A Longitudinal Analysis on the Incidence of Over-education among Immigrants and its Impacts on Earnings

Le Wen† Sholeh Maani‡

The University of Auckland

Paper Presented at the New Zealand Association of Economists (NZAE)

55th Annual Conference, 2-4 July 2014

Abstract

This paper uses longitudinal analyses based on HILDA* to investigate the extent of matching between education and occupation and resulting earnings effects among immigrants in Australia. The panel approach based on nine years of longitudinal data addresses individual heterogeneity effects that are important to over-education analysis, and thereby extends the international literature. Correlated Random effects logit results suggest that both ESB (English speaking background) and NESB (Non-English speaking background) immigrants have high incidence rates of over-education. Longitudinal analyses show that overeducated NESB workers suffer a large earnings penalty from education-occupation mismatches and skill under-utilisation impedes their assimilation process.

† PhD Candidate, Department of Economics, the University of Auckland, New Zealand, l.wen@auckland.ac.nz

‡ Professor of Economics, Graduate School of Management, the University of Auckland, s.maani@auckland.ac.nz

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

1. Introduction

It is a commonplace occurrence to hear of immigrants being employed in occupations that are below the level of their educational attainment; such as those from professional occupations driving taxis or working as kitchen hands. What is the extent of this phenomenon across host countries? What are the determinants of this disadvantageous situation among immigrants? How can immigrants' skills be used to full advantage? A topic of significant debate among researchers and policy makers has been immigrant adjustment, assimilation, and success in their new labour market.

This study uses longitudinal data to examine labour market outcomes for immigrants in the Australian labour market.

During a two-year period (2005-2006), approximately 48,865 skilled migrants, 45,290 family migrants and 14,140 humanitarian migrants arrived in Australia. The number of skilled migrant visas issued in 1998-99 was 35,000; this increased to 97,340 in 2005-06. Of these, 17% of permanent arrivals came from the United Kingdom and 11% came from New Zealand.

> *"Skilled visa holders were the most likely to be employed after arriving in Australia. Humanitarian visa holders were the least likely to be employed. However, the longer an immigrant remained in Australia, no matter what their visa class, the more likely they were to be in employment." (DIMA 2007)*

The evidence shows that the Australian immigration policy has placed greater emphasis on skill based immigration because skilled immigrants are more employable and productive than their unskilled counterparts. Thus, they are therefore likely to increase Australia's productive capacity. However, if immigrants cannot work in occupations that fully utilise their skills, this productivity gain is reduced. The cause of "the unrecognised skills of immigrants" is the mismatching of educational attainment and the educational requirements for migrants prospective occupations in the host country, generally referred to as over-education. When compared to the native-born, immigrants are more likely to be over-educated and to suffer an earnings loss and therefore experience individual earnings disadvantage (See for example, Chiswick and Miller, 2008). Moreover, a potential loss to the economy as well as a significant burden on new arrivals may be caused (Ferrer & Riddell, 2008).

Over-education is defined as the extent of someone's actual education exceeding the educational requirement to perform his/her job. Because the HILDA data does not provide any questions on over-education, workers' self-reports (SR) are not applicable. Thus, the required years of education for a particular occupation can be defined by using a cross-wave Mode measure; this measures the number of years of education required to undertake a position of employment; the number varies between waves. The amount of education that most commonly occurs within an occupational category is calculated for each wave. The required years of education for all nine waves, are derived by combining the Mode education of all the waves; next, the years of over-education and the years of under-education are obtained by comparing the actual years of education with the required years of education.

By employing the procedure described, it was found that the incidence of over-education differed considerably between the native born population and the immigrants. In particular, immigrants were shown to have a higher probability of being overeducated than natives[1]. Non-English-Speaking background (NESB) immigrants were found to suffer especially from extremely high levels of educational mismatch. For example, the incidence of over-education ranged from 24 per cent to 28 per cent among natives. However, among English-Speaking background (ESB) immigrants it was 3 to 10 per cent higher, ranging from 28 per cent to 36 per cent; the incidence of over-education was 36 per cent to 48 per cent among NESB immigrants. This is 17 per cent to 21 per cent higher than for the native born population, depending upon the specific year of assessment. It was also found that ESB immigrants earn a premium wage and NESB immigrants suffer loss of earnings when compared to natives.

A number of questions arise from these findings. Why is it that immigrants have a higher incidence of over-education than natives? What are the determinants of educational mismatch? What is the relationship between earnings and over-education? Does over-education have a negative effect on earnings? Why do NESB immigrants earn less than natives? Can NESB immigrants reduce their earnings disadvantage with years since migration? These questions have motivated the research reported in this paper.

To date, immigrants' over-education is under-researched in Australia. This study makes the following contributions to the international literature: It investigates the determinants of over-education among immigrants in Australia, and the extent of the impact of over-education on earnings after accounting for individual heterogeneity. We use the Correlated Random Effects (CRE) logit model, and fixed effects earnings models to address endogeneity and individual heterogeneity. To the best of our knowledge, it is the first examination of the determinants of over-education and its impact on earnings among immigrants using longitudinal techniques based on panel data.

The over-education of immigrants is examined from the following perspectives:

Country of origin and language proficiency

In a study of immigrant assimilation, country of origin is of importance. Immigrants from different countries have differing assimilation rates in the host country. Immigrants from a background that is similar to that of the host country are more likely to have similar incidence rates of over-education due to the higher rate of transferability of human capital. However, those from a non-English speaking background may find it more difficult to settle down, which could produce serious over-education rates. The over-education rate of immigrants may not converge with the rate of natives, even after a lengthy period of residence.

As English is the main language in Australia, the English proficiency of immigrants may help them to obtain education-occupation matched jobs. Compared to NESB immigrants, in the host country, ESB immigrants would expect to face similar labour market conditions to those of their country of origin. Their prior migration experience and education may be portable to the host countries. As a result, relative to NESB immigrants, ESB immigrants may adapt to new environments quickly, and be more likely to find a matched job.

In this study, an immigrant is defined as a person who was born overseas. People born overseas are asked whether English is the first language they learned to speak as child[2].

[1] Natives in this paper refer to people who were born in Australia, and this applies to the entire study.

[2] This variable is constructed for the population born overseas. The survey asked: Is English the first language you learned

If English was the first language learned, the immigrant is defined as an ESB immigrant, otherwise, as a NESB immigrant. Thus, the sample is divided into three subsamples: Natives, ESB immigrants and NESB immigrants.

Chiswick and Miller (2009) provided evidence of strong positive relationships between English-speaking proficiency and occupational attainment.

Transferability of human capital

Immigrants generally demonstrate high rates of over-education due to the imperfect transferability of human capital in the host country. Thus, the over-education rates of immigrants signify education-occupation matching difficulties in the host countries' labour market, and they reflect an important dimension of immigrants' assimilation (Friedberg, 2000).

With time, gaining local experience or investing in local education may help immigrants to improve educational and job matches, reduce the rates of over-education, and decrease the potential earnings' penalty.

Age at migration

Migrating as a child or as an adult may give rise to differing effects on the incidence of over-education. Young immigrants are more likely than adults to adapt to their new country of residence. Thus, they behave similarly to a member of the local population.

The following questions are addressed in this study:

> *To what extent are immigrants and natives over-educated? Does the incidence of over-education among immigrants vary by country of origin, English proficiency, and age on arrival?*
>
> *Are there differing impacts of over-education on earnings between sub-groups based on country of origin and English proficiency?*

To estimate the effects of over-education on immigrants' assimilation effects, we examine the following hypotheses:

1. NESB Immigrants are more likely to be overeducated in relation to ESB immigrants at the time of arrival.
2. As time passes, by gaining local experience, the over-education rates of immigrants converge to the rates of the native born population, and the immigrant earnings differential relative to that of the native born decreases. Therefore, the coefficients of years of experience and YSM (years since migrating to Australia) are predicted to be negative with over-education, and positive with earnings.

The remainder of this paper is organised as follows: Section 2 provides an overview of recent immigrants' over-education literature, and it identifies the main factors affecting immigrant mismatch and labour market outcomes in the host country. Section 3 develops the econometric framework. Section 4 outlines the data and variables. The Results are presented in Section 5, and in Section 6, conclusions are drawn.

---

to speak as a child? Answer 1-English was first language learned; 2- English was not first language learned.

2. Review of the literature

A number of studies have examined over-education among immigrants in different countries. Regardless of host country and official language, these studies have shown that immigrants have a high incidence rate of over-education, ranging, from 16 per cent (Kler, 2007) in Australia to 96 per cent (Aringa and Pagani, 2010) in Italy. In addition, there is emerging evidence that immigrants suffer an earnings loss from education-occupation mismatches (Chiswick and Miller, 2006; Kler, 2007; Green, Kler and Leeves, 2007; Lindley, 2009; Wald and Fang, 2010).

To date, few studies have been conducted on immigrant assimilation in the Australian labour market. Based on the 2001 Census of Population and Housing, Chiswick and Miller (2006) reported that NESB immigrants have a lower rate of return to schooling accompanied by over-education and under-education. The payoff to years of schooling for Australian-born males is 8.8 percent. For ESB immigrants and NESB immigrants it is 8 percent and 5.9 percent, respectively. However, there is the same payoff to required years of schooling of 15.2 percent for these three groups. The earning effects of over-education (under-education) is 5 to 6 (-3 to -4) percent for both the Australian-born and ESB immigrants, and it is about 3 (-1) percent for NESB immigrants.

Based on longitudinal data for immigrants to Australia (LSIA), Green, Kler and Leeves (2007) examined the determinants of employment and over-education. They also studied the return to required schooling and surplus schooling by two cohorts among male immigrants aged 15-64. They found that immigrants, even those with skill-assessed visas are more vulnerable to over-education than natives. NESB Immigrants are more likely to be over-educated, with the incidence of over-education being between 32% and 49%. NESB Immigrants also have lower returns to required and surplus education than do natives. Tighter welfare and support policies[3] for immigrants may increase employment at the expense of under-utilising their skills. However, their sample is limited to recent immigrants in their sample (arriving in 1993, 1995, 1999, and 2000). The analysis employed OLS estimation.

Using the same LSIA dataset with the addition of the inclusion of both genders, Kler (2007) examined the effects of over-education among tertiary educated immigrants. The evidence is in line with Green, Kler and Leeves (2007). The incidence of over-education is similar between ESB immigrants and the native born population, and is higher among Asian NESB immigrants. The rate of over-education is around 16% for ESB immigrants. Among Asian immigrants, approximately 50% are over-educated. Among other NESB immigrants, the rate of over-education is close to 40%. The payoff for over-education is much smaller than the payoff to required education. There is no significant effect of over-education on earnings among Asian immigrants.

Green, Kler and Leeves (2007) and Kler (2007) used a bivariate probit model to examine the incidence of over-education, and an augmented human capital earnings model (Frenette,2004) to examine earning effects in the Australian labour market. They focused on the effects of visa category and labour market conditions.

This study extends Green, Kler and Leeves' (2007) work and it contributes to the Australian literature as follows. We extend the analysis to panel data, and we employ a Correlated

[3] For example, stringent entry standards were applied to skill and English language ability test scores and eligibility to claim welfare and unemployment benefits was extended from 6 to 24 months (except for humanitarian visa holders).

Random Effects (CRE) logit model with Mundlak correction (1978) to examine the incidence of over-education by focusing on the effects from years since migration and age at migration. The endogeneity due to the correlation between explanatory variables and error terms is addressed by Mundlak correction. We also employ both panel fixed effects (FE) and random effects (RE) models to examine the effects of over-education on earnings from years since migration. The latter aspect of our study on the effects of years since migration on over-education and earnings and the panel feature of the analysis extend the international literature on the subject.

In Spain, the effects of years since migration have been examined by Fernández and Ortega (2008). They used data from the Spanish Labour Force Survey for the period 1996-2006, and showed that compared to the rates for the native-born, immigrants experience initially higher participation and unemployment rates, and have a higher incidence of over-education and temporary contracts. Over a 5-year period, immigrants' participation rate was shown to be reduced to that of those who are native-born and unemployment rates to levels even lower than those of the native-born. The incidence of over-education and temporary contracts however remained constant.

Moreover, the portability of the human capital of immigrants into the Spanish job market was studied by Sanroma, Ramos, and Simon (2008) .They suggested that geographic origin has an influence on the transferability of human capital. Immigrants from countries that are highly developed, or have a similar culture or language to that of the host countries, have higher transferability levels.

Similar evidence is also found in the study of the Italian labour market by Aringa and Pagani (2010). Based on data from the Italian Labour Force Survey for the years between 2005 and 2007, Aringa and Pagani found that foreigners arriving in Italy are much more likely to be over-educated than are the natives. Furthermore, experience acquired in Italy did not help to improve their occupation-education match. The researchers suggested that foreigners struggle to catch up with the natives even if they adapt their skills to the host countries.

Age at arrival is expected to have a negative effect on immigrant earnings. This was shown by Friedberg (1992), who found that there was an 11.6 per cent earnings disadvantage between an immigrant who arrived in the United States at age 30 and a comparable immigrant who had migrated at age 10.

3. Econometric framework

A longitudinal analysis is applied in this study to address the potential problem of "omitted unobservable bias" from cross-sectional analysis, which is important to identifying both the incidence and potential earning penalty to over-education.

In order to obtain the estimates for comparison between the Australian-born and immigrants, two samples are examined in this study. One sample consists of the Australian-born (natives) and ESB immigrants, and the other sample is natives and NESB immigrants.

This approach allows us to examine results for each immigrant group compared to the same base category of the Australian-born.

Part 1: Determinants of over-education

We apply the correlated random effects logit model to examine the likelihood of over-education with panel data. In this model a number of important variables, such as immigrant status, are time-invariant. A conditional logit (or fixed effects logit) model which was also considered, sacrifices time-invariant but potentially important information on any individual who presents no change in dependent variables by eliminating time-invariant variables. However, this model benefits from controlling for the endogeneity from individual effects. The random effects logit model, is in comparison able to estimate the coefficient of time invariant variables whilst also allowing for dynamic adjustment. Thus, based on these considerations, we have chosen the random effects logit model to examine the determinants of over-education.[4]

A potential problem arises from the biases occurring in the correlation between explanatory variables and error terms. We address this problem by using the Mundlak correction (Mundlak, 1978).

In the following latent model, β is unbiased if explanatory variables x_{it} and individual specific effects μ_i are independent, that is

(1) $$y_{it}^{*} = x_{it}\,\beta + \mu_i + \varepsilon_{it}\text{ , Where } E[\mu_i|X_i] = 0, and\ \varepsilon_i|\ X_i \sim \text{N}\ (0,\sigma_\varepsilon^2).$$

To relax this assumption, the Mundlak model (Mundlak, 1978) proposes individual effects μ_i as a function of individual means, that is $\mu_i = \bar{X}_i\,\delta + \eta_i$, where $\eta_i|\ X_i \sim \text{N}(0,\sigma_\eta^2)$. It assumes zero correlation between $\bar{X}_i$ and η_i.

Thus, we have $E[\mu_i|X_i] = \bar{X}_i\delta$, where $\bar{X}_i$ is an average of x_{it} over time for individual i, and it is time invariant.

We rewrite the above latent model as

(2) $$y_{it}^{*} = x_{it}\,\beta + \bar{X}_i\delta + \left[\varepsilon_{it} + \mu_i - E[\mu_i|X_i]\right] = x_{it}\,\beta + \bar{X}_i\delta + u_{it}\text{ , where}$$

u_{it} is the new error term. By this construction, we have

(3) $$E[u_{it}|X_i] = E[\varepsilon_{it} + \mu_i - E[\mu_i|X_i]\ \ |X_i] = 0$$

Mundlak's approach is used to control for endogeneity effects due to unobserved individual effects. It is considered as a compromise between the fixed and random effects models. It also provides a test for adjustment for endogeneity as an alternative to the Hausman test--If the coefficient on group mean δ is non-zero, that suggests that individual effects are not to be ignored (Greene, 2010).

Applying the Mundlak correction (1978), the unobserved individual effect μ_i is conditional on the means of time varying explanatory variables.

(4) $$\mu_i = \bar{X}_i\delta + \eta_i\ \ where\ \eta_i|X_i \sim \text{N}\left(0,\sigma_\eta^2\right)$$

Thus, the model is written as:

[4] Chamberlain (Chamberlain, 2010) has shown that 'logit' rather than 'probit' can achieve root and consistency in a fixed effects model. In other words, the probit setup is not available in a fixed effects model, and would not allow the test of fixed-effects versus random-effects.

$$(5) \qquad y_{it}^{*} = x_{it}\,\beta + \bar{X}_{\iota}\delta + \eta_i + \varepsilon_{it}$$

It is noted that coefficients δ will differ between panels of different lengths T and they are specific to the particular sample. The estimates of β approximate the fixed effects estimators, as shown by Wooldridge (2009).

In this study, we employed both a random effects logit model and a correlated random effects logit model with Mundlark correction to estimate the determinants of over-education. As noted earlier, we consider effects for natives and ESB immigrants, and among natives and NESB immigrants, respectively. The random effects logit model is applied as a benchmark. The endogeneity issue due to the individual effects is corrected by the correlated random effects logit model with Mundlak correction. If the results from these two models are significantly different, then endogeneity is adressed by the correlated random effects logit model.

We examine the hypothesis that the incidence of over-education for immigrants may decrease with their duration of stay (YSM) in Australia. This less-examined hypothesis has important implications for understanding the labour market assimilation of immigrants in earnings models.

Model 1: Determinants of over-education

The functional form of logit model is written as:

$$(6) \qquad ln\left(\frac{Pr(overeducation_{it})}{1 - Pr(overeducation_{it})}\right)$$
$$= \delta_0 + \delta_1 \mathrm{Z_{it}} + \delta_2 \mathrm{M_i} + \delta_3 ED_{it} + \delta_4 DQUA_{it} + \delta_5 (DQUA_{it} * M_i) + \delta_6 YSM_{it}$$
$$+ \delta_7 YSM_{it}^2 + \delta_8 EXP_{it} + \delta_9 EXP_{it}^2 + \sum_{j=1}^{m} [\bar{X}_{\iota} \;\; \delta_j] + \eta_i + \varepsilon_{it} \;\; ,$$

$$\eta_i \sim \mathrm{N}(0, \sigma_\eta^2); \qquad \varepsilon_i \sim \mathrm{N}\,(0, \sigma_\varepsilon^2)$$

$$i = 1, \dots, N;\; t = 1, \dots, T; j = 1, \dots, m$$

By this logit model setup, the natural log of the odds ratio of over-education is explained by a quadric function of years since migration (YSM) with other explanatory variables. The observed variable $overeducation_{it}$ takes the value of 1 if worker *i* is overeducated and is defined as 0 otherwise. $\mathrm{Z_{it}}$ denotes a set of personal or job characteristics of individual *i* at time period *t*; ED_{it} denotes actual years of education obtained by individual *i* at time *t*. M_i is a dummy variable, and it takes the value of 1 if individual i is an immigrant, 0 otherwise. The coefficient of M_i, δ_2, measures the initial over-education gap of immigrants upon arrival relative to comparable natives. YSM_{it} denotes the number years of residence since migrating to the host country. The coefficient of YSM_{it}, δ_6, measures the way in which the over-education gap varies as immigrants spend time in the host country. The over-education rates of immigrants are expected to signify their levels of assimilation. Therefore, the coefficient of YSM_{it} is predicted to be negative. δ_7, the coefficient of YSM_{it}^2 examines the rate of over-education in a linear or quadric style over time. $\sum_{j=1}^{m}[\bar{X_\iota}\ \delta_j]$ represents the Mundlak adjustments (where is *m* is the number of explanatory variables).

The unobservable individual specific μ_i as a function of individual means, that is

$\mu_i = \sum_{j=1}^{m}[\bar{X_\iota}\ \delta_j]\ + \eta_i$, where $\eta_i \sim$ N(0,σ_η^2).

It assumes zero correlation between the means of time varying explanatory variables and η_i. And $\varepsilon_{i,t}$ denotes the disturbance terms, which are assumed to be independent and identically distributed (*iid*).

To further examine the effects of age on arrival on the probability of being overeducated, we replace YSM_{it} and YSM_{it}^2 with age on arrival dummy variables, respectively, in order to avoid an over-specification problem.

Part 2: Impacts of over-education on earnings

Unobserved heterogeneity, such as unobserved ability, motivation or work efforts influence earnings, and also are correlated with observed education and skills. If these unobserved individual effects, u_i, are correlated with explanatory variables, cross-section analysis would result in omitted unobservable biases. Longitudinal data captures the same individual over time. Thus, unobservable individual effects are eliminated by using a panel fixed effects model. Thus, estimation results from fixed effects models are consistent. However, this model cannot evaluate the time-invariant explanatory variables because they are removed by within-group transformation. In contrast, a random effects Generalised Least Squares (GLS) model assumes that u_i is uncorrelated with explanatory variables in which GLS uses the optimal combination of within-group and between-group variations. If individual effects do not matter, then the GLS estimator is equal to the ordinary least squares (OLS) estimator. A Hausman test is used to identify whether the random effects GLS estimator is biased.

In this section, the analysis focuses on the link between over-education and earnings. The following questions are of interest in the empirical analysis. How does over-education impact, directly or indirectly, on earnings via years since migration and migration status? Is the impact of over-education on earnings affected by unobserved heterogeneity, such as, personal ability or variable quality or under-valuation of immigrant qualifications?

Based on the standard Over-education, Required-education, Under-education (ORU) earnings model, the extended earnings' model is applied into this study for issues of interest.

The standard ORU (Over-education, Required education and Under-education) earnings model (as originally proposed by Duncan and Hoffman (1981)) is widely used in 'over-education' empirical research. It was proposed by. The ORU model decomposes actual years of education (S_a) into required years of education (S_r), years of over-education (S_o), and years of under-education (S_u). Thus $S_a = S_r + S_o - S_u$, where $S_o = S_a - S_r$ for the over-educated (i.e. if $S_a > S_r$), and 0 otherwise. Similarly, $S_u = S_r - S_a$ for the under-educated if (i.e. $S_r > S_a$), and 0 otherwise.

Then the log of earnings in the ORU model can be written as:

$$(7) \qquad lny = \alpha_1 + \beta_r S_r + \beta_o S_o + \beta_u S_u + \delta_1 X_1 + \varepsilon$$

lny is the natural logarithm of earnings, X_1 is a vector of a variety of other control variables that generally includes personal characteristics and job characteristics, S_r, S_o, S_u are, respectively, the years of required education, over-education, and under-education. $\alpha_1$is the intercept term, and ε is an error term.

Equation (7) estimates β_r, β_o, β_u continuously, and β_r, β_o, β_u are the rates of returns to required education, over-education and under-education respectively.

Prior literature on 'over-education' has consistently found that $\beta_r > \beta_o$ and $\beta_o > 0$, such that the return of over-education is lower than the return to required education; and the return to over-education is positive (Cohn, 1992; Groot, 1996; Rumberger, 1987; Sicherman, 1991). In contrast, they also found that $\beta_u < \beta_r$ and $\beta_u < 0$, which means the return to under-education is lower than the return to required education; and that it is a negative return (Hartog, 2000).

In panel data settings the ORU model is expressed as follows:

$$(8) \qquad lny_{i,t} = \beta_r S_{i,t}^r + \beta_o\left(S_{i,t}^a - S_{i,t}^r\right) + \beta_u\left(S_{i,t}^r - S_{i,t}^a\right) + \delta X_{i,t} + \alpha_i + \varepsilon_{i,t}$$

$$i = 1, \ldots, N; \; t = 1, \ldots, T$$

Where $lny_{i,t}$ denotes the hourly wage from main job of individual i at year t; $X_{i,t}$ is personal characteristics and job characteristics of individual i at year t; α_i denotes the unobservable individual-specific effect and $\varepsilon_{i,t}$ denotes the remainder disturbance, assumed independent and identically distributed i.i.d (0, σ_ε^2). $S_{i,t}^a$ denotes the years of actual education for individual i at year t and $S_{i,t}^r$ is the years of required education for individual i at year t. Thus, ($S_{i,t}^a - S_{i,t}^r$) is the years of over-education when $S_{i,t}^a > S_{i,t}^r$; 0, otherwise. Likewise, $\left(S_{i,t}^r - S_{i,t}^a\right)$ is years of under-education when $S_{i,t}^r > S_{i,t}^a\,; 0\; otherwise$. β_r is the rate of returns to required education, β_o is the rate of return to over-education and β_u is the rate of penalty to under-education.

The extended ORU earnings model is built by adding interaction terms to Equation (8) to examine the impacts of educational mismatch, years since migration and migrant status on the return to over-education, after controlling for the individual effects. By doing so, we can examine the earnings gap between immigrants and natives via educational mismatch.

These results reveal an added and less-studied explanation for the existing earnings disadvantage for immigrants in the Australian labour market.

Model 2: The extended ORU earnings model (over-education earnings impact via years since migration and occupation)

$$(9) \quad lny_{i,t} = \beta_r S_{i,t}^r + \beta_o(S_{i,t}^a - S_{i,t}^r) + \beta_u(S_{i,t}^r - S_{i,t}^a) + \beta_{rM}(S_{i,t}^r \times \mathrm{M}) + \beta_{oM}[(S_{i,t}^a - S_{i,t}^r) \times \mathrm{M}] + \beta_{uM}[(S_{i,t}^r - S_{i,t}^a) \times \mathrm{M}] + \theta_1 \mathrm{Z_{it}} + \theta_2 \mathrm{M_i} + \theta_3 YSM_{it} + \theta_4 YSM_{it}^2 + \sum_{k=1}^{2}[\theta_{kYSM}(TYP_{k,it} \times \mathrm{YSM}_{it}) + \theta_{kYSM2}(TYP_{k,i} \times YSM_{it}^2)] + \mu_i + \varepsilon_{\mathrm{it}}$$

$$i = 1, \dots, N; \; t = 1, \dots, T; k = 1,2$$

The error term is denoted by $\mu_i + \varepsilon_{i,t}$. The unobservable individual-specific effect μ_i is assumed not to change over time, and the random disturbance, ε_{it}, is assumed to be independent and identically distributed, i.i.d (0, σ_ε^2).

$\mathrm{Z_{it}}$ denotes a set of personal characteristics, such as years of experience. $lny_{i,t}$ is the natural log of hourly wage from main job in constant (2009) dollars for the *i* th individual in period t.

β_o, β_u, and β_r estimate the magnitude of earnings effect of a one unit change in the years of over-education, years of under-education, and the required years of education, respectively among natives.

The coefficient of the interaction terms, $\beta_{\mathrm{oM}}, \beta_{\mathrm{uM}}, \beta_{\mathrm{rM}}$ evaluate the difference of earnings effects between natives and migrants who have the same type of educational mismatch.

$TYP_{k,it}$ is a binary variable[5], which corresponds to the three types of educational mismatch. k takes the value of 1 if individual *i* is overeducated at time period t, and 0 otherwise. k equals 2 if individual *i* at time t is undereducated, 0 otherwise. Educationally matched is the reference category.

The coefficient of M_i, θ_2, denotes the initial earnings gap between immigrants and natives.

YSM_{it} denotes the number of years of residence since migrating to the host country for individual i at time t. The coefficient of YSM_{it}, θ_3, denotes assimilation effects. Based on previous studies, θ_3 is expected to have a positive sign. The significance of coefficient YSM_{it}^2 reveals the linear or quadratic relationship between earnings and the number of years since migration. If immigrants work in jobs requiring qualifications that are below their educational attainment, this may lengthen their assimilation process with the consequence that they catch up with the natives' earnings more slowly over time, or not at all. Thus, the

[5] See Verdugo and Verdugo (1989) for the applications of these dummy variables specifications in cross-section data.

coefficient of interaction terms, $TYP_{1,it} \times YSM_{it}$, θ_{1ysm}, is negative if over-education slows immigrants' earnings assimilation in the host-country.

4. Data and Variables

4.1 Data

The data used to examine the incidence of over-education and immigrants' assimilation in Australia is taken from the wave 1 to wave 9 (year 2001-year 2009), *responding person* file of the Household, Income and Labour Dynamics in Australia (HILDA) Survey. The HILDA Survey, Australia's first nationally representative household panel survey, began in 2001 and interviews are now conducted annually. This longitudinal survey overcomes the disadvantages of cross-section surveys. It is designed to follow the same individuals over time, and it allows researchers to analyse the dynamics of change at individual and household levels.

The sample for the current study includes all full-time[6] male workers, who were aged from 23 to 64 in the initial survey year. In order to make full use of the panel data features we used a balanced data set to select the observations who had taken part in each year of the survey. With pooled 2001-2009 data, the full sample size used in this study is composed of 18,250 observations of 2,732 individuals. Among the employed (17,644 observations), 90% is employed full-time.

Workers in part-time jobs may have chosen to do so for reasons of family or other personal commitments or preferences. Therefore, part-time workers may be more likely to accept mismatched jobs in terms of education and occupation match in exchange for other job characteristics, such as the flexibility of hours of work, or shorter distances to work. These supply side job mismatches are less likely to affect workers' work attitudes and behaviour. Thus, these mismatches are less likely to reduce workers' productivity and result in wage penalties. In addition, part-time jobs are also shown to have a different pay structure which adjusts for other job-related fringe benefits. Therefore, we consider full-time workers for a more comparable group of employees and earnings scales. In the initial stages of the study the potential impact of selection into both employment and also full-time employment was examined using a Heckman selection adjustment. The results showed that control for selection for either selection did not change the results.

Of this full-time sample, 79 per cent are native-born and 21 per cent are immigrants. Among the immigrants surveyed, 13 per cent have English as their first language and 8 per cent did not learn English as their first language. Most ESB immigrants come from developed countries such as the United Kingdom (50 per cent), New Zealand (23 per cent), South Africa (3 per cent) and the United States of America (3 per cent). Unlike the ESB immigrants, NESB immigrants are diverse, coming from over 60 different countries, including Vietnam (13 per cent), China (including Rep, Hong Kong and Taiwan, 10 per cent), India (6 per cent), Philippines (5 per cent), and The Netherlands (4 per cent).

The mean characteristics of the full-time samples of Natives, ESB and NESB immigrants are shown in Table 1. There is significant difference between these three groups across a number of personal characteristics. The average age of ESB immigrants is 44.70 years,

[6] At an early stage, we also examined the incidence of over-education and its effects on earnings for the entire employed sample; this was achieved by using the Heckman selection model to control for sample selection issue. Results are not sensitive to the sample selection. The Heckman adjustment did not alter the results. Results are available on request.

about one year older than the NESB immigrants (43.81) and four years older than natives (41.12). The mean of age on arrival in Australia is at age 18 for ESB immigrants, and at age 23 for NESB immigrants. Thus, the mean of years since migration is 26.32 for ESB immigrants and 20.65 for NESB immigrants, which indicates that ESB immigrants have been in Australia six years longer than NESB immigrants. The hourly wages for the main job are found to be slightly higher for natives ($29.81) than for NESB immigrants ($29.50) but lower than for ESB immigrants ($32.49).

Table 1: Summary Statistics by Country of Birth

	Native		**ESB**		**NESB**	
VARIABLES	mean	sd	mean	sd	mean	sd
Personal Characteristics						
Age	41.12	9.96	44.70	9.91	43.81	9.79
Disability/Impairment	0.13	0.33	0.14	0.34	0.11	0.31
Poor English	/	/	/	/	0.04	0.21
Age on Arrival	/	/	18.38	12.08	23.16	11.56
Age 0-12	/	/	0.40	0.49	0.22	0.41
Age 13-22	/	/	0.17	0.37	0.24	0.42
Age 23-34	/	/	0.34	0.47	0.38	0.49
Age 35-60	/	/	0.10	0.29	0.16	0.37
Years since Migration-YSM	/	/	26.32	13.08	20.65	12.65
YSM2/100	/	/	8.64	7.36	5.86	6.94
Job Characteristics						
Unemployment Rate	0.05	0.01	0.05	0.01	0.05	0.01
Hourly wage of main job	29.81	15.75	32.49	18.23	29.50	15.24
Log(Hourly wage of main job)	3.28	0.50	3.34	0.54	3.27	0.49
Human Capital						
Years of experience (total)-EXP	21.36	10.30	24.63	10.38	23.25	10.44
$EXP^2/100$	5.62	4.74	7.14	5.22	6.49	4.98
Years of actual education (total)-ED	13.76	2.40	14.08	2.55	14.57	2.52
With Qualification	0.69	0.46	0.71	0.45	0.74	0.44
Highest Qualification						
Postgraduate	0.11	0.31	0.17	0.37	0.20	0.40
Bachelor	0.15	0.36	0.16	0.37	0.24	0.43
Diploma	0.10	0.30	0.10	0.31	0.13	0.33
Certificate	0.34	0.47	0.28	0.45	0.16	0.37
Australian qualification	0.69	0.46	0.41	0.49	0.43	0.49
Overseas qualification	/	/	0.30	0.46	0.31	0.46
Without Qualification	0.31	0.46	0.29	0.45	0.26	0.44

Table 1 presents further detailed information with respect to education and experience. Both ESB and NESB Immigrants in full-time employment have more years of work experience than natives, and ESB immigrants have slightly higher years of work experience than NESB immigrants. On average, immigrants are better educated than natives. The educational attainment is highest among NESB immigrants (14.57). Among ESB immigrants, the average number of years of education is a total of 14.08 years. Natives have 13.76 years of educational attainment which is lower that for immigrants.

It is worth noting that although the NESB group has a higher average years of education (14.57) than the ESB (14.08) group and natives (13.76), their average number of required years of education to perform a job is slightly lower for NESB (14.30) immigrants than it is for ESB (14.39) immigrants and higher than for natives (14.25). NESB workers earn less than ESB workers and natives. This evidence encourages the test of the hypothesis that NESB immigrants are more likely to undertake jobs in which they are over-educated in comparison with ESB immigrants and natives. In contrast, among both natives and ESB immigrants, the required years of education to perform a job exceed their actual years of education. This implies that they are more likely to have higher level jobs and earn more.

Other measures of qualification are used based on credentials obtained also show that immigrants are highly educated. 44 per cent of NESB immigrants and 33 per cent of ESB have qualifications above a Bachelor degree; in contrast, only 26 per cent of Australians have obtained these qualifications.

Most ESB immigrants come from advanced countries and their qualifications are valued in Australia. However, NESB immigrants may experience more difficulty in adapting to their new lives even if they work in skilled categories. Furthermore, NESB immigrants may work in occupations that require lower levels of educational attainment in instances in which their overseas credentials are not recognised by Australian employers.

4.2 Variables

HILDA does not provide direct information for variables of interest, thus they are derived from the relevant variables.

The earning variables used in this study are log hourly wage from main job. To derive the hourly wage for main jobs, the first step is to convert nominal earnings to real earnings. We use 2009 as a base year, reference ABS CateNo6345.0 labour price index, and generate real earnings for each year by using nominal earnings divided by the wage price index. To account for non-responding (in responding households) persons' wages which are presented as missing data, the variable we choose is imputed weekly gross wages and salary for the main job[7]. After converting the imputed nominal weekly gross wages and salary from the main job to real imputed weekly gross wages and salaries, the hourly wage from main job is derived by using imputed real weekly gross wages and salary from the main job divided by combined hours per week usually worked in the main job. Then we convert the hourly wage into log hourly wage.

[7] Imputation methods are used to deal with missing cases. Since income is a sensitive issue for some people who do not report their income in interview, thus missing data occurs. Nearest Neighbour Regression imputation and little and Su imputation are applied to the imputation of data for responding persons. A full description of the imputation process for the income variables is provided by Hayes and Watson (2009).

Years since migration (YSM) measures years of duration in Australia for immigrants.

Years of actual education are derived by four variables from HILDA. To evaluate the effects of qualification, we categorise qualifications into five categories: Postgraduate, Bachelor, Diploma, Certificate, and No qualification. Postgraduate includes Doctorate, Masters, Graduate Diploma, Graduate certificate and Bachelor with Honours; this requires over 17 years of education. Bachelor covers a Bachelor degree without Honours and takes 16 years of education to achieve. Diploma includes Advanced Diploma and Diploma and requires15 years of education. Certificate includes Certificate I, Certificate II, Certificate III or Certificate IV; these require over 13 years of education. 'No qualification' covers workers without qualifications, representing less than 13 years of education.

Age at migration is assumed to have an effect on assimilation. Wilkins (2003) examined the impact of age at migration for Australian immigrants by using data from the Australian Bureau of Statistics Education and Training Survey (ETS) 1997. Empirical results show that younger arrivals have lower initial earnings but faster earnings growth compared to older arrivals. If the age on arrival is between 1 and 6, then this group of young immigrants is more likely to come to Australia with their adult parents who are the migration decision makers. This young arrival group is assumed to have no initial stock of human capital and to accumulate their human capital after migration, thus they become more likely to perform similarly to natives. If the age on arrival in Australia is over 6 years, immigrants are more likely to have received education overseas and have an initial stock of human capital, but their human capital obtained elsewhere may be less valued in Australia. They are more likely to face difficulties when entering into the labour market, such as; having unrecognised educational qualifications, poor knowledge of the domestic labour market, and a low level of English proficiency. Previous research has found that elementary school education is equally valued and is quite portable across national boundaries (Friedberg, 2000). Therefore, we define four cohorts based on their age at migration: 0-12, 13-22, 23-34, and 35-60. Notably, the distribution of poor English among NESB immigrants increases with age at migration, which suggests as expected that language proficiency is affected by age on arrival.

As English is the main language in Australia, NESB immigrants with difficulties in English are more likely to decrease their expectations while job searching, and to accept jobs which require education below their level of attainment. Therefore, proficiency in spoken English may have a significant effect on the rate of over-education and on immigrants' assimilation. We collapse four classifications into two: those who speak English well, and those who speak English poorly[8].

The unemployment rate represents the percentage of the labour force that is currently unemployed and actively looking for work. It is also a common indicator of a country's economic conditions. It is used as a control for labour market conditions. We have collected the annual unemployment rate (year 2001 to year 2009) from the Australian Bureau of Statistics (ABS) as a reference. Higher unemployment rates may force some workers to accept mismatched employment positions due to the limited availability of positions. Alternatively, when the unemployment rate is high, those who remain in employment may be those who are in better matched position, such that the incidence of mis-match decreases with unemployment. This variable is an annual rate.

[8] The variable Hgeab in HILDA asks 'How well do you speak English?' among the population who speaks other language at home. Answer 1-very well; 2-well; 3-Not well; 4-Not at all.

4.3 Extent of over-education

The over-education measure in our analysis is based on the Mode method and it is derived at the two digit occupational category level for greater accuracy. In the initial stages of this study we evaluated four alternative measures of over-education. The Mode method was adopted as the preferred method based on the literature that generally favours the Mode method. In particular, in panel analysis the cross-wave mode is more appropriate for defining the required education when compared to the other three measures.

Alternative measures are based on: cross-wave mode (Mode) as adopted here, mean plus one standard deviation (Range-one), mean plus half standard deviation (Range-half) and Job Analysis (JA). Job Analysis (JA) is not updated over time, and there is lack of consideration for the heterogeneity of jobs. Range-one (mean plus one standard deviation) and Range-half (mean plus half standard deviation) represent the symmetry between over-education and under-education; and the cut-off points of one standard and half standard deviation are arbitrary.

The Job Analysis (JA) measure is a systematic evaluation by professional job analysts who specify the level and type of education required based on grading the occupation. This measure is derived from information in regard to the respondents' occupations. For example, the *Dictionary of Occupational Titles* (DOT) (U.S. Department of Labour 1965) developed by the United States (U.S.) Employment Service, contains detailed descriptions of all occupations in the U.S. economy and information on a number of occupational characteristics, the Standard Occupational Classification System (SOCS) in the United Kingdom (UK), and the Australian and New Zealand Standard Classification of Occupation (ANZSCO). The ANZSCO is referred to for defining the required education in a number of studies (Chiswick and Miller, 2006; Kler, 2007; Green,Kler and Leeves 2007). JA fails to account for the educational variations in jobs within occupations because of job aggregation, which is where the job analyst considers the same job title requiring the same educational requirement. The heterogeneity error is generated by aggregating error, where the heterogeneity within an occupation is ignored (Halaby, 1994). In addition, due to the large amount of expenditure required for updating new codes, existing codes may lack depth and be out of date, which will bias the criteria of the required qualification.

Self-Reported (SR) or Worker Self-Assessment (WA) is a subjective measure which evaluates over-education by asking the respondents the required educational level for their job. Because this method measures the required level of education based on the answers of workers, on the one hand, SR measure "has the advantage of drawing on all local, up-to-date information. The assessment deals, in principle, precisely with the respondent's job, not with any kind of aggregate". On the other hand, an SR measure could be biased due to classification error (Verhaest & Omey, 2006a), where workers might overstate job requirements or merely recite hiring practice standards (Hartog, 2000; Kler, 2005).

Realised Match (RM) includes Mean measure and Modal Education (Mode) measure. It is referred to as the empirical or the statistical measure of over-education. It was first introduced by Verdugo& Verdugo (1989) who defined that a worker is over-educated if his education is higher than one standard deviation above the average for his/her occupation (in the 1980 census occupation code). Conversely, a worker is under-educated if his education is lower than one standard deviation below the average for his 1980 census occupation code. The advantage of this measure is that the mean is derived directly from the existing data, so it is always available. However, this measure also has its drawbacks.

For example, RM only assesses frictional mismatches but fails to consider structural sources of over and under-education (Kiker, Santos, & De Oliveira, 1997; Verhaest & Omey, 2006b). Kiker, et al. (1997) noted concerns as this measure is more sensitive to technological change and changes in workplace organization than others. It is likely to be misinformed by the development of insufficient schooling over time. “one-standard deviation away from the mean” implies the symmetry between over-education and under-education, which is not rational. And the cut-off point is arbitrary. Moreover, as is similar to JA, the mean method ignores job variations within occupations (Halaby, 1994).

The Modal method (Mode) is the other Realised Match (RM) measure. It was proposed by Kiker, et al.(1997). Mode measure estimates the level of required education by computing the amount of education that most commonly occurs within an occupational category (Rubb, 2003). Mode measure proves more accurately than the mean method by considering the asymmetry between over-education and under-education and by being less sensitive to outliers or technological change. Kiker, Santos and Oliveira (1997) proved that Mode criterion is preferred to Verdugo& Verdugo’s mean criterion by using a very simple example. They found Verdugo& Verdugo’s mean criterion to be changing gradually and that it could produce classification errors before correcting itself but that the Mode changes more freely reflecting each period’s educational requirements of most workers at a given time.

In the initial stage of the study, the above measurements were evaluated, with the exception of Self-Reported (SR) and Worker Self-Assessment (WA) due to lack of related information in the HILDA data. The analyses provide support that the different methods are generally comparable, and that the mode is a reasonable measure to define the required years of education.

Based on this cross-wave Mode method, there is a very high incidence rate of over-education in Australia. Evidence can be found from Table 2 that migrants are more likely to be over-educated than natives. In addition, NESB immigrants are more vulnerable to over-education than their ESB counterparts. Among full-time workers aged 23 to 64, Table 2 shows that NESB immigrants have the highest rate of over-education, 42 per cent compared to 31 per cent for ESB and 25 per cent for natives. It reveals that mismatch is very serious among NESB immigrants. Almost half (42 per cent) of full-time NESB migrant workers are employed in positions that there are below their educational attainment levels.

Table 2: The Extent of Over-education by Country of Origin

	Native		ESB		NESB	
VARIABLES	mean	sd	mean	sd	mean	sd
Educational mismatch						
Over-educated	0.25	0.43	0.31	0.46	0.42	0.49
Under-educated	0.36	0.48	0.33	0.47	0.28	0.45
Matched	0.39	0.49	0.36	0.48	0.30	0.46
Years of over-education	0.58	1.33	0.70	1.38	1.07	1.70
Years of under-education	1.07	1.70	1.02	1.71	0.80	1.51
Years of required education	14.25	1.87	14.39	1.98	14.30	2.04

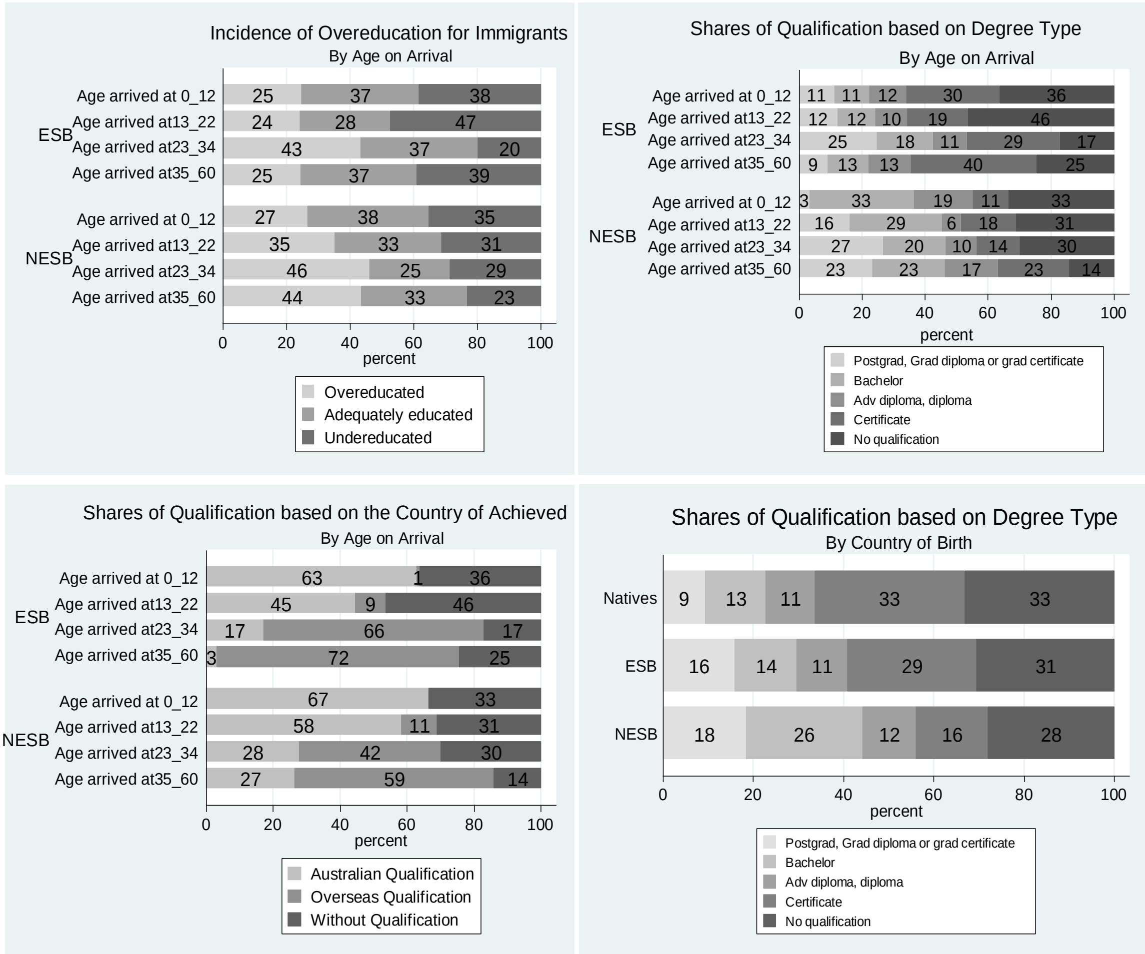
This evidence is consistent with Green, Kler and Leeves (2007) and Kler (2007). Both papers use the immigrant longitudinal data to Australia (LSIA). Green, Kler and Leeves (2007) applied Job Analysis (JA)[9] to measure the required education. They found that NESB immigrants are more likely to be over-educated, with the incidence of over-education being between 32% and 49%. Kler (2007) examined the effects of over-education among tertiary educated immigrants. The rate of over-education was found to be around 16% for immigrants from English Speaking Countries. Among Asian immigrants, approximately 50% are over-educated. Among other NESB immigrants, the rate of over-education is close to 40%.

Figure 1 further presents the incidence of over-education and the share of qualifications based on degree type and country of study by age at migration. Figure 1 shows insignificant effects on the incidence of over-education for ESB immigrants from migrating as a child or as an adult. However, significant impacts are found among NESB immigrants. Younger NESB immigrants who migrate to Australia between age 0 and age 12 are more likely to find a job which matches their level of education, with a 27 per cent incidence of over-education.

However, when individuals migrate at an older age, the incidence of over-education increases from 35 per cent (when migrating at age 13 to 22) to 44 per cent (when migrating at age 35 to 60). ESB immigrants who migrate at age 23 to 34 are a highly educated group relative to the other three age arrival cohorts, with 43 per cent of them achieving above Bachelor degree and only 17 per cent of them are without qualifications. Thus, they are expected to have better labour market performance outcomes than the other three groups.

[9] Job Analysis (JA) is a systematic evaluation by professional job analysts who specify the required level (and type) of education based on grading the occupation and deriving from information in regard to the occupations of respondents. It was originated to measure the required education by Eckhaus (1964).

Figure 1: Incidence of Over-education and Share of Qualifications based on Highest Qualification and Age on Arrival



5. Results

As noted earlier, to better use the features of panel data, a Correlated Random Effects logit[10] model is employed to examine determinants of over-education among natives and immigrants. We employ two samples separately for comparison purposes. The first sample contains ESB immigrants and the native-born, and the second consists of NESB immigrants and the native-born. Thus, we can determine specific effects for ESB and NESB immigrants respectively by comparing them with natives, as the common base. The dependent variable for the outcome equation is the odds ratio of being over-educated.

Our earnings model, in turn, examines over-education effects on earnings via years since migration. This model examines potential earnings penalties associated with over-education and it demonstrates the effects of over-education on immigrants' assimilation.

To control for potential unobserved heterogeneity effects on earnings, we employ a fixed effects model. We also report results for pooled OLS and random effects models.

[10] We also employed a random effects probit model to examine the determinants of over-education. The results are consistent with the results obtained from the random effects logit model.

5.1 Determinants of over-education

Based on Model 1 in Equation (6), the results of the estimations for natives and ESB immigrants and for natives and NESB immigrants are reported in Tables 3A and 3B, respectively.

As discussed previously, duration of residency and age on arrival may influence the rate of over-education. To examine these effects, based on Equation (6), two specifications are employed respectively. For each specification, we employ both random effects logit model and correlated random effects logit model. The first set of results reported for each specification is the base random effects logit model, and the second controls for Mundlak adjustment, as our preferred model.

The comparisons of these results reveal the endogeneity issue which is addressed by the Correlated Random Effects logit model with Mundlak correction (1978).

Marginal effects are reported. Marginal effects are derived as the coefficient multiplied by the density function (the probability of a positive outcome), evaluated at sample mean values of explanatory variables. In each table, columns (1) and (3) report marginal effects results from random effects logit model. Columns (2) and (4) report marginal effects results from correlated random effects logit model.[11]

Overall, immigrants are 28 to 54 per cent more likely to be over-educated than natives, in particular, a high incidence of over-education is found among NESB immigrants (54.2% in Column (1) of Table 3B). However, once the endogeneity issue is controlled by Mundlak correction, results from columns (3) and (4) of Tables 3A and 3B present that the propensity of over-education for immigrations is 88 to 94 per cent higher than for natives. This reveals immigrants have a serious education-occupation mismatch in Australia.

Workers who hold postgraduate are more likely to be over-educated than others. Individuals with a Bachelor degree or specific educational Certificate achieve better education-occupation matches than those with other types of qualifications. This result does not change after Mundlak correction. It is robust. Immigrants with diplomas reduce the probability of being over-educated by 7 per cent and NESB immigrants with certificates reduce the probability of being over-educated by 6 to 7 per cent compared to natives with the same qualifications. However, these effects become insignificant with Mundlak correction.

Years since migration, representing the duration of residency in Australia does help an NESB immigrant to achieve a better education-occupation match; [12] this is shown in the negative sign on YSM in Column (1) of Table 3B. Results show that there is a negative significant effect of linear YSM on the incidence of over-education, and this effect applies only to NESB immigrants. On the contrary, after accounting for the endogeneity due to the correlation between individual effects and error term, years since migration do not improve education-occupation mismatch for NESB immigrants. The coefficient of YSM is insignificant in Column (2) of Table 3B.

[11] We also applied a fixed effect logit for comparison. The fixed effects logit does not estimate the distribution of individual effects or the coefficients of time invariant variables. We have found very small coefficients from the fixed effects logit regression.

[12] The random effects (RE) logit results show that there is no significant effect of the quadratic YSM on the probability of over-education, thus this result is not reported but is available upon request.

Among NESB immigrants, Columns (3) and (4) in Table 3B shows that migrating as a child helps migrants to reduce the probability of being over-educated in employment. Immigrants who migrate at less than 12 years of age have 10 per cent lower probability of over-education rate in comparison to others who migrate between 34 to 60 years of age. These effects do not apply to ESB immigrants.

The evidence from random effects logit estimations is consistent with previous study. Years since migration and younger entries have a significant effect on reducing the probability of over-education among NESB immigrants. However, once we account for the endogeneity issue, immigrants have extremely higher incidence of over-education than natives. And years since migration do not help them to improve their education-occupation mismatch situation.

5.2 The Impact of over-education on Earnings

Pooled OLS analysis is based on the assumption of homogenous individuals and the random assignment of workers to jobs. Therefore, its result may be biased due to the unobserved heterogeneity of individuals and jobs. In contrast, longitudinal analysis allows the evaluation of unobserved heterogeneity on the earnings. In this section, we apply fixed effects models to address individual heterogeneity. We also report pooled OLS estimation as a benchmark to examine unobserved heterogeneity effects[13].

Following Model 2 in Equation (9), estimation results are given in Table 4. There are six columns for two specific subsamples. The first three columns report results from pooled OLS, fixed effects and random effects estimation for full-time ESB immigrants and natives. The last three columns present the results for full-time NESB immigrants and natives. The Hausman test rejects the null hypothesis that individual specific error is uncorrelated with the explanatory variables of the wage equation. Therefore, fixed effects estimates are preferred to random effects.

After accounting for individual effect, fixed effects estimations reveal that years since migration (YSM) have a stronger effect on earnings for ESB immigrants than for NESB immigrants. That is, an ESB immigrant improves his earnings by 2.4 per cent for each year of staying in Australia, which is 1 per cent higher than for NESB immigrants (1.4 per cent). Longitudinal estimations suggest a much stronger effect on assimilation for ESB immigrants than for NESB immigrants.

Compared to over-educated natives with the same characteristics, over-educated ESB immigrants seemingly have similar returns to years of over-education. This effect is shown by the insignificant effects on interaction terms between years of over-education and immigrant status. In contrast, according to the panel fixed effects estimation in column (5), NESB immigrants suffer a 9 per cent lower return for the additional years of over-education than comparable natives. This suggests that educational mismatch is a serious problem among NESB immigrants, and that it can explain the earnings penalty from education-occupation mismatch. Similar effects are also found in the returns to years of required education, which is shown in Column (5), for each year of required education, as NESB immigrants have a 9 per cent lower return than natives. This indicates that NESB immigrants suffer earnings penalties not only from education-occupation mismatch but also

[13] Comparing cross-section and panel evidence on mismatch wage penalties, using first eight waves of HILDA, Mavromaras et al (2012) found that cross-section estimates were considerately higher, which indicated the presence of unobserved heterogeneity in the data.

when they possess adequate years of education.

Table 3A: Determinants of Over-education among Natives and ESB Immigrants
Correlated Random Effects Logit Estimations
(Model 1)

Dependent variable =1 if workers are observed to be over-educated Sample: Natives (N) and ESB Immigrants					
	(1) RE Logit	(2) CRE Logit	(3 RE Logit)	(4) CRE Logit	Pr(over-education|u_i=0) =11.8%
Explanatory Variables	Marginal Effects	Marginal Effects	Marginal Effects	Marginal Effects	Mean of X
Immigrant (M)	0.379**	0.924***	0.277*	0.937***	0.143
Human Capital	(0.156)	(0.079)	(0.154)	(0.059)	
Years of education	0.078***	-0.029	0.078***	-0.027	13.720
	(0.010)	(0.584)	(0.010)	(0.597)	
Postgraduate	0.477***	0.223	0.480***	0.229	0.115
	(0.145)	(0.362)	(0.145)	(0.365)	
Bachelor	-0.120***	-0.116*	-0.120***	-0.115*	0.146
	(0.022)	(0.064)	(0.022)	(0.064)	
Diploma	0.133*	-0.064	0.135*	-0.063	0.102
	(0.074)	(0.076)	(0.075)	(0.077)	
Certificate	-0.179***	-0.215***	-0.179***	-0.214***	0.320
	(0.024)	(0.069)	(0.024)	(0.069)	
Postgraduate × M	-0.039	-0.025	-0.041	-0.035	0.024
	(0.039)	(0.147)	(0.039)	(0.132)	
Bachelor × M	-0.010	0.090	-0.011	0.076	0.022
	(0.044)	(0.275)	(0.044)	(0.259)	
Diploma × M	-0.072***	-0.081	-0.074***	-0.084*	0.016
	(0.023)	(0.050)	(0.022)	(0.047)	
Certificate × M	-0.012	0.087	-0.016	0.078	0.040
	(0.040)	(0.235)	(0.039)	(0.226)	
EXP	-0.005***	-0.149	-0.005***	-0.147	22.264
	(0.002)	(0.585)	(0.002)	(0.599)	
EXPSQR	0.012***	0.010	0.012***	0.010	6.106
	(0.004)	(0.007)	(0.004)	(0.007)	
Disability or impairment	0.005	-0.002	0.005	-0.002	0.148
	(0.012)	(0.012)	(0.012)	(0.012)	
Years since Migration-YSM	-0.002	-0.003	/	/	26.320
	(0.001)	(0.005)	/	/	
Age on Arrival					
Age 0-12	/	/	-0.010	-0.032	0.399
	/	/	(0.050)	(0.043)	
Age 13-22	/	/	-0.036	-0.052	0.166
	/	/	(0.045)	(0.038)	
Age 23-34	/	/	0.013	-0.001	0.340
	/	/	(0.059)	(0.055)	
Control for States	YES	YES	YES	YES	
Control for unemployment	YES	YES	YES	YES	
Control for time periods	YES	YES	YES	YES	
Mundlak Correction	NO	YES	NO	YES	
Observations	14,711	14,711	14711	14,711	
Individuals	2,313	2,313	2,313	2,313	
Log likelihood	-4536	-4504	-4536	-4504	
Wald chi-squared	1304	1291	1305	1291	
rho	0.530	0.529	0.530	0.529	
Likelihood-ratio test of rho=0:	1468	1456	1454	1441	
significance	0	0	0	0	

Notes:
Dependent variable in outcome equation is the probability of over-education in full-time job. Constant is included.
Standard errors in parentheses; *10 per cent level of significance, **5 per cent level of significance, ***1 per cent level of significance;
Base-categories are Natives, no qualification, Age 35-60, Year 2009, and QLD.
The models include qualifications dummy variables, States dummy variables (NSW, VIC, SA, WA, TAS, NT, ACT), Unemployment, Unemployment × M, time periods dummy variables. Full Results are available upon request.
Sample: Natives and English Speaking Background (ESB) immigrants.

Sources: HILDA-Release 9 (Wave 1-Wave 9).

Table 3B: Determinants of Over-education among Natives and NESB Immigrants
Correlated Random Effects Logit Estimations
(Model 1)

Dependent variable =1 if workers are observed to be over-educated Sample: Natives (N) and NESB Immigrants					
	(1) RE Logit	(2) CRE Logit	(3) RE Logit	(4) CRE Logit	Pr(over-education|u_i=0) =11.9%
Explanatory Variables	Marginal Effects	Marginal Effects	Marginal Effects	Marginal Effects	Mean of X
Immigrant (M)	0.542***	0.879***	0.260	0.937***	0.093
Human Capital	(0.193)	(0.167)	(0.199)	(0.030)	
Years of education	0.082***	0.117***	0.081***	0.116***	13.746
	(0.011)	(0.031)	(0.011)	(0.031)	
Postgraduate	0.468***	0.258	0.474***	0.261	0.114
	(0.156)	(0.395)	(0.155)	(0.396)	
Bachelor	-0.127***	-0.113	-0.127***	-0.113	0.154
	(0.023)	(0.071)	(0.023)	(0.071)	
Diploma	0.130	-0.058	0.132*	-0.058	0.102
	(0.079)	(0.085)	(0.079)	(0.085)	
Certificate	-0.184***	-0.211***	-0.183***	-0.211***	0.312
	(0.026)	(0.070)	(0.026)	(0.070)	
Postgraduate × M	0.013	0.690	0.039	0.843**	0.018
	(0.088)	(1.141)	(0.103)	(0.352)	
Bachelor × M	-0.003	0.278	0.049	0.642	0.023
	(0.057)	(1.666)	(0.076)	(1.393)	
Diploma × M	-0.067**	-0.076	-0.060	-0.065	0.012
	(0.034)	(0.180)	(0.038)	(0.241)	
Certificate × M	-0.069**	-0.001	-0.061*	-0.003	0.015
	(0.033)	(0.143)	(0.036)	(0.142)	
EXP	-0.005***	-0.006**	-0.006***	-0.007**	21.924
	(0.002)	(0.003)	(0.002)	(0.003)	
EXPSQR	0.013***	0.016**	0.012***	0.015**	5.955
	(0.004)	(0.008)	(0.004)	(0.008)	
Disability or impairment	0.006	-0.002	0.006	-0.002	0.143
	(0.013)	(0.013)	(0.013)	(0.013)	
Poor English	-0.050	-0.046	-0.048	-0.056	0.04
	(0.055)	(0.078)	(0.058)	(0.070)	
Years since Migration-YSM	-0.007***	0.000	/	/	20.650
	(0.002)	(0.006)	/	/	
Age on Arrival					
Age 0-12	/	/	-0.095***	-0.098***	0.220
	/	/	(0.020)	(0.018)	
Age 13-22	/	/	-0.051	-0.065*	0.235
	/	/	(0.045)	(0.037)	
Age 23-34	/	/	-0.043	-0.046	0.379
	/	/	(0.047)	(0.045)	
Control for States	YES	YES	YES	YES	
Control for unemployment	YES	YES	YES	YES	
Control for time periods	YES	YES	YES	YES	
Mundlak Correction	NO	YES	NO	YES	
Observations	13808	13808	13808	13,808	
Individuals	2,185	2,185	2,185	2,185	
Log likelihood	-4235	-4208	-4238	-4211	
Wald chi-squared	1178	1181	1172	1179	
rho	0.553	0.552	0.555	0.554	
Likelihood-ratio test of rho=0:	1492	1483	1496	1489	
significance	0	0	0	0	

Notes:
Dependent variable in outcome equation is the probability of over-education in full-time job. Constant is included.
Standard errors in parentheses; *10 per cent level of significance, **5 per cent level of significance, ***1 per cent level of significance;
Base-categories are Natives, no qualification, Age 35-60, Year 2009, and QLD.
The models include qualifications dummy variables, States dummy variables (NSW, VIC, SA, WA, TAS, NT, ACT), Unemployment, Unemployment × M, time periods dummy variables. Full Results are available upon request.
Sample: Natives and Non-English Speaking Background (NESB) immigrants.

Sources: HILDA-Release 9 (Wave 1-Wave 9).

Table 4: The Effect of Over-education on Earnings for Natives and Immigrants
(Model 2)

Dependent Variable : The natural logarithm of hourly wage from main job in 2009 dollars

	Natives (N) and ESB Immigrants			Natives (N) and NESB Immigrants		
	(1)	(2)	(3)	(4)	(5)	(6)
	N&ESB	N&ESB	N&ESB	N&NESB	N&NESB	N&NESB
Explanatory Variables	Pooled OLS	Panel-FE	Panel-RE	Pooled OLS	Panel-FE	Panel-RE
Immigrant (M)	0.879***	/	0.346	0.070	/	0.319
	[0.208]	/	[0.290]	[0.296]	/	[0.350]
Human capital						
Years of over-education	0.091***	0.049***	0.080***	0.088***	0.048***	0.077***
	[0.007]	[0.016]	[0.011]	[0.007]	[0.015]	[0.011]
Years of under-education	-0.063***	-0.036**	-0.066***	-0.060***	-0.036**	-0.063***
	[0.006]	[0.015]	[0.010]	[0.006]	[0.015]	[0.010]
Years of required education	0.117***	0.045***	0.080***	0.114***	0.045***	0.077***
	[0.006]	[0.015]	[0.010]	[0.006]	[0.015]	[0.010]
Years of over-education × M	-0.081***	-0.020	-0.061**	-0.030	-0.089**	-0.073**
	[0.019]	[0.035]	[0.025]	[0.025]	[0.040]	[0.029]
Years of under-education × M	0.105***	-0.009	0.038	0.037*	0.070	0.049
	[0.016]	[0.037]	[0.025]	[0.021]	[0.045]	[0.030]
Years of required education × M	-0.075***	-0.004	-0.046*	-0.033	-0.087**	-0.070**
	[0.016]	[0.035]	[0.024]	[0.023]	[0.041]	[0.029]
Postgraduate × M	0.481***	0.036	0.286*	0.238	/	0.443**
	[0.107]	[0.238]	[0.168]	[0.146]	/	[0.203]
Bachelor × M	0.413***	0.165	0.325**	-0.008	-0.024	0.337**
	[0.085]	[0.203]	[0.141]	[0.112]	[0.100]	[0.165]
Diploma × M	0.450***	-0.035	0.210*	0.041	0.305	0.170
	[0.073]	[0.166]	[0.119]	[0.092]	[0.337]	[0.154]
Certificate ×M	0.185***	-0.053	0.051	-0.021	-0.018	0.045
	[0.057]	[0.155]	[0.096]	[0.080]	[0.153]	[0.111]
EXP	0.023***	0.040***	0.030***	0.024***	0.039***	0.028***
	[0.001]	[0.003]	[0.002]	[0.002]	[0.003]	[0.002]
EXPSQR/100	-0.042***	-0.052***	-0.051***	-0.046***	-0.050***	-0.050***
	[0.003]	[0.005]	[0.004]	[0.003]	[0.005]	[0.004]
Years since migration-YSM						
YSM	0.003	0.024***	0.015***	0.023***	0.014*	0.021***
	[0.004]	[0.007]	[0.005]	[0.005]	[0.008]	[0.006]
YSMSQR/100	-0.008	-0.035***	-0.026***	-0.031***	-0.008	-0.020*
	[0.008]	[0.011]	[0.009]	[0.010]	[0.015]	[0.012]
Over-educated ×YSM	0.003	0.004	0.004	-0.010**	0.004	0.003
	[0.004]	[0.004]	[0.003]	[0.005]	[0.004]	[0.004]
Over-educated ×YSMSQR/100	-0.011	-0.010	-0.009	0.026**	-0.015	-0.011
	[0.010]	[0.009]	[0.008]	[0.012]	[0.011]	[0.011]
Under-educated ×YSM	-0.007*	0.004	0.002	-0.007	0.000	0.000
	[0.004]	[0.004]	[0.003]	[0.006]	[0.005]	[0.005]
Under-educated ×YSMSQR/100	0.006	-0.008	-0.004	0.011	-0.002	-0.002
	[0.008]	[0.008]	[0.008]	[0.012]	[0.013]	[0.012]
Disability impairment	-0.077***	-0.008	-0.016*	-0.073***	-0.007	-0.014*
	[0.011]	[0.008]	[0.008]	[0.012]	[0.009]	[0.009]
Poor English	/	/	/	-0.190***	0.052	-0.008
	/	/	/	[0.066]	[0.078]	[0.072]
Control for States	YES	YES	YES	YES	YES	YES
Control for unemployment	YES	YES	YES	YES	YES	YES
Constant	1.536***	1.997***	1.876***	1.570***	2.124***	1.935***
	[0.082]	[0.165]	[0.123]	[0.081]	[0.169]	[0.122]
F-test	93.74	18.47	/	84.63	16.75	/
R2	0.174	0.0290	0.149	0.173	0.00832	0.148
Individuals	2313	2313	2313	2185	2185	2185
Observations	14711	14711	14711	13808	13808	13808
R2_within	/	0.0456	0.0364	/	0.0442	0.0344
rho	/	0.805	0.733	/	0.833	0.730
Wald-test chi2	/	/	0	/	/	0
Hausman fe re test: Chi^2	/	324.6		/	347.8	
Prob>chi2=	/	0		/	0	

Notes: The Hausman test rejects random effects results and accepts the fixed effects result. The result is consistent. *10 per cent level of significance, **5 per cent level of significance, ***1 per cent level of significance; Standard errors in brackets. Based-categories are natives, no qualification, being matched × YSM, being matched YSM SQR/100, QLD. The models include qualifications dummy variables, States dummy variables (NSW, VIC, SA, WA, TAS, NT, ACT), Unemployment, Unemployment × M. Results are available upon request.
Sources: HILDA-Release 9 (Wave 1-Wave 9).

6. Conclusion

Based on recent Australian immigration policy, flows of skilled immigrants to Australia are increasing, such that endorsed skills would help immigrants to become more employable and thereby increase Australian productive capacity. However, if skilled immigrants disproportionately work at jobs that under-utilise their educational attainment, do they still contribute to the host country's economic development, or do they become a burden to the local economy?

This paper has provided evidence on the above question. Based on nine years of HILDA and longitudinal analyses, results show that NESB immigrants have a significantly higher incidence of over-education and that they receive a large earnings penalty from over-education. Using over-education as an indicator in explaining immigrant assimilation, our results are summarised as below:

Firstly, 42 per cent of NESB immigrants have been found to work in jobs which require a lower educational standard than the one they possess. The determinants of over-education are examined by a correlated random effects logit model with Mundlak correction. After accounting for endogeneity, immigrants demonstrate extremely higher rates of over-education than the native-born. As time passes, the education-occupation mismatch situation for immigrants does not change with increased years since migration. Among NESB immigrants, younger entrants (who have migrated at younger than the age of 12) are more likely to reduce the probability of over-education than are older entrants.

Secondly, the impact of education-occupation mismatch on earnings is examined by Longitudinal (panel fixed effects) Analysis. The results reveal that, in general, ESB immigrants earn more, and NESB immigrants earn less than natives. After controlling for unobserved heterogeneity (such as motivation, ability, etc.), years since migration is shown to have a significant impact on earnings for both ESB and NESB immigrants. ESB immigrants have a faster earnings growth rate than natives.

However, educational mismatches worsen the NESB earnings outcomes. With panel fixed effects estimation, NESB immigrants are shown to suffer a 9 per cent lower return to each additional year of over-education and a 9 per cent lower return to required years of education than natives. This evidence suggests that the earnings penalty among NESB immigrants is due, not only to skill under-utilisation, but perhaps also to an earnings disadvantage that cannot be accounted for by the extensive human capital variables included in our models. There is a persistent earnings gap between natives and NESB immigrants, even when NESB immigrants who migrate beyond age 12.

These findings have implications for Australian immigration assimilation policies, which focus, not only on attracting skilled immigrants, but also on the likelihood and facilitation of employment into matched positions.

References

Aringa, Carlo. Dell., & Pagani, Laura. (2010). *Labour Market Assimilation and Over Education: The Case of Immigrant Workers in Italy*. No.178. Universita degli Studi di Milano-Bicocca. Dipartimento di Economia Politica.

Chamberlain, Gary. (2010). Binary Response Models for Panel Data: Identification and Information. *Econometrica, 78*(1), 159-168.

Chiswick, B.R., & Miller, P.W. (2006). *An Explanation for the Lower Payoff to Schooling for the foreign born in Australia*. Paper presented at the 35th Australian Conference of Economists, Curtin University of Technology, Perth.

Chiswick, Barry R., & Miller, Paul W. (2008). Why is the Payoff to Schooling Smaller for Immigrants? *Labour Economics*.

Chiswick, Barry R., & Miller, Paul W. (2009). The International Transferability of Immigrants' Human Capital. *Economics of Education Review, 28*(2), 162-169.

Cohn, Elchanan. (1992). The Impact of Surplus Schooling on Earnings. *Journal of Human Resources, 27*(4), 679-682.

DIMA. (2007). Department of Immigration and Multicultural Affairs. *Population Flows: Immigration Aspects*. Canberra: AGPS.

Duncan, Greg J., & Hoffman, Saul D. (1981). The Incidence and Wage Effects of Overeducation. *Economics of Education Review, 1*(1), 75-86.

Fernández, Cristina, & Ortega, Carolina. (2008). Labor Market Assimilation of Immigrants in Spain: employment at the expense of bad job-matches? *Spanish Economic Review, 10*(2), 83-107.

Ferrer, Ana, & Riddell, W. Craig. (2008). Education, Credentials, and Immigrant Earnings. *Canadian Journal of Economics, 41*(1), 186-216.

Friedberg, Rachel M. (2000). You Can't Take It with You? Immigrant Assimilation and the Portability of Human Capital. *Journal of Labor Economics, 18*(2), 221-251.

Green, Colin, Kler, Parvinder, & Leeves, Gareth. (2007). Immigrant Overeducation: Evidence from Recent Arrivals to Australia. *Economics of Education Review, 26*(4), 420-432.

Greene, William. (2010). Maximum Simulated Likelihood Methods and Applications: Introduction. In W. Greene & R. C. Hill (Eds.), *Maximum Simulated Likelihood Methods and Applications* (pp. ix-xiv): Advances in Econometrics, vol. 26. Bingley, U.K.: Emerald.

Groot, Wim. (1996). The Incidence of, and Returns to Overeducation in the UK. *Applied Economics, 28*(10), 1345-1350.

Halaby, Charles N. (1994). Overeducation and Skill Mismatch. *Sociology of Education, 67*(1), 47-59.

Hartog, Joop. (2000). Over-education and Earnings: Where are We, Where should We Go? *Economics of Education Review, 19*(2), 131-147.

Hayes, C, & Watson, N. (2009). HILDA Imputation Methods. *HILDA Project Technical Paper Series*(2/09).

Kiker, B. F., Santos, M. C., & De Oliveira, M. M. (1997). Overeducation and Undereducation: Evidence for Portugal. *Economics of Education Review, 16*(2), 111-125.

Kler, Parvinder. (2005). Graduate Overeducation in Australia: A Comparison of the Mean and Objective Methods. *Education Economics, 13*(1), 47-72.

Lindley, Joanne. (2009). The Over-education of UK Immigrants and Minority Ethnic Groups: Evidence from the Labour Force Survey. *Economics of Education Review, 28*(1), 80-89.

Mavromaras, K., Sloanr, P., & Zheng, W. (2012). the Role of Education Pathways in the Relationship between Job Mismatch, Wages and Job Satisfaction: a Panel Estimation approach. *Education Economics*, 20(3), 303-321.

Mundlak, Yair. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica, 46*(1), 69-85.

Rubb, Stephen. (2003). Post-College Schooling, Overeducation, and Hourly Earnings in the United States. *Education Economics, 11*(1), 53.

Rumberger, Russell W. (1987). The Impact of Surplus Schooling on Productivity and Earnings. *Journal of Human Resources, 22*(1), 24-50.

Sanroma, Esteban, Ramos, Raul, & Simon, Hipolito. (2008). The Portability of Human Capital and Immigrant Assimilation: Evidence for Spain. *SSRN eLibrary*.

Sicherman, Nachum. (1991). "Overeducation" in the Labor Market. *Journal of Labor Economics, 9*(2), 101.

Verdugo, Richard R., & Verdugo, Naomi Turner. (1989). The Impact of Surplus Schooling on Earnings. *Journal of Human Resources, 24*(4), 629-643.

Verhaest, Dieter, & Omey, Eddy. (2006a). Discriminating between Alternative Measures of Over-education. *Applied Economics, 38*(18), 2113-2120.

Verhaest, Dieter, & Omey, Eddy. (2006b). Measuring the Incidence of Over- and Undereducation. *Quality & Quantity, 40*(5), 783-803.

Wald, Steven, & Fang, Tony. (2008). Overeducated Immigrants in the Canadian Labour Market: Evidence from the Workplace and Employee Survey. *Canadian Public Policy, 34*(4), 457-479.

Wilkins, Roger. (2003). Immigrant Earnings Adjustment: The Impact of Age at Migration. *Australian Economic Papers, 42*(3), 292-315.

Wooldridge, Jeffrey (2009). *Introductory Econometrics*. 4^{th} Edition, Mason OH: South-Western.

Appendix: Definition of Variables

<u>Personal Characteristics</u>	
age	Continuous age variable, expressed in years
Disability or impairment	Dummy variable, 1 if has Long term health condition, disability or impairment, zero otherwise
Poor English	Dummy variable, 1 if immigrant speaks English poorly, zero otherwise
Age on Arrival	
Age 0-12	Dummy variable, 1 if immigrant migrated at ages between 0 and 12, zero otherwise
Age 13-22	Dummy variable, 1 if immigrant migrated at ages between 13 and 22, zero otherwise
Age 23-34	Dummy variable, 1 if immigrant migrated at ages between 23 and 34, zero otherwise
Age 35-60	Dummy variable, 1 if immigrant migrated at ages between 35 and 60, zero otherwise
Years since Migration-YSM	Continuous variable, years of duration in Australia for immigrants
Country of birth	
Natives	Dummy variable,1 if born in Australia, zero otherwise
Immigrant	Dummy variable,1 if born overseas, zero otherwise
ESB immigrant	Dummy variable,1 if born in an English speaking country, zero otherwise
NESB immigrant	Dummy variable,1 if born in an non-English speaking country, zero otherwise
Region	
NSW	Dummy variable, 1 if living in NSW, zero otherwise
VIC	Dummy variable, 1 if living in VIC, zero otherwise
QLD	Dummy variable, 1 if living in QLD, zero otherwise
SA	Dummy variable, 1 if living in SA, zero otherwise
WA	Dummy variable, 1 if living in WA, zero otherwise
TAS	Dummy variable, 1 if living in TAS, zero otherwise
NT	Dummy variable, 1 if living in NT, zero otherwise
ACT	Dummy variable, 1 if living in ACT, zero otherwise
<u>Job Characteristics</u>	
Jbmo62	jbmo62 provides 2-digit Australian and New Zealand Standard Classification of Occupations (ANZSCO2006) occupations category
Unemployment	Unemployment rate annually, refer to 6202.0 - Labour Force, Australia, Australian Bureau of Statistics
Hourly Wage	Continuous variable, current weekly gross wages and salary from main job divided by combined hours per week usually worked in main job in 2009 dollars
Log Hourly Wage	Continuous variable, the natural logarithm of Hourly Wage from main job
<u>Human Capital</u>	
Years of experience (total)-EXP	Continuous variable, potential years of work experience (Age – years of education - 6)
Years of actual education (total)-ED	Continuous educational attainment variable, expressed in years
Highest Qualification	
Postgraduate	Dummy variable,1 if highest qualification is doctorate, masters, grad diploma, grad certificate or bachelor with honours, zero otherwise
Bachelor	Dummy variable, 1 if highest qualification is bachelor without honours, zero otherwise
Diploma	Dummy variable, 1 if highest qualification is advanced diploma or diploma, zero otherwise
Certificate	Dummy variable, 1 if highest qualification is certificate I II III or IV, zero otherwise
Educational Mismatched <u>based on cross -wave Mode measure</u>	
Over-educated	Dummy variable, takes the value 1 if over-educated, zero otherwise
Under-educated	Dummy variable, takes the value 1 if under-educated, zero otherwise
Matched	Dummy variable, takes the value 1 if adequately educated, zero otherwise
Years of over-education	Continuous variable, the years of over-education
Years of under-education	Continuous variable, the years of under-education
Years of required education	Continuous variable, the years of adequate education