Household Structure and Overeducation: Wage Study

By: Cameron Holmes

An essay submitted to the Department of Economics

In partial fulfillment of the requirements for

The degree of Master of Arts

Queen's University

Kingston, Ontario, Canada

August 2021

Abstract:

This paper examines the effects that martial status, having children in the home, and having young children has on a person's likelihood of being over or under educated for their job. We classify a person overeducated if for example the mean level of education for their occupation is a bachelor's degree, but they have a PhD then they are overeducated. The opposite is what undereducation is classified as, say a person with a bachelor's degree is working at a job where the mean level of education is a PhD, then that person would be classified as undereducated. Then taking these new over and undereducation variables, they are put into another model to determine the effect they may have on a person's income. We find that single people are more likely to be overeducated and single people with children are more likely to be both overeducated and undereducated. As for wage effects, being overeducated has a much stronger negative effect on wage than undereducation but result in a loss of income compared to a well-matched person.

Introduction:

Education mismatch can affect anyone no matter where you live, or what industry you work in, it is shown to negatively effect a worker's wages, reduce job satisfaction, and raise a worker's intent to leave their current job (Duncan & Hoffman, 1981: Leuven & Oosterbeek, 2011). These are important reasons to study the factors that could lead to educational mismatch so that workers and firms can be more informed about what may cause a person to be mismatched, and to help those who are most likely to end up in a bad fitting job. Mismatch can be in two directions, overeducated for your current job and undereducated for your current job. This paper will break down the results into these two categories, this paper will also examine the wage effects of being over and under educated in the results section. In other papers the effects on job satisfaction are examined more closely see Allen & van der Velden (2001).

In this paper I will examine the factors that may be associated with the likelihood for a person to find themselves mismatched in the labour force. This paper has a focus on marital status, the effect of having children in the household, and the final question this paper will look at is what industries are the most susceptible to education mismatch. The second model depicted in this paper will look at the income effects of over and undereducation, taking the information learned from the first model about the factors as well as showing the negative income effects of being mismatched. The seminal paper for this area of research is "The Incidence and Wage Effects of Overeducation" by Duncan and Hoffman (1981). This paper also uses data from the United States of America, but its data comes from 1976, and the focus of their research is on race and gender. Focusing on four groups of people, black males, white males, black females, and white females. Racial segregation had only been officially banned in the USA in the Civil Rights Act of 1964, racial discrimination was a widely researched and focused topic. This paper will look at those demographics as well but with more recent data from 2014-2017.

This paper will add to the existing literature by using more recent data as stated above and with a deeper focus on household variables such as marital status, the number of children in the home and the age of these children. These household structure variables have been studied in the past but only for how they effect a person's labour market decisions. These household structure variables as well as an industry break down to see which on the industries are the most susceptible to education mismatch.

The literature review will cover some different ways that mismatch has been measured in previous literature and will set this study up by comparing methods to some previous papers. Then moving onto collecting the data, and how we defined some of the key variables, such as marriage status, and number of children. Methodology will explain the key model we use for the results section as well as some robustness checks that were performed. Then the results are summarized and presented in table 2, and finally the conclusion wraps up the findings and closes with some future research that could be done.

Literature Review:

Mismatch has previously been measured in three different ways, self-assessment, job analysis and realized matches. Self-assessment is measured by asking people if they believe they are mismatched or not. This can be done with either direct or indirect questions(Verhaest & Omey, 2006). Duncan & Hoffman (1981) used this method in which respondents to the Panel Study of Income Dynamics responded to the question, "How much formal education is required to get a job like yours?", then comparing the answers reported to this question with individuals' education level they determine the level of a person's mismatch. Allen & van der Velden (2001) also use the self-assessment method, one key difference between these two papers is that Allen & van der Velden (2001) also consider educational field differences as well as educational level mismatch, respondents were asked both questions about educational level as well as field of study, and if they believed their level and/or field of study were a good fit for their occupation. Self-assessment brings to light the issue of subjectivity in how people believe if they are mismatched or not. (Hartog Joop, 2000) To combat this subjectivity, Turmo-Garuz, Bartual-Figueras, & Sierra-Martinez (2019) used the self-assessment method but instead of only looking at one time period conducted an analysis on the same group of people over 6 years. This longitudinal study allowed them to reduce the risk of subjectivity by always relating back to the first year the survey was conducted, comparing results from the two future surveys with the base one. The second measure of mismatch is called the Job analysis method. This method compares the educational level of a person with a recommended level set by some job analysis. One common point of reference is the Dictionary of Occupational Titles (DoT).

The third method and the one I will employ in my study, is the realized matches method. This method first calculates the mean or mode educational level for a certain occupation and then compares a person's educational level with the mean or mode level for their occupation. The typical measure when comparing mean education levels, is that anyone more than one standard deviation away from the mean educational level is considered mismatched. (Verhaest & Omey, 2006) The realized matches method can also be used with educational field as the determining factor if a person is mismatched or not, but for the purpose of this study I will be looking at solely education level.

Leuven & Oosterbeek (2011) also looked at a person's gender as a factor in their likelihood of being mismatched. They find that women have a higher incidence of being overeducated for a position than their male counterparts. One counter to this study is offered by Rubb (2014) when he looks at data from the 2000 United States census, finding that males are more likely to be overeducated than women. Groot & Maassen van der Brink (2000) in their meta-analysis, find the same result, that women are more likely to be over educated but males are more likely to be undereducated. Boudarbat & Chernoff (2010) in their study of recent Canadian university graduates that gender has no significant effect on the likelihood of finding a good match. They also state that marital status has no significant effect on a person being mismatched or not. Rubb (2014) finds that marital status does affect the likelihood of mismatch. People who are married are more likely to be overeducated and less likely to be undereducated.

Other studies refer to the influence that gender has on the likelihood of being mismatched but choose to exclude it from their analysis, so they can get clearer results. Bowlus (1995) is looking at the quality of someone’s match between their education and job. Using job tenure to determine mismatch they state that;

“Females are excluded because of the nonnegligible probability of their spells terminated for reasons other than poor quality such as pregnancy, marriage, and moving due to spousal job relocation.”

The purpose of my study is focused on household structure and the effects of being married and/or having children, have on your likelihood of being over or under educated, so it is important to include both men and women to capture these effects correctly.

Different studies use data from different countries with different government policies, different cultures and different work to job lifestyles. Duncan & Hoffman (1981) took their sample from the Panel Study of Income Dynamics in the United States of America. They find that 42% of the American work reported they were overeducated for their job while 12% reported they were undereducated, resulting in a 54% rate of educational mismatch. I will also look at American data, I will update these findings, using the realized matches method and with the additional household structure variables. Another study by Summerfield & Theodossiou (2017) examined German data, taking interest in what the effects are on job mismatch when you graduate during an economic recession. They determine mismatch in two ways, comparing a person’s education level with the median level in either their occupation or industry. They find using the occupational median that 24% of people are mismatched and 35% are mismatched when comparing to the industry median education level. This study employs a version of the realized matches method of measuring job mismatch, the same method I will use but with a spin, comparing the median level instead of the mode level of education. They also compare on two different scales, industry and occupational levels. I will use occupational level in my study.

Yuen (2010) conducted a study using the self-assessment method this time with Canadian subjects, and asked respondents to select whether they found that their education was closely related, somewhat related or not at all related to their current job. They broke the results down into three categories, people with greater than a bachelor's degree, people with a bachelor's or below and people with non-university postsecondary. Of these three groups she found that they responded that their education was closely related or somewhat related to their job, 68%, 60% and 54% of the time. Meaning that people with a university degree above a bachelor's degree are the lease likely to be mismatched in their workplace. Whereas those without a university degree are the most likely to be mismatched. In Allen and van der Velden's (2006) paper they looked at two different levels of education, university and higher vocational education. They found that 50% of university graduates and 56% of vocational education graduates considered a level/field of education other than their own appropriate. Providing the same result as Yuen (2010). Using Canadian data Boudarbat & Chernoff (2010) examined a group of recent university graduates with either a bachelor, masters or doctorate degree. They used data collected by Statistics Canada in the Follow-up of Graduates Survey (2000). They found that education level played a significant role in the likelihood of mismatch. Setting the likelihood of finding a good education-job match for a bachelor's degree equal to 1, they found that with a master's degree the probability goes up to 1.325 and a doctorate degree increases the likelihood of a good education-job match to 2.539. Providing further support for the theory that higher education attainment leads to increase probability of finding a good job match.

Duncan & Hoffmen (1984) examined every education level from having only completed grade 5 up to having an advanced degree and respondents were aged from 18 to 64. Finding that people with advanced degrees were the second highest group reporting that they had more education than what was needed for their current job, 59.6%. The only higher group was the 13-15 years of completed education, reporting in at 66.7% mismatched. The difference in the mismatch figures between Duncan & Hoffmen (1984) and the two more recent studies, Allen & van der Velden (2006) and Yuen (2010) might be explained by the change in how society views education. Even in recent years with the way technology is advancing and with societies changing views, from 2001 to 2006 the percentage of people with some sort of postsecondary education, be it a certificate, diploma or degree increased from 53% to 61% in Canada. The demand for higher educated people is on the rise, in 2006 Statistics Canada said that only 16% of jobs required a university degree or equalivant. That is an increase of 14% from 2000 (Yuen, 2010). Groot & van den Brink (2000) also noted this changing education attainment level.

"One of the most remarkable social developments in past decades in all Western countries has been the increase in educational level of the population"

They found that in 1992 about 38% of the population of OECD countries aged between 55 and 64 had obtained at least upper secondary education. For the population aged 25 to 34 years, 65% had at least upper secondary education. This is a good example of why this study must be repeated often as the population attains higher levels of education.

The likelihood of mismatch also varies across different occupational fields, this may be linked to the different types of education that each field requires. Duncan & Hoffman (1981) found that being overeducated was most prevalent among blue collar workers (semi/unskilled). Between 55-60% of this group of workers identified as being overeducated for the position they were working in. The least likely group to be mismatched was professionals, such as managers and clerical workers with less than 33% of them identifying as mismatched. Audra Bowlus (1995) finds in her study of the United States of America finds that those most effected by mismatch are those in professional jobs, financial services, Insurance and real estate. She operationalizes mismatch in a different way though, looking at skills matchups instead of education levels. Ortiz & Kucel (2008) doing a study in Germany and Spain find that in Germany the highest sectors for mismatch were the services industry and agriculture, while the lowest was the health and welfare sector. These results are like those of Duncan & Hoffman (1981), but still draws to the assumption that the underlying factor isn't the field of work, but the education required for the field. Cappelli (2015) in his discussion paper believes that the problem of mismatched workers is rooted in the education system and not in these underlying factors such as education level and ovulational field but rather in what is taught in schools and how people go about the selection process of what to study. Showing that more research is needed to determine if the field of work is really a factor or just another way that the level of education required is affecting the likelihood of mismatch.

Data/Methodology:

The data for this study comes from the United States American Community Survey from the years 2014 to 2017 accessed from the IPUMS data base, this survey is a 1-in-100 random sample of the US population. This is a mandatory survey if your household is selected allowing for the closest approximation of the population that is publicly available, respondents are mailed a data packet and then instructed to fill out their responses online to minimize data processing time and the risk of data breach or lost packages in the postal system. The survey includes observations from all states and includes approximately 3 million observations each year.

Being mismatched is defined as being either overeducated or undereducated for a worker's current job. The required level of education is determined by using the mode level of obtained education for a certain occupation. Grouping by occupation codes from the American Community Survey we can find a mode level education for all occupations represented in the survey. For example, if the mode level of education for a coffee shop barista is a high school diploma, then someone with a master's degree would be classified as over educated, while someone in the same job with no high school diploma would be classified as under educated.

One of the main variables of interest in this study are marital status, which is a survey question in the ACS, respondents are asked about their marital status and can respond "Married, Spouse Present" which we recode as a 1 for married. The other options for marital status "Single, Never Married", "Divorced", "Separated" and "Widowed" are all coded as "Single". "Married, Spouse Absent" was recoded as single because this paper is more interested in the family constraints on mobility, for example your spouse having a job would make it more difficult to search for a better fitting job as they do not want to leave their job. The next research question deals with having children, this is a picked from a couple different questions in the ACS. Respondents are asked if they have children living in the home with them, as well they are asked to answer how many children under 5, and how many total children they have in the home. From these questions we determine the variables, number of children under 5 and number of children over 5. The age of 5 years is used to break this variable as it is what is available in the survey, but it also makes a logical breaking point. Most children start attending school at age 5, once a child starts school the family becomes more rooted in their local community, and it may make searching for a better fitting job more difficult. A dummy variable was also included in some of the results removing the age break and setting 1 for if any children are present and 0 if there is no children in the home.

The final research questions this paper hopes to examine is about industries, what industry a person works in is recorded in the ACS as a 4-digit code according to the standard occupational classification system. This 4-digit code is then reclassified into 15 larger categories according to the American Community Survey Occupation Codes: Agriculture, mining, construction, manufacturing, wholesale trade, retail trade, transportation/warehousing, utilities, information, finance, professional, education, health care, arts/entertainment, service, public administration, and military. I choose agriculture as the base group for the first few iterations of my model as it is easy to picture what farming is. Later in the results section by choosing agriculture as the base group this study finds that almost all other industries are more likely to be undereducated and less likely to be overeducated when compared to farmers. So, in future work it may be more informative to choose a different base group or to look at each industry specifically to get a better picture.

Some other demographic variables are also included in the model, and they come directly from the American Community survey. Sex is coded for females as 1 and males as 0, making males the base group. Age is copied directly from the survey but people under 16 are removed from the dataset. These people are removed as 16 is generally the age as to when young people are aloud to start working full time if they so desire. As well people over the age of 65 are also removed from the data set. This is the average age for retirement, this will come up again later in the results section, but as older people are examined the average education level begins to fall, causing them to become more likely to be undereducated but they have a higher level of experience and skills to offset this lower educational attainment. So, to negate this effect of education to skills trade-off the age is cut off at 65. A dummy variable for rase is also included, white is set as the base group but there are 8 other groups in the survey that respondents could self-identify within. A person's ability to speak English comes from the variable speaking in the ACS and is recoded into a dummy variable for either "Yes can speak English" or "No can't speak English". Being able to speak english is the base group, as english is the dominate language in the United States of America, so a person's ability to fluently speak english could greatly affect the type of job their receive and therefore effect the likelihood of a person being mismatched.

Education level is also included in the data, it was recoded from the survey into 5 different groups according to highest level of education achieved. The groups are, No High School Diploma, High School Diploma, Community Collage, Undergraduate Degree, and Graduate Degree. High School Diploma is set as the base group as nearly 73% of the American population have attained a Highschool diploma from our sample.

Table 1: Summary Stats

	Undereducated		Match		Overeducated		Total
Male	48.95%		53.45%		52.88%		51.88%
Female	51.05%		46.55%		48.12%		48.12%
Age	41.9		42.2		41.5		41.8
Married	55.40%		55.18%		55.01%		56.37%

Children	45%		43.93%		42.80%		43.82%
Child Un 5	0.7166019		0.64176		0.6262163		0.6567
Income	39931.41		56251.93		46606.08		48291.65

From the summary stats table we can see that people in the undereducation category have on average more young children under 5 and are more likely to have children in the home. Also, on average we find under, and over educated people have a lower income than their well-matched counterparts. Leading to motivate the second model presented in this paper examining the effect of over and under education on a person's income. We also see that well matched people have a higher average wage than the total sample 56,251 to 48,291. Providing us with a hypothesis into the wage effects that over and undereducation may have on employees.

We see that Females are more likely to be both over and under educated for their jobs than males. As well we can see that the highest educated people with master's degrees and higher are the most likely to be over educated, this could be due to a fault in the coding of education as they have hit an education ceiling, based on how the data is coded it is impossible for them to be undereducated. Surprisingly, people with a community collage degree are more likely to be mismatched than both high school and university graduates. Normally in Canada we view community collage as being very job specific education, a Plummer who receives a plumbing certificate from the Nova Scotia Community Collage has been trained very specifically for a plumbing job and we would expect them to be very well matched.

Methodology:

The main model of interest in this study is testing the impact of a person's marital status (α), if they have children at home (γ), the age of these children, broken up into either over 5 (δ) or under 5 (θ) as well as their industry of employment (μ), on their likelihood of being mismatched (M). Also included is the interaction term between marital status and having children present in your home , it is reported in the regression with a base group of married people without children. The mismatched variable is further broken down into over and undereducation, these results are reported along with the mismatch variable. Other demographic variables are also included, a person's sex (ρ), their age (τ), their age squared (φ) as there is evidence that a person's age has a parabolic effect on a person's mismatch. Their race (π) is also included as well as their ability to speak English (ω). Educational attainment (α_2). I use OLS to estimate the relationships between the independent variables and the dependent variable, mismatch. The mismatch variable is reported either as a 0 or a 1 so this model is using a linear probability model, robust standard errors are reported, the model also makes use of the weights supplied by the ACS to get a better representation of the population in question. The results were also tested by using both a logit and a probit model, these results are depicted in the appendix. Calculating marginal effects for the logit and probit yielded very similar results as the OLS model, so the paper uses OLS for the main model as it has an easier to interpret coefficients.

$$M_i = B_o + B_1 * \alpha + B_2 * \gamma + B_3 * \delta + B_4 * \theta + B_5 * \mu + B_6 * \pi + B_7 * \rho + B_8 * \sigma + B_9 * \tau + B_{10} * \varphi + B_{11} * \omega + B_{12} * \alpha_2 + \varepsilon$$

Robustness checks were also performed by only testing the main variables of interest and slowly adding in the demographic variables and other control variables. Industry variables are added in last as they represented a large number of variables so to control for any volatility they may have added, they were saved till last. These results can be found at the end in the appendix. As the other control variables are added in the results still remain significant and the same sign.

This next model looks at the effect over (α) and under (θ) education has on a person's wage (Y), the dependent variable is only income earned from wages, to see the effect over and under education has primarily on your job's income and not any additional investments a person may have. Also included are the household structure variables, marital status (ρ), if there are children in the house (μ), the interaction term of being single and having children in the home (τ), having children younger than 5 years old (γ) and sex (∂). Also included in the model are education, using a bachelor's degree as the base group there is a variable for having above a bachelor's and below a bachelors. A dummy variable for race either White or other, and the persons age are also included. These full results are in the appendix section.

$$Y = B_0 + B_1 * \alpha + B_2 * \theta + B_3 * \rho + B_4 * \mu + B_5 * \tau + B_6 * \gamma + B_7 * \partial + \varepsilon$$

Results:

In table 3 the results from the main regression are presented.

Table 3: OLS Regression

VARIABLES	(2) undereduc	(1) mismatch	(3) overeduc
Single	-0.0245***	0.000109	0.0246***
	(0.000540)	(0.000628)	(0.000613)
Child Present	0.00415***	-0.00726***	-0.00311***
	(0.000735)	(0.000831)	(0.000815)
Single with Children (Interaction Term)	0.00181**	0.0112***	0.00935***
	(0.000854)	(0.000966)	(0.000949)
Children over 5	-0.00147***	0.00164***	0.00312***
	(0.000294)	(0.000328)	(0.000325)
Children Under 5	0.000218	-0.00441***	-0.00463***
	(0.000488)	(0.000567)	(0.000553)
age	0.00701***	0.00209***	-0.00492***
	(0.000105)	(0.000121)	(0.000117)

age2	-7.20e-05***	-1.87e-05***	5.33e-05***
	(1.24e-06)	(1.42e-06)	(1.39e-06)
Female	0.0215***	0.0248***	0.00324***
	(0.000394)	(0.000459)	(0.000447)
Self Employed	0.0176***	0.0158***	-0.00174**
	(0.000618)	(0.000711)	(0.000713)
Military	0.121***	0.0684***	-0.0530***
	(0.00242)	(0.00290)	(0.00295)
retail	0.00793***	0.0417***	0.0337***
	(0.00114)	(0.00168)	(0.00158)
wsale	0.218***	0.0163***	-0.202***
	(0.00159)	(0.00196)	(0.00187)
mining	0.0602***	-0.0204***	-0.0806***
	(0.00247)	(0.00308)	(0.00303)
Construction	0.0421***	-0.00781***	-0.0499***
	(0.00121)	(0.00172)	(0.00162)
Transportation	0.0150***	0.00696***	-0.00800***
	(0.00129)	(0.00185)	(0.00176)
Utilities	0.0698***	0.00983***	-0.0600***
	(0.00205)	(0.00262)	(0.00263)
infocom	0.244***	-0.00652***	-0.251***
	(0.00166)	(0.00209)	(0.00200)
finance	0.287***	0.0102***	-0.277***
	(0.00130)	(0.00175)	(0.00165)
professional	0.181***	-0.0311***	-0.212***
	(0.00117)	(0.00167)	(0.00158)
educational	0.383***	-0.00631***	-0.389***
	(0.00133)	(0.00177)	(0.00166)
health	0.177***	-0.0468***	-0.224***
	(0.00120)	(0.00170)	(0.00161)
entertainment	0.104***	0.0219***	-0.0820***
	(0.00120)	(0.00172)	(0.00161)
service	0.0738***	0.00800***	-0.0658***
	(0.00131)	(0.00181)	(0.00172)
public	0.234***	0.0220***	-0.212***
	(0.00140)	(0.00183)	(0.00177)
manufact	0.0748***	-0.0149***	-0.0897***
	(0.00116)	(0.00167)	(0.00158)
No HighSchool	0.757***	0.526***	-0.231***
	(0.000468)	(0.000524)	(0.000465)
Community College	0.00330***	0.416***	0.413***
	(0.000570)	(0.000534)	(0.000578)
Undergraduate	-0.254***	-0.0808***	0.173***

	(0.000487)	(0.000692)	(0.000605)
Graduate School	-0.366***	0.185***	0.551***
	(0.000539)	(0.000773)	(0.000702)
Black	-0.0484***	-0.000360	0.0480***
	(0.000636)	(0.000736)	(0.000727)
Native	-0.0284***	0.0158***	0.0442***
	(0.00219)	(0.00239)	(0.00244)
Chinese	0.0150***	0.0185***	0.00351**
	(0.00128)	(0.00171)	(0.00159)
Japanese	0.0110***	-0.00318	-0.0142***
	(0.00318)	(0.00400)	(0.00390)
Otherasian	-0.0122***	0.0230***	0.0352***
	(0.000778)	(0.00103)	(0.000998)
OtherRace	-0.0222***	-0.00183*	0.0203***
	(0.000871)	(0.000992)	(0.000936)
TwoMajor	0.000783	0.0240***	0.0232***
	(0.00131)	(0.00150)	(0.00151)
ThreeMajor	0.000934	0.0169***	0.0160***
	(0.00352)	(0.00416)	(0.00415)
Cant Speak English	0.0449***	0.0316***	-0.0133***
	(0.000794)	(0.000982)	(0.000853)
Constant	-0.0166***	0.389***	0.406***
	(0.00250)	(0.00306)	(0.00292)
Observations	6,806,472	6,806,472	6,806,472
R-squared	0.334	0.213	0.238

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

For our main variables of interest, we find that single people are 2.5% more likely to be overeducated but 2.5% less likely to be undereducated, both these are significant at the 0.01 level. Boudarbat & Chernoff (2010) in their study of recently graduated Canadian university students found that marital status had no effect on under or over education, their study was limited to recent graduates and many of those are not established in the work force and may already be at a higher likelihood to be mismatched. We find the opposite results from Rubb(2014) who found that single people are more likely to be undereducated and less likely to be overeducated. In the study done by Rubb (2014) they do not include any interaction term between marital status and having children present which could be a key part of understanding the relationship between the variables. Having a child present in the household raises the chances of being undereducated and decreases the chances of being overeducated. These effects are depicted further with the inclusion of the interaction term of being single with children present, which increases the likelihood of being overeducated and undereducated. I have a possible explanation for this, single parents may not have a spousal income to rely on while they are searching for a better fitting job so they must take whatever they can get to provide for their children. As for Industry variables, it was touched on previously that agriculture may not have been the best base group for this regression, as almost every other industry is more likely to be undereducated and less likely to be overeducated, we can still interpret the industries with the bigger coefficients. Some of the largest coefficients are on the finance, health and educational variables, one suggestion for these results is that these groups cover a wide range of different professions, for example in the education industry it included teachers from kindergarten all the way to university professors. These two professions would differ widely as well as the finance category includes a wide range, and more research could be done to break down these larger categories into smaller ones for a more in-depth analysis.

Women are more likely in all three categories, where Leuven & Oosterbeek (2011) and Groot & Maassen van der Brink (2000) found that they were only more likely to be over educated and men were more likely to be under educated. When looking at the different dummy variables for race, the results show that there is a consistent positive coefficient for all races compared to white, regarding overeducation. We see the predicted effects of graduate school; they are 36.6% less likely to be undereducated and 55.1% more likely to be overeducated as they represent the most educated people in the population and as you get higher education it becomes more difficult to be overeducated. These are similar to what was found by Duncan & Hoffman (1981) that higher educated people are also more likely to be mismatched.

Table 4: Income as a function of over and under education

VARIABLES	(1) Income
Overeducated	-8,002***
	(54.16)
Undereducated	-764.1***
	(51.02)
Single	-8,202***
	(68.78)
Has Children	6,045***
	(90.92)
Single w Children	435.9***
	(97.90)
Young Children	385.8***
	(36.48)
Female	-18,521***
	(45.89)
Constant	59,089***
	(150.6)
Observations	6,806,472
R-squared	0.214

Robust standard errors in
parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

We can see that in the data, overeducation has a much greater negative effect on income than being undereducated. If someone is working a job where the required education level is a high school diploma and they have a master's degree, comparing them to other people with master's degrees then it makes sense that them being over educated would lead to a negative income effect. The other variables of interest are marital status and having children in the home. We see that single people earn less than married people but having children in the home have a positive effect on income. The interaction term of single people with children also has a positive effect on income comparing to married people with no children. This decrease in income for single people matches up with the possible explanation offered earlier for why single people are more likely to be overeducated, due to not having a spousal income to fall back on while searching for a better fitting or better paying job, they end up in a mismatched job and therefore earn less on average than married people. As for having children increasing income, there is a possibility here for reverse causality, as people earning a higher income could choose to have kids, would have more disposable income for childcare and possibly better perks at their job such as maternity leave. Therefor the income could lead to having children and not the children causing the increase in income.

Another cause of concern is omitted variable bias, when looking at any model concerning income, there is no way to measure a person innate ability or their intelligence or for example their work effort and motivation to do well at a job. All these things may influence their income at a certain job and since we cannot measure them directly in the data in this paper, there may be some bias in the estimation of this model. Using a method depicted in "Unobservable selection and coefficient stability: Theory and Evidence" by Oster (2019) it is possible to get a sense of the possible bias in the model. Using two models one with controls and another without, we can examine the explanation power of the variables we do have access to in the data and see how much of the variance in income is explained by their inclusion. Using the formula mentioned in the above paper we can solve for the coefficient on overeducation, and I found there was only a slight downwards bias on the term in my model. The models used for this are presented in the appendix.

The final issue to discuss is the selection bias, we have already filtered the data by age (only selecting those aged 16-65) and employment status (Full time workers). Therefore, we no longer have a random selection process. If we could follow people over time, we could employ the Heckman selection process which adjusts models for bias when observations are missing due to selection and not randomness. This is a spot for future research, with the data available to me for this project there is no way I can identify individual people in the ACS and follow them to different time periods.

Conclusion:

In conclusion the OLS regression found that single people with children are more likely to be both over and under educated. Having children present yields a mixed result but breaking this down by the age of the children we see that having children over 5 years old increases overeducation while children under 5 decreases overeducation. These results are only one way to look at mismatch, I stated earlier that it can be calculated in 3 different ways and can also look at educational fields along with education level. Another limitation of my results is that I compared education level across occupations, but it is also possible to compare across industries such as in Summerfield & Theodossiou (2017).

Both over and under education were found to have a negative effect on people's incomes, with being single also a negative effect on income. As mentioned above there is a concern for reverse causality between having children and higher income, since people with higher income may choose to have children rather than by having children they make more money.

These are important findings which will hopefully enable policy makers to implement more programs and funding opportunities for single parents and to help them find a better fitting job. To further this research, I would like to apply this same study to Canadian data, to find if these same results hold for Canadians. I would also be interested in changing the measure of mismatch I used to also check the robustness of the results under the different methods of calculating mismatch, as well consider peoples fields of study to get an even deeper look. I mentioned above the wide industry categories, and with more time it would be interesting to break these into smaller, more detailed categories to get a finer picture. It could also be done to run the model on just one industry group at a time to see how the results differ across the different industries.

Appendix:

	Probit	Logit
VARIABLES	mismatch	mismatch
age	0.00424***	0.00400***
	(0.000424)	(0.000703)
age2	-3.29e-05***	-2.62e-05***
	(4.96e-06)	(8.21e-06)
Female	0.0483***	0.102***
	(0.00139)	(0.00231)
Community Collage	1.315***	2.247***
	(0.00202)	(0.00376)
Undergraduate	-0.216***	-0.346***
	(0.00173)	(0.00278)
Graduate	0.450***	0.727***
	(0.00203)	(0.00330)
Single	0.00199	0.00633*
	(0.00202)	(0.00332)

Child Present	-0.0211***	-0.0338***
	(0.00277)	(0.00456)
Single with Child	0.0287***	0.0489***
	(0.00326)	(0.00539)
Children >5 Years Old	0.00353***	0.00644***
	(0.00115)	(0.00190)
Children <5 Years Old	-0.0176***	-0.0309***
	(0.00191)	(0.00315)
Constant	-0.170***	-0.220***
	(0.00849)	(0.0141)
Observations	6,332,315	6,332,315

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Marginal Effects	Probit	Logit
Single	0.0047206	0.0054656
Children Present	-0.0025122	-0.0021891
Children >5 Years Old	0.0011587	0.001281
Children <5 Years Old	-0.0057856	-0.0061486

Appendix Table 2: Robustness Checks

VARIABLES	mismatch	mismatch	mismatch
age	0.000675***	0.000954***	0.00217***
	(0.000118)	(0.000118)	(0.000120)
age2	-3.75e06***	-6.51e-06***	-1.91e-05***
	(1.40e-06)	(1.41e-06)	(1.42e-06)
Women	0.0192***	0.0191***	0.0254***
	(0.000420)	(0.000421)	(0.000456)
Single	0.00462***	0.00507***	0.00477***
	(0.000492)	(0.000496)	(0.000496)
ChildrenPresent	-0.00277***	-0.00299***	-0.00231***
	(0.000747)	(0.000747)	(0.000746)
Children <5 Years Old	-0.00570***	-0.00532***	-0.00435***
	(0.000568)	(0.000568)	(0.000567)
Children >5 Years Old	0.00103***	0.00139***	0.00134***
	(0.000328)	(0.000328)	(0.000328)
No High School	0.518***	0.526***	0.526***
	(0.000449)	(0.000513)	(0.000524)
Community Collage	0.414***	0.413***	0.416***
	(0.000531)	(0.000531)	(0.000534)
Undergraduate	-0.0851***	-0.0874***	-0.0812***
	(0.000671)	(0.000676)	(0.000691)
Graduate	0.176***	0.173***	0.184***
	(0.000754)	(0.000761)	(0.000772)
Black		-0.00241***	-4.16e-05
		(0.000732)	(0.000735)
Native		0.0166***	0.0161***
		(0.00239)	(0.00239)
Chinese		0.0197***	0.0182***
		(0.00172)	(0.00171)
Japanese		-0.00134	-0.00340
		(0.00400)	(0.00399)
Otherasian		0.0215***	0.0228***
		(0.00104)	(0.00103)
OtherRace		-0.00273***	-0.00169*

		(0.000991)	(0.000992)
TwoMajor		0.0250***	0.0241***
		(0.00151)	(0.00150)
ThreeMajor		0.0191***	0.0169***
		(0.00416)	(0.00416)
spkeng		0.0318***	0.0316***
		(0.000977)	(0.000982)
mil			0.0684***
			(0.00290)
retail			0.0415***
			(0.00168)
wsale			0.0162***
			(0.00196)
const			-0.00786***
			(0.00172)
mining			-0.0203***
			(0.00308)
manufact			-0.0150***
			(0.00167)
transport			0.00686***
			(0.00185)
utilit			0.00983***
			(0.00262)
infocom			-0.00671***
			(0.00209)
finance			0.0101***
			(0.00175)
professional			-0.0311***
			(0.00167)
educational			-0.00653***
			(0.00177)
health			-0.0468***
			(0.00170)
entertainment			0.0217***
			(0.00172)
service			0.00790***
			(0.00181)
public			0.0220***
			(0.00183)
classw			0.0159***
			(0.000711)
Constant	0.454***	0.416***	0.384***
	(0.00233)	(0.00263)	(0.00303)
Observations	6,806,472	6,806,472	6,806,472
R-squared	0.209	0.209	0.213

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As the variables for race and occupation are added into the regression, we can see that the estimations for the relationships between the household structure variables and mismatch remain similar. The coefficients on the household structure variables are all statistically significant at the 0.01 level regardless of the demographic variables added in.

Appendix Table 3: Omitted Variable Bias

VARIABLES	Long incwage	Short incwage
overeduc	-8,382***	-9,775***
	(48.79)	(50.68)
undereduc	-293.3***	-1,489***
	(45.06)	(56.21)
sex	-20,192***	
	(41.04)	
age	633.5***	
	(1.465)	
1.marst2	-8,259***	
	(57.76)	
1.childp	6,554***	
	(79.70)	
0b.marst2#0b.childp	0	
	(0)	
0b.marst2#1o.childp	0	
	(0)	
1o.marst2#0b.childp	0	
	(0)	
1.marst2#1.childp	663.7***	
	(85.07)	
speakeng	-3,698***	
	(23.12)	
White	2,590***	
	(44.00)	
gbch	28,554***	31,578***
	(110.4)	(74.98)
ubch	-29,506***	-29,956***
	(52.96)	(54.59)
nchild5p	577.8***	
	(33.39)	
Constant	60,303***	67,497***
	(132.6)	(48.52)
Observations	6,806,472	6,806,472
R-squared	0.209	0.137

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

The formula for the true coefficient from the Oster (2019) paper is

Btrue= Blong-(Bshort-Blong)(R2max-R2long)/(R2long-R2short)
So using the above data I found a Btrue of -7156.8169, so in our model we should examine the results with a slight upwards bias for the coefficient on overeducation.

References:

Allen, J., & van der Velden, R. (2001). Educational mismatches versus skill mismatches: Effects on wages, job satisfaction, and on-the-job search. *Oxford Economic Papers, 53*(3), 434-452. doi:10.1093/oep/53.3.434

Bowlus, A. J. (1995). Matching workers and jobs. *Journal of Labor Economics, 13*(2), 335-350. Retrieved from http://www.econis.eu/PPNSET?PPN=26058410X

Cappelli, P. H. (2015). Skill gaps, skill shortages, and skill mismatches: Evidence and arguments for the united states. *ILR Review, 68*(2), 251-290. doi:10.1177/0019793914564961

Chernoff, V., & Boudarbat, B. (2010). *The determinants of education-job match among canadian university graduates.* (). Retrieved from http://econpapers.repec.org/paper/circirwor/2010s-14.htm

Duncan, G. J., & Hoffman, S. D. (1981). The incidence and wage effects of overeducation. *Economics of Education Review, 1*(1), 75-86. doi:10.1016/0272-7757(81)90028-5

Groot, W., & Maassen van den Brink, Henriette. (2000). Overeducation in the labor market: A meta-analysis. *Economics of Education Review, 19*(2), 149-158. doi:10.1016/S0272-7757(99)00057-6

Hartog Joop. (2000). Over-education and earnings: Where are we, where should we go? *Economics of Education Review, 19*(2), 131-147. doi:10.1016/S0272-7757(99)00050-3

Leuven, E., & Oosterbeek, H. (2011). Overeducation and mismatch in the labor market. *Discussion Paper Series // Forschungsinstitut Zur Zukunft Der Arbeit, 5523* Retrieved from https://www.econstor.eu/handle/10419/51653

Rubb, S. (2014). Factors influencing the likelihood of overeducation: A bivariate probit with sample selection framework. *Education Economics, 22*(1-2), 181-208. doi:10.1080/09645292.2011.555971

Summerfield, F., & Theodossiou, I. (2017). The effects of macroeconomic conditions at graduation on overeducation. *Economic Inquiry, 55*(3), 1370-1387. doi:10.1111/ecin.12446

Turmo-Garuz, J., Bartual-Figueras, M. -., & Sierra-Martinez, F. (2019). Factors associated with overeducation among recent graduates during labour market integration: The case of catalonia (spain). *Social Indicators Research,* , 1-29. doi:10.1007/s11205-019-02086-z

Verhaest, D., & Omey, E. (2006). The impact of overeducation and its measurement. *Social Indicators Research, 77*(3), 419-448. Retrieved from http://www.econis.eu/PPNSET?PPN=516249525

Yuen, J. (2010). Job-education match and mismatch: Wage differentials. *Perspectives on Labour and Income, 22*(2), 26-36. Retrieved from https://search.proquest.com/docview/609462210

Oster, E. (2019). Unobservable selection and coefficient stability: Theory and evidence. Journal of Business and Economic Statistics, 37(2), 187-204.

Retrieved from https://www.tandfonline.com/doi/full/10.1080/07350015.2016.1227711?scroll=top&needAccess=true