Pablo López Aparicio

GENDER DIFFERENCES IN JOB SEARCH: How men and women are affected by labor problems in the early stages

BACHELOR'S THESIS

Labour Economy

Economics Degree

FACULTAT D'ECONOMIA i EMPRESA
Universitat Rovira i Virgili

Reus

Academic Year: 2024/2025

Table of Contents

List of Tables

Abstract

English

This thesis analyses the differences between the jobs that men and women execute on average during the first stages of their labor market career, based on a database of different alumni from the URV. There, we can find the jobs the individuals had before, during, and after graduation, along with the conditions of these positions. In the beginning, we contextualize the gender differences that we can find between men and women on average, explaining situations and concepts like the school-to-work transition and risk-aversion. After that, the paper examines through an econometric study three different situations or problems that, according to previous studies, affect men and women differently: The overeducation problem (probit model), the part-time job (probit model), and the time that the individuals expend before getting a high-qualification job (linear regression model). The results showed that being a woman increases the possibility of working at a part-time job and the time individuals spend before getting a high-qualified job, but decreases the possibility of being overeducated. During the study, as we use them to isolate the effect of gender on the dependent variables, we also examine how other variables, such as age or previous job experience, affect the problems that we analyse.

Key words: gender, job search, school-to-work transition.

Spanish

Esta tesis analiza las diferencias entre los trabajos que acostumbran a desempeñar hombres y mujeres durante las primeras etapas de su carrera en el mercado laboral, basándose en una base de datos de distintos exalumnos de la URV. En ella, se pueden encontrar los trabajos que los individuos tuvieron antes, durante y después de graduarse, junto con las condiciones de dichos puestos. Al inicio, se contextualizan las diferencias de género que existen, en promedio, entre hombres y mujeres, explicando situaciones y conceptos como la transición de la escuela al trabajo o la aversión al riesgo. Posteriormente, el estudio examina, a través de un análisis econométrico, tres situaciones o problemas distintos que, según estudios previos, afectan de manera diferente a hombres y mujeres: el problema de la sobre educación (modelo probit), el empleo a tiempo parcial (modelo probit), y el tiempo que los individuos tardan en acceder a un empleo de alta cualificación (modelo de regresión lineal). Los resultados muestran que ser mujer aumenta la probabilidad de trabajar en un empleo a tiempo parcial y el tiempo que se tarda en conseguir un empleo de alta cualificación, pero disminuye la probabilidad de estar sobre educada. Durante el estudio, también se examina cómo otras variables, como la edad o la experiencia laboral previa, afectan a los problemas analizados, ya que se utilizan para aislar el efecto del género sobre las variables dependientes.

Palabras clave: género, búsqueda de empleo, transición entre el mundo académico y el laboral.

Catalan

Aquesta tesi analitza les diferències entre les feines que acostumen a realitzar homes i dones durant les primeres etapes de la seva trajectòria al mercat laboral, mitjançant una base de dades de diferents antics alumnes de la URV. En aquesta base de dades es poden trobar les feines que els individus van tenir abans, durant i després de graduar-se, juntament amb les condicions d'aquests llocs de treball. A l'inici, es contextualitzen les diferències de gènere que es poden trobar, normalment, entre homes i dones, explicant situacions i conceptes com la transició de l'escola al món laboral i l'aversió al riks. Tot seguit, l'estudi examina, mitjançant una anàlisi econometric, tres situacions o problemes diferents que, segons estudis previs, afecten de manera diferent a homes i dones: el problema de la sobre educació (model probit), la feina a temps parcial (model probit) i el temps que els individus tarden en accedir a una feina d'alta qualificació (model de regressió lineal). Els resultats mostren que ser dona augmenta la probabilitat de treballar en una feina a temps parcial i el temps que es triga a aconseguir una feina d'alta qualificació, però disminueix la probabilitat d'estar sobre educada. Durant l'estudi, també s'examina com altres variables —com l'edat o l'experiència laboral prèvia— afecten els problemes analitzats, ja que s'utilitzen per aïllar l'efecte del gènere sobre les variables dependents.

Paraules clau: gènere, cerca de feina, transició entre el món acadèmic i el laboral

Presentation

The reason behind the choice of this topic is mainly our interest in all the problems that occur in the labor market. Of course, the discussion over gender disparities has played a major role in the social discussions, but even more in the academic research. Despite that, we observed that economists have focused their interest on which are the effects that this disparity has on the wages that men and women earn on average, and not on which are the causes behind the differences. Also, as we are finishing our degrees and entering the labor market, the school-to-work transition is something that we will be doing soon, and no one is talking about what we must do to have a good start to it. With this Thesis, we wanted to link these two concepts and analyse those differences that we will be observing first hand during the next years.

During this bachelor’s thesis, there are multiple subjects of the degree that we will need to work on to explain and comprehend all the ideas and concepts that we will be talking about, but there are two that will have the most importance in it: Labor economics and econometrics. The concepts about labor economics will play a clue role in the theoretical part of this work, since we will be talking about job search, overeducation, and part-time work, among others. And econometrics will be fundamental to execute the empirical part, since we will be running two different models (probit and linear), with different kinds of variables (dummy, linear, and count), and applying multiple econometric tests to make sure that our models are valid.

During this work, we will be able to show different “hard skills” that we have been developing during the bachelor's (Empirical application of the concepts that we learned and the methods that we studied, for example), but also multiple “soft skills” such as analytical capacity, reading comprehension of economic studies and clearness of writing expression, etc.

As a conclusion, this study aims to explore the reasons behind the differences that we can find in the jobs that men and women execute on average on the labor market during the first years in the labor market, and analyse if being a woman increases the possibilities of suffering some problems that the jobs positions usually have. All that while applying part of the concepts that we learned during the degree.

Chapter 1: Introduction

1.1. Introduction and objectives

Mistakenly, in the last few decades, economists who have tried to answer the question of what the causes behind the gender wage gap are, have focused their studies just on the differences that men and women have on average in household responsibilities. When we look for information on why, even nowadays, that the gender gap has been significantly reduced, there is still a wage gap on the labor market, we find a substantial quantity of papers and studies that analyse the different effects that childbirth has on the parents. For example, the existence of a "dominant" wage (usually of men) that encourages women to work in part-time jobs so they can take care of the children(Kleven et al., 2019), instead of distributing the responsibilities between the two parents, it causes women to accept lower salaries since the earnings of part-time work appear to be substantially lower than those from full-time work (Francesconi, 2003), not just because of a difference in earnings per hour, but also because working part-time gives the individual access to a different job position than working full-time.

The problem with this idea is that, if after childbirth is the only moment where men and women have different behaviour facing the labor market (motivated by social norms or personal decisions), it would be assumed that the wage gap will only appear after the moment when the first children are born. Even if we consider the expectations of having a child, we could consider that it will appear from the moment that the couple knows that they will be having a baby, but not before. And that is not the case. Although this difference increases at the point when a child is born, when we observe the differences in wages between men and women in the early stages of their professional careers, we can observe that even some recent graduate women of academic careers that count with a larger number of them, receive a lower average daily wage than men (Sandner & Yükselen, 2024).

For that reason, a smaller number of economists have been asking themselves if there is another turning point that could also be responsible for causing this gender gap. Among the points that they had found, the one that seems to be the most relevant, and the one that we will be analysing, is the job search.

During the job search, we can observe multiple situations where individual preferences have a very important role in the decisions that each person makes and the paths that they select. This fact made economists ask themselves if there is a possibility that the differences that we can find between men and women could affect the behaviour that they have when facing the entrance to the labor market, regardless of the origin of these differences, which could be caused by psychological traits or established social norms.

This study aims to look beyond what we can find when we observe just the problem. There is a big part of the literature that talks about the gender gap that focuses its interest on the wages that men and women receive on average. Despite this, we believe that restudying the problem will not end it, but that the important part is knowing which are the causes that are inducing the wage disparity. During this bachelor's thesis, the objective will be to study the differences that we can find when observing the traits of the job positions that recent graduates have on average, and see which differences we can find between the ones that men and women usually have.

1.2. Content and structure

This study is divided into five different chapters that will help us analyse the problems that men and women find in the first stages of their labor career, and how each gender is affected by them. These five chapters can be grouped as two main sections: theoretical (chapters 1, 2, and 5) and empirical (chapters 3 and 4).

The first chapter works as an introduction to the study. In this part, we will try to explain the problem that we find and why we are trying to do research around it, how the study will be structured, and lastly, the methodology that we will be following.

After that, we find chapter two, where we will contextualize the subjects that we will be analysing in the empirical part and review all the literature that we found around it.

In Chapter 3, we talk about the data and all the variables that we will be using to run the econometric models, along with an explanation of those models.

In the fourth chapter, we will show the results of running our models and we will explain all the processes that we followed.

Finally, the fifth and last chapter will work as an explanation of the conclusions of the study.

1.3. Methodology

The empirical part of this thesis shows an econometric analysis of the problems that men and women face in their first years of professional career, and how these problems affect them depending on their gender. The dataset that we use during the empirical study is constructed by several researchers from the University Rovira i Virgili, and it shows the multiple job positions that each individual had during their first years at the labor market (before, during, and after they graduated).

For the estimation of the econometric models, we have used the statistical software "Stata".

Chapter 2: Literature Review

To approach the subject of this study in the best possible way, we have read different literature around the gender gap in the job search. During this chapter, we will discuss the most essential information we found and analyse it to summarise it in the best possible way.

When we try to find the causes behind the gender gap, we always find that economists blame the same factor: the household distribution of responsibilities. The problem with this idea is that economists usually blame it as the only cause. Despite this, during the past years, another group of economists has realized that if this is the only cause behind the gender gap, this will be expected to appear only when a new family is established, but not before. For example, after childbirth(Kleven et al., 2019). A moment where the parents will have to share their responsibilities to take care of the child, at the same time that they are still working, so they can live in the best possible way. When the economists looked at the wages of recent graduates and compared them between men and women, they found what they were suspecting, that the gender gap also appears in the early stages of work life(Sandner & Yükselen, 2024).

With that information, they realized that outside of household distribution, there needs to be another turning point where men and women act differently, and that the way of behaving creates a gender gap. That turning point ended up being the job search.

This moment has a clue effect over the gender gap, down to the fact that it is directly affected by individual preferences and constraints. During our work, we will observe how individuals approach the labor market based on their psychological and sociological traits, but we will mainly focus on the gender gap that results from the job search, which is caused during this process by the differences that men and women have in those psychological and sociological traits.

2.1. Early career patterns

The job search is an event that occurs at multiple moments in the working life of individuals. Despite that, now we will be focusing on the job search for recent graduates. The job search is a fundamental part of the entrance to the labor market for recent graduates. This moment of change between academic life and entrance to the labor market is known as the “school-to-work transition”. Down to the general misunderstanding of the causes behind the gender gap in the labor market during the first years after graduation, there exists a lack of information about the importance of those years for the future jobs that graduates will hold.

Before talking about the school-to-work transition, it is important to know that the differences between men and women date back some years. Specifically, from the moment students choose which academic path they want to take. The tendency of men to opt for STEM careers (Science, technology, engineering, and mathematics), while women tend to choose more health, education, or humanities studies, has a huge effect on the kind of jobs that they will have in the future. Now that we know this relevant fact, we can talk about the importance of the school-to-work process.

The job search is a fundamental part of the school-to-work transition because it represents the beginning of some differences between men and women, since, down to the different psychological and sociological traits that men and women have on average, the results that they will get from this moment are different. At the same time, it plays a crucial role in determining not only the jobs that graduates will have at the moment of entering the labor market, but also the future job positions that they will have. This is due to the influence that current job positions have on the evolution that individuals will have during their work life.

We can observe the importance of school-to-work transition and the job search at this moment when we look at the average GPA(Grade Point Average) that men and women tend to have nowadays, and how, despite having a better starting point, this may be affected by different personal traits. During the last few years, we have been able to observe how nowadays women tend to have a higher GPA than men(Becker et al., 2010). Knowing that, we could think that women will even have a better job position after graduation. The problem comes when we observe that despite these, there still exists a gender gap (in the benefit of men) between the jobs that men and women execute on average. The explanation for this fact could be that women tend to negotiate less in their job interviews than men(Bertrand, 2011). This problem does not appear just in the job search of recent graduates, but if we add the fact that when recent graduates are facing their first job interviews, they have less confidence in their abilities and are in a worst situation to demand better conditions, the situation of women negotiating less make it even worse.

During this school-to-work transition, we can also observe another factor that allows us to comprehend how these different personal traits between men and women will affect them. When we look at the time between the moment that graduated students start the job search and the moment when they accept a job position, we can find a clear difference. Women tend to accept jobs about one month earlier than men (Cortés et al., 2023). The problem with this difference is the same as the previous one: The lack of negotiation from women on the labor market. Accepting a job position earlier, whether due to a lack of negotiation or an absence of a comparative analysis of the job position with other ones, will probably lead to a worse-paying job. Another important idea about this subject is the fact that a big part of students usually accept jobs before graduation. When we look at the gender levels of accepted jobs before graduation, we see that while only 52% of males have accepted a job before graduation, 60% of women have done it (Cortés et al., 2023).

Despite not being what we aim to analyse, we can observe the consequences of this situation by looking at the wage gap of graduates immediately after entering the labor market and the evolution that it has in the first years of a professional career. For example, we can observe how women who graduated in humanities and social sciences, despite having a higher rate in this sector, receive a lower average daily wage(Sandner & Yükselen, 2024). Anyways, this is not a situation that repeats itself in all the different sectors, since when we look at science-related jobs, there is no existing evidence that shows us that there may be a gender wage gap.

2.2. Personality Traits and Their Importance Over the Job Search and the Gender Gap

As we have been saying, one of the main factors that affects the gender gap and has nothing to do with any direct discrimination form is the different personality traits that men and women have on average. Despite the existence of many studies that talk about how companies appreciate or value the traits of the people that they are going to hire (wages), there is a lack of literature around how these personal traits affect the job search behaviours.

The two main concepts used in the study of personal traits based on the differences in preference and constraints are the levels of risk tolerance and over-optimism around the job search. The levels of risk tolerance and over-optimism have an essential role in the decisions that men and women make in their professional careers, since searching for a job is a process that involves a considerable number of uncertainties. On average, men have a higher risk tolerance and an over-optimistic concept about job search(Cortés et al., 2023). This idea influences the kind of jobs that men and women will end up having on average down to the fact that if women have a higher risk-aversion level, intuitively, they will have a lower reservation wage, which means that they will necessarily start searching for jobs and accepting job offers before, which will lead them to worse-paying jobs. In the same way, if men have lower risk-aversion levels, they will have a higher reservation wage, and in the end, they will accept jobs later.

For the same reason that risk tolerance and over-optimism play an essential role in job search, down to the levels of uncertainty that this moment has, different literature that exists on this subject has shown that individuals usually have subjective beliefs about how their efforts in the search process affect the quantity of job offers arrivals (Caliendo et al., 2010). While some people believe that the rate of job offer arrival depends on their perceived "locus of control" (a measure of how much they think their success rate depends on the things that they can control), other people think that the most important decisions are more related to external factors. In the end, they found out that the individuals who have this "locus of control" search for jobs more intensively and have a higher reservation wage.

One problem that has been found is that despite having the same personality traits, men and women are affected differently by them. While women on average are more conscientious than men, men receive a higher wage compensation for being it than women. Also, both men and women are affected negatively by being agreeable, but the channel through which this "punishment" act by is different for each gender. While men are affected in their productivity, women take a disadvantage in the bargaining process (Flinn et al., 2019).

The problem with this last statement is that the idea that we had about personality traits and them having nothing to do with any direct discrimination form disappears. If it is true that some men's and women's traits are taken into consideration differently, just based on gender, this will show a possible discrimination case.

2.3. Differences in the job position

All these different personality traits that we discussed in the past section have a clue effect on the decisions that men and women make over their labor market careers, and at the same time, these decisions play a crucial role in shaping the jobs that they will end up having.

When we look at the job positions that men and women have on average, we can observe differences that put women in a worse position to close the wage gap. A problem that we can find looking at the studies that have been made on this subject is that most of them have been made around the idea that the point where the gender gap occurs is just the moment when a family is formed, and not before. In our study, we will check if this idea also appears in the early stages of a professional career to see if the explanation that it is caused by a recent childbirth is the only reason behind it.

For example, there is a part of the literature that talks about the different things that men and women are willing to trade off on average (and for what they are willing to trade off), with the idea of finding the job that better fits their necessities and objectives in the long term. When we analyse them, we can see how women usually are favourable to trade off a higher portion of their potential earnings in exchange for shorter commutes to job places (Le Barbanchon et al., 2021). The cost of going to the job places has increased significantly over the last decades. For example, in the US, the average two-way commute for a full-time job has increased from 45 minutes in 1984 to an average of 54 minutes in 2016 (an increase of 26%). Moreover, more than 20% of workers spend over 90 minutes on their commute (Farré et al., 2022). Also, if we look at the European case, the costs are comparable in size and increment (Gimenez-Nadal & Molina, 2014).

Anyways, this idea of women accepting shorter commutes in exchange for a lower wage is only supposed to appear when the so-called child penalty occurs. This term refers to the penalty that women usually suffer on the labour market after having a child, due to the time that they must dedicate to unpaid childcare instead of dedicating it to any paid employment. The fact that they prefer shorter commutes may be explained by the facility to drop off or pick up the children from daycare or school. This penalty only exists for women and not for men, on average, due to some social norms that give the responsibility of taking care of the children only to the mother. This can have many long-term effects on the economic opportunities that women receive. As we advance in this study, we will be able to comprehend the impact that the child penalty has on the decisions that women make over their professional careers.

Another point of difference between the jobs that men and women execute on average is the hours that they invest in them. Part-time jobs have been something much more prevalent among women since the first years of this decade. In 2000, while just 2% of men who worked were doing it in a part-time job, 17% of women were (Farré et al., 2022). Also, if we look at the hours worked by each gender on average, we can observe that while women were working an average of 39.55 hours, men worked 46.3 hours on average.

In the same way as the commute time, this has been shortly studied on the early stages of labor life, given that it is only supposed to appear once a couple has responsibilities to share at the household.

One difference that does not have a clear explanation with the child penalty is the overeducation problem in the labour market. This problem refers to the achievement of an educational level that is higher than the one the current job requires to perform it (Boto-García & Escalonilla, 2022). This problem has usually been attributed to the increase in the number of graduates and the impossibility of incorporating them into the labor market.

When we look at the gender differences that we can find over that problem, we can see that there is no consensus on whether there is a gender gap or not. While some studies show that women are more likely to be overeducated (Barone & Ortiz, 2011; Charalambidou & McIntosh, 2021), others do not find significant differences (Caroleo & Pastore, 2018; Devillanova, 2013). Also, in the debate on whether this problem is caused by the different household distribution between the parents or not, there is no clear agreement. While some people think that women are more exposed to the overeducation problem since they are affected by the child penalty (Frank, 1978), others do not find clear evidence on this (Mcgoldrick & Robst, 1996).

Chapter 3: Methodology and Data

3.1. Objectives and database

During the empirical work of this research, we will try to observe if there is any evidence that a gender gap exists on the jobs that men and women execute on average once they have finished the job search of the school-to-work transition. For that, we will look after any evidence that for example there exists a difference on the possibility of being overeducated between genders or how much being a woman increases (or decreases) the time that the individuals had to spend before getting their first high-qualified job. Also we will look if it is true that women tend to work more at part-time jobs than men on their first stages of career.

To achieve this, the empirical part will consist of an econometric analysis of our database through different models and other statistical measures. We will be combining the main explanatory variable of our work (women) with other explanatory variables that may help us isolate the effects of being a women on the dependent variables, controlling for any external effects that those may cause.

Our study is based on a dataset constructed by several researchers from the University Rovira i Virgili. This dataset shows the first work years of different university graduates of the University Rovira i Virgili who graduated between 2007 and 2020 (specifically graduates from 2007, 2008, 2010, 2013, 2016, 2018, 2019, and 2020). All the data that this database contains has been extracted from five different sources: University Quality Agency of Catalunya (AQU), Ministry of Science, Innovation and Universities of Spain - combining the integrated system of university information and working life from the Social Security Treasury of the Ministry of Work, Migration and Social Security -, Public Service Stat Employment of Spain (SEPE), Iberian Balance Analysis System (SABI) and finally Academic Management of the University Rovira i Virgili (URV).

3.2. The Variables

For our model, we studied the variables of a database with 151.283 observations and 318 variables. The dataset showed the first years of work life of different graduates of the URV who graduated in some years between 2007 and 2020 (specifically graduates from 2007, 2008, 2010, 2013, 2016, 2018, 2019, and 2020). Obviously, we will not be using all the variables that this dataset contains. Table 1 shows the variables that we will use and outlines brief characteristics of those variables and data sources.

Table 1: Description of the variables

Function	Variables	Name in the model	Description	Unit
Dependent	Part-Time	part_time	Indicates 1 if the individual works part-time. If not 0	Dummy
	Overeducation	overeducation	Indicates 1 if the individual is overeducated in their work once he has graduated. If not 0.	Dummy
	Time before highly qualified work	daystohighqcontr	How much time has passed after graduation till the individual has had a highly qualified job position	Count

Explanatory	Gender (Women)	women	Indicates 1 if the individual is a woman. If it is a Man 0.	Dummy
	Age	age	Age of the individual at the moment when he has/had the work	Count
	Experience	experience	How many job positions has the individual had before this one	Count
	GPA	GPA	Indicates the average grade of the individual over 10	Count
	Education	edu	The variable indicates the academic path that the individual has followed[1]	Categorical
	Sector	Sector	The variable indicates the sector of the job position[2]	Categorical

[1] Natural Sciences, Social Sciences, Engineering, Humanities, Health
[2] Primary Sector, Extractive industries, Manufacturing, Energy, gas, water and waste management, Construction, Trade and vehicle repair, Transport and storage, Hospitality and food services, Information and communication, Finance and insurance, Real Estate activities, professional and scientific services, Administrative and support services, Public administration. Education, Health and social services, Arts and entertainment, Other services.

3.2.1. Dependent Variables

During this part of the work, we will study 2 different traits of the job positions, and one situation where the individual traits have a high importance: The overeducation problem, the part-time work, and the time between the moment that the individual graduated and the moment that he got a high-qualified work. For that reason, we have selected 3 different dependent variables that will have their own model. Next, we are going to explain each one:

- **Part-time work**: Part-time work is a recurrent situation that workers face in their labor careers, and the studies have shown that women usually tend to work more at part-time jobs. In our study, we will consider part-time work as everything that is under 30 hours worked per week. For studying this variable, we have chosen to create a dummy variable that indicates 1 if the individual works less than 30 hours per week, and 0 if they work more than 30 hours per week.

- **Overeducation:** The Overeducation problem refers to the situation in which the worker's level of education is higher than the level that the job position that he has requires to execute it optimally. For studying this variable, we have chosen to create a dummy variable that indicates 1 if the level of education from the individual is higher than the level of education that his job position requires, and 0 if it's lower or the same as the level that he has.

- **Time before highly qualified work:** One personality trait that has been discussed by economists, which has a high level of importance in the job search process, is the level of negotiation that individuals do. Usually, studies have shown that women tend to negotiate less in their job interviews and that this fact makes them end up in worse jobs. To study if that is true, we are going to analyse the time that each individual has spent before getting a highly qualified job.

3.2.2. Explanatory Variables

The variables that will work as explanatory variables in our study will be the gender (the most important one), the age of the individuals, the quantity of job positions before graduation, the academical career that they have studied (if it is a STEM career or not) and the GPA of that career. We have chosen these variables because we consider them as the most important ones in the school-to-work transition. Now, we are going to explain each one:

- **Gender:** The gender variable will be created as a dummy variable, where the observations will indicate 1 in the case that the individual is a woman, and 0 in the case that it is a man. This variable will be the main explanatory variable of the study, and we will use the rest of the variables to isolate the effects of gender from other factors that also influence.

- **Age:** The variable age collects the age of the individual at the moment he is working in the job position. We include it to be able to capture any effect that age could have on the explanatory variables. Despite that, in the school-to-work transition, its effect could be limited, as some of the individuals are older than others; we think it could have consequences. We will also consider age^2 to contemplate any non-linear effect that could appear.

- **Experience:** This variable will count the number of job positions that the individual has had before the job that he has in that moment, so we can have control of the effects of work experience on the jobs that they will have after, despite some of these jobs being unrelated to the studies that this person has had.

- **GPA:** The variable GPA will represent the grade that the individuals have achieved in their degree. As we have discussed in the literature review, studies have shown that women currently have a higher GPA on average than men. This fact may influence the first jobs of the individuals after graduation.

- **Education:** The variable edu indicates the academic path that the individual has followed. In our model, we will use the command "i." in Stata, which creates a dummy variable for each academic path and introduces it into the model as different variables.

- **Sector:** Similarly to the variable edu, the variable sector indicates the sector of the job position, and we will introduce it in the model in the same format.

3.3. The Hypothesis

Our work will follow the same main hypothesis on the three models that we will run:

H0: Being a woman shows a positive statistically significant relationship to Overeducation, part-time work, and the time before getting a high-qualified work.

H1: Being a woman does not show a positive statistically significant relationship to Overeducation, part-time work, and the time before getting a high-qualified work.

On the table below, we can observe the expected sign of the explanatory variables over each dependent variable, and the justification for each one:

Table 2: Expected sign of the variables

Dependent variable	Explanatory variable	Expected relationship	Justification
Overeducation (dummy)	Gender (dummy)	+	If the gender gap exists against women in the overeducation problem, the gender will have a positive effect since we are showing a dummy variable with women.
	Age	-	The older the individual is, the more experience and trajectory he is expected to have, so the overeducation problem should

			decrease. Despite that, we are not sure of the effect since we are looking just at the years after graduation, and not long-term.
	Experience	-	Having more working experience, despite this not being in the same sector as the next job, should help in the working transition.
	GPA	-	A higher grade at the degree should help the school-to-work transition and lower the overeducation problem in the first stages of the job search process.
	edu	+/-	Depending on the academic path that we will look at, we will get a positive or a negative impact
	Sector	+/-	Depending on the academic path that we will look at, we will get a positive or a negative impact
Part-time work (dummy)	Gender (dummy)	+	If the gender gap exists against women in the part-time work levels, the gender will have a positive effect since we are showing a dummy variable with women.
	Age	-	As the individual becomes older, the job positions they will get should be more stable and full-time. Despite that, we are not sure of the effect since we are looking just at the years after graduation, and not long-term.
	Experience	-	Having more working experience, despite this not being in the same sector as the next job, should help in the transition to jobs with better conditions.

	GPA	-	A higher grade in the degree should help the school-to-work transition to job positions with better conditions.
	edu	+/-	Depending on the academic path that we will look at, we will get a positive or a negative impact
	Sector	+/-	Depending on the academic path that we will look at, we will get a positive or a negative impact
Time before getting a highly qualified job	Gender (dummy)	+	If the gender gap exists against women in access to higher-qualified jobs, the gender will have a positive effect since we are showing a dummy variable with women.
	Age	-	The older the individual is, the more experience and trajectory he is expected to have, so access to higher-qualified jobs should be easier. Despite that, we are not sure of the effect since we are looking just at the years after graduation, and not long-term.
	Experience	-	Having more working experience, despite this not being in the same sector as the next work, should help in accessing higher-qualified jobs.
	GPA	-	A higher grade at the degree should help the school-to-work transition and lower the expected time of waiting for higher education jobs.
	edu	+/-	Depending on the academic path that we will look at, we will get a positive or a negative impact
	Sector	+/-	Depending on the academic path that we will look at, we will get a positive or a negative impact

As we can see in the table above, except for gender (and edu and sector that may have both), all the other variables have a negative effect on the dependent variable. All the variables that we added over the principal one, that is, the gender, have the objective of dissipating and controlling for individual factors that could influence the result, allowing us to isolate the effects of the gender on the dependent variables.

3.4. The Model

Our model studies the effects of gender on three different variables, using 4 extra explanatory variables to isolate the impact of gender on the dependent variables. Our first two dependent variables (overeducation and part-time) will be dummy variables (is overeducated or not, and whether works part-time or not). At the same time, the third one (Time before getting a high-qualified job) will be continuous. For that reason, we will be running a probit model for the first two and a linear regression model for the last one.

For the first two, the model will be identical, just varying the dependent variable:

- $P(\text{Part\_time}=1 \mid X) = \Phi(\beta_0 + \beta_1 \text{ Gender} + \beta_2 \text{ Age} + \beta_3 \text{ Experience} + \beta_4 \text{ GPA} + \beta_5 \text{ edu} + \beta_6 \text{ Sector})$

- $P(\text{Overeducation}=1 \mid X) = \Phi(\beta_0 + \beta_1 \text{ Gender} + \beta_2 \text{ Age} + \beta_3 \text{ Experience} + \beta_4 \text{ GPA} + \beta_5 \text{ edu} + \beta_6 \text{ Sector})$

And for the third one, we propose the following linear model:

- $\text{daystohighqcontr} = \beta_0 + \beta_1 \text{ Gender} + \beta_2 \text{ Age} + \beta_3 \text{ Age2} + \beta_4 \text{ Experience} + \beta_5 \text{ GPA} + \beta_6 \text{ edu} + \beta_7 \text{ Sector} + \varepsilon i$

It is important to say that we have a big database with over 151.000 observations. This fact may cause some problems in the estimation of the effects of each explanatory variable that we have on the models.

Also, previous information that we want to explain is that we will be using the statistical software Stata for running all the models.

Chapter 4: Empirical Work

During this part of the work, we will explain all the steps we followed during the empirical part. We will follow the next process:

1. Descriptive statistics and previous tests for the explanatory variables
2. Part-time model (probit)
3. Overeducation model (probit)
4. Days to high-qualified contract (linear regression)

4.1. Descriptive statistics and previous tests

First of all, we decided to analyse the main descriptive statistics of each variable (the number of observations, the mean, the standard deviation, the minimum, and the maximum) depending on whether the individual is a man or a woman. We will first run the descriptive statistics of all the dependent variables and all the explanatory variables, but without specifying the categories of the variables education and sector. After that, we will run the descriptive statistics separately for all the education and sector categories.

Table 3: Descriptive statistics by gender

Men	mean	sd	min	max	N
part time	.375	.484	0.000	1	43436
overeducation	.627	.484	0.000	1	43436
daystohighqcontr	494.718	630.705	2.000	4470	2199
age	24.901	5.298	6.409	64.679	43436
experience	15.288	20.615	1.000	186	43436
GPA	6.886	.716	5.000	11.296	43325
edu	3.846	1.195	1.000	6	43436
sector	12.357	5.371	1.000	20	43396

Women	mean	sd	min	max	N
part time	.493	.5	0.000	1	107847
overeducation	.561	.496	0.000	1	107847
daystohighqcontr	442.238	614.385	2.000	4492	4774
age	24.823	5.206	2.513	62.283	107847
experience	17.696	21.773	1.000	187	107847
GPA	7.184	.705	5.000	10.904	107810
edu	4.095	1.437	1.000	6	107847
sector	14.023	4.646	1.000	20	107768

Women (edu)	mean	sd	min	max	N
Engineering	.236	.425	0.000	1	14071
Health	.818	.386	0.000	1	43729
Humanities	.733	.442	0.000	1	7868
Natural Sciences	.651	.477	0.000	1	6137
Social Sciences	.742	.437	0.000	1	79471

Men (edu)	mean	sd	min	max	N
Engineering	.764	.425	0.000	1	14071
Health	.182	.386	0.000	1	43729
Humanities	.267	.442	0.000	1	7868
Natural Sciences	.349	.477	0.000	1	6137
Social Sciences	.258	.437	0.000	1	79471

	Women	Men
Sector	mean	mean
Primary Sector	.495	.504
Extractive industries	.514	.485
Manufacturing	.472	.528
Energy, gas, water and waste management	.455	.545
Construction	.387	.613
Trade and vehicle repair	.756	.244
Transport and storage	.557	.443
Hospitality and food services	.686	.314
Information and communication	.504	.496
Finance and insurance	.689	.311
Real estate activities	.722	.278
Professional and scientific services	.706	.294
Administrative and support services	.685	.315
Public administration	.801	.199
Education	.817	.183
Health and social services	.827	.173
Arts and entertainment	.643	.357
Other services	.764	.236

Also, we ran the frequencies of our three principal dummy variables (women, overeducation, and part-time). We did not run the frequencies of the days to a high-qualified job since it would be impossible to draw any conclusion from it because it is a continuous variable:

Table 4: Frequencies of Variable part_time by gender

	part_time		
women	**0**	**1**	**Total**
0	27,161	16,275	43,436
1	54,642	53,205	10,7847
Total	81,803	69,480	151,283

Table 5: Frequencies of Variable Overeducation by gender

	overeducation		
women	**0**	**1**	**Total**
0	16,214	27,222	43,436
1	47,395	60,452	10,7847
Total	63,609	87,674	151,283

Now, we will run a multicollinearity test to check if there is any multicollinearity problem between our data:

Table 6: VIF test for part-time and overeducation models

	VIF	1/VIF
women	1.252	.799
age	1.103	.907
experience	1.102	.907
GPA	1.203	.831
3.edu	4.961	.202
4.edu	3.075	.325
5.edu	1.829	.547
6.edu	5.013	.199
2.sector	1.134	.882
3.sector	30.127	.033
6.sector	3.336	.3
7.sector	6.931	.144
8.sector	20.534	.049

9.sector	3.614	.277
10.sector	9.698	.103
11.sector	23.854	.042
12.sector	4.551	.22
13.sector	2.615	.382
14.sector	45.368	.022
15.sector	19.283	.052
16.sector	27.689	.036
17.sector	80.828	.012
18.sector	92.616	.011
19.sector	28.75	.035
20.sector	15.086	.066
Mean VIF	17.422	.

Table 7: VIF test for days to highly qualified job model

	VIF	1/VIF
women	1.143	.875
age	47.775	.021
experience	46.79	.021
GPA	1.223	.818
3.edu	1.187	.842
4.edu	1288.217	.001
5.edu	8287.909	0
6.edu	2217.9	0
2.sector	1752.68	.001
3.sector	7654.609	0
6.sector	1.068	.936
7.sector	24.363	.041
8.sector	2.284	.438
9.sector	4.429	.226
10.sector	29.978	.033
11.sector	7.562	.132
12.sector	25.472	.039
13.sector	11.155	.09
14.sector	9.15	.109
15.sector	2.646	.378
16.sector	27.651	.036
17.sector	25.342	.039
18.sector	21.239	.047
19.sector	47.756	.021
20.sector	88.345	.011
Mean VIF	22.499	.044

As we can see, we ran a VIF test to check for any problems. Despite this test being possible only to run it with a linear regression model and not a probit, since we are just looking at the collinearity between our variables, there is no difference between the part-time and overeducation models with the time before getting a highly qualified job model.

When we look at the values, we can observe that there does not exist a collinearity problem on the simple variables (women, age, experience, and GPA), but that when we look at the variables edu and sector depending on the category that we look at, we may see pretty high collinearity. As we know, a low collinearity level at the VIF test is any number shorter than 5, a level between 5 and 10 is a moderate collinearity, and a level of 10 or more is considered a high collinearity.

Despite some dummy variables inside our variables edu and sector (especially sector, since the higher level of collinearity at edu is 7.50, and we will consider it acceptable) having pretty high values of collinearity, this is a normal thing since we are running a model with a lot of dummy variables, which are highly correlated with the others of the same variable.

We can observe that we will not have a "serious" problem with collinearity if we run the same test, but with the variables edu and sector compacted:

Table 8: VIF test without categories

	VIF	1/VIF
sector	1.182	.846
GPA	1.133	.882
women	1.092	.915
age	1.085	.922
experience	1.08	.926
edu	1.072	.933
Mean VIF	1.107	.

	VIF	1/VIF
sector	1.15	.869
experience	1.149	.87
edu	1.134	.882
age	1.126	.888
GPA	1.113	.899
women	1.053	.95
Mean VIF	1.121	.

Finally, we will run a correlation matrix between our explanatory variables to check for any correlation that they may have:

Table 9: Correlation matrix

Variables	(1)	(2)	(3)	(4)	(5)
(1) women	1.000				
(2) age	-0.007	1.000			
(3) experience	0.051	0.402	1.000		
(4) edu	0.082	0.099	0.161	1.000	
(5) sector	0.153	0.128	0.141	0.188	1.000

As we can observe, there are no correlation problems between the explanatory variables.

4.2. Econometric Models

Table 10: Summary of Econometric models

Variable	Model 1 part_time	Model 2 overeducation	Model 3 daystohighqcontr
women	0.185***	0.114***	33.90**
age	-0.00811***	-0.0140***	256.7***
experience	0.00219***	0.00105***	-6.744***
NOTA_10	0.0584***	-0.167***	-35.89***
Natural Sciences	0	0	0
Social Sciences	0.183***	0.572***	-11.94
Engineering	-0.189***	-0.333***	-134.1***
Humanities	0.308***	0.679***	55.98
Health	0.231***	0.0635*	83.82***
Primary Sector	0	0	0
Extractive industries	0.842**	0.438	-168.7

Manufacturing	-0.249**	-0.124	-65.01
Energy, gas, water and waste management	0.175	-0.889***	-65.73
Construction	0.230*	-0.341***	-158.3
Trade and vehicle repair	0.741***	0.255**	-159.9
Transport and storage	0.266**	0.584***	242.6
Hospitality and food services	0.989***	0.243**	-273.9
Information and communication	-0.0908	-1.598***	-118.7
Finance and insurance	0.0144	-0.162	-137.2
Real estate activities	0.395***	-0.0597	-193.0
Professional and scientific services	0.657***	-0.663***	-98.88
Administrative and support services	0.789***	-0.306***	-92.27
Public administration	0.250**	-1.616***	-267.2
Education	1.296***	-2.183***	-197.8
Health and social services	0.486***	-2.043***	-292.0*
Arts and entertainment	1.113***	-0.917***	-234.8
Other services	0.980***	-1.326***	-232.3
age2			-2.998***
Constant	-1.276***	1.830***	-3630.2***

Pseudo R2/ R-squared	0.0926	0.3703	0.2779
Chi2	0.0000	0.0000	
N Observations	66447	66447	6953

4.3. Part-time Model

The model will have the following form:

- $P\ (Part\_time=1 \mid X) = \Phi\ (\beta_0 + \beta_1\ Gender + \beta_2\ Age + \beta_3\ Experience + \beta_4\ GPA + \beta_5\ edu + \beta_6\ Sector)$

And we will be using the next command in Stata to run it:

probit part_time women age experience GPA i.edu i.sector if edu !=1 & aftergradu==1, vce(robust)

As you can observe, we will omit directly the category edu 1 ***(if edu !=1)***, since it represents the category "None" and with the low quantity of observations that it has (7 of 151.283), it gives us some problems running it.

Also, we will use the variable aftergradu ***(aftergradu==1)***, which indicates 1 if the job position is after graduation and 0 if it is before, to run the model only with the job positions once the individual has finished their degree. This reduces the number of observations that the model is using directly (66.552 job positions are after graduation, i.e., 43.99%), but anyway, we used all the variables since, for example, for the variable "experience" we count also the job positions that the individual has had before graduation.

Finally, we anticipated the possibility of being obligated to reject the hypothesis of homoskedasticity, which makes us assume that the variance of the errors will be constant (something typical in probit models), and accept the presence of heteroskedasticity in the model by calculating the robust standard errors ***(vce(robust))***.

When we first run it, we can observe this:

Table 11: Part-time model

part_time	Coef.	St.Err.	t-value	Sig
women	.185	.013	14.72	***
age	-.008	.001	-7.40	***
experience	-.002	0	-9.80	***
GPA	.058	.008	7.33	***
edu				
base Natural Sc~s	0	.	.	
Social Sciences	.183	.029	6.41	***
Engineering	-.189	.035	-5.41	***
Humanities	.308	.035	8.78	***
Health	.231	.03	7.61	***
sector				
base Primary Se~r	0	.	.	
Extractive industr~s	.842	.408	2.06	**
Manufacturing	-.249	.118	-2.12	**
Energy, gas, water~n	.175	.151	1.16	
Construction	.23	.129	1.78	*
Trade and vehicle ~r	.741	.116	6.38	***
Transport and stor~e	.266	.122	2.18	**
Hospitality and fo~s	.989	.117	8.48	***
Information and co~n	-.091	.121	-0.75	
Finance and insura~e	.014	.121	0.12	
Real estate activi~s	.395	.139	2.84	***
Professional and s~v	.657	.116	5.66	***
Administrative and~i	.789	.117	6.77	***
Public administrat~n	.25	.117	2.13	**
Education	1.296	.116	11.19	***
Health and social ~s	.486	.116	4.20	***
Arts and entertain~t	1.113	.117	9.53	***
Other services	.98	.119	8.21	***
Constant	-1.276	.134	-9.54	***

Mean dependent var	0.436	SD dependent var	0.496
Pseudo r-squared	0.093	Number of obs	66447
Chi-square	7709.302	Prob > chi2	0.000
Akaike crit. (AIC)	82656.818	Bayesian crit. (BIC)	82893.526

*** *p<.01, *** *p<.05, ** *p<.1*

When you run a probit model, unlike when you run a linear regression one, you can't observe directly from the coefficients the effect that the explanatory variables will have on the dependent one. Anyway, you can already observe the significance and the direction of the effect in each explanatory variable.

Another important fact to know is that, as we are using the command i. to run the variables "edu" and "sector" in the model as categorical, Stata deletes one category of each one to us it as a reference for the other ones. Natural Sciences at edu and Primary Sector at sector.

Now, if we look at the chi-squared statistic, which shows us the level of significance of the model, we can see how the model in general is also significant (Chi2 = 0.0000).

When we observe the significance level of each variable (P-value), we can observe how all the simple variables that we chose to isolate the effect of being a woman on the dependent (age, experience, and GPA), and also all the categories of edu are significant with a P-value of 0.000. Also, some of the categories of the variable sector (such as Trade and vehicle repair and hospitality, and food services, among others) have this level of significance.

Summarizing the results, we have 21 of 25 variables (counting all the dummy variables generated by Stata with categories of the variables edu and sector) being significant at a level of confidence of 95%, 1 variable being significant at a level of 90%, and 3 variables (Energy, gas, water and waste management; Information and communication; and Finance and insurance) that are not significant for the possibility of working part-time.

Now, if we look at the sign of the variables, the first thing that we can appreciate is that the variable women has the effect that we expected. As we said before, our variable is a dummy one that indicates 1 for women and 0 for men. With this information, we know that being a woman will increase the probability of working part-time since the variable has a positive effect on the dependent one.

If we look at the rest of the explanatory variables, we observe how most of them have a positive effect on the variable, but we will focus on the variables age, experience, and

GPA, which are the ones that we concentrated more on which sign we expected them to have.

As we expected, the variables age and experience have a negative effect, helping to reduce the possibility of working at a part-time job as they increase. Surprisingly, the variable GPA has a positive effect on the dependent variable. That could be explained since we are considering a part-time job, any job position where the individual works less than 35 hours per week. This fact could cause a mix between job positions that are part-time and job positions that have better conditions.

To check that idea, we ran the model again, decreasing the hours that the individual has to work to be considered part-time. Surprisingly, the effect of the variable did not change till the part-time job was considered as working less than 10 hours, a moment where the variable stopped being significant. You can find all those tests in the appendix.

It also exists the possibility that the reason behind GPA increasing the part-time job possibilities is due to the number of students who are doing a Master's and working at the same time. Due to the relationship between GPA and studying for a master's degree.

Now, continuing with the study of the model, if we want to observe the exact effect that the variables that we chose have on the probability of working part-time, we must check the marginals:

Table 12: Marginals of part-time model

	dy/dx	Std.Err.	P>z
women	0.066	0.004	0.000
age	-0.003	0.000	0.000
experience	-0.001	0.000	0.000
GPA	0.021	0.003	0.000
edu			
Social Sciences	0.065	0.010	0.000
Engineering	-0.063	0.012	0.000
Humanities	0.110	0.012	0.000
Health	0.082	0.011	0.000
sector			
Extractive industries	0.299	0.157	0.057
Manufacturing	-0.063	0.033	0.055
Energy, gas, water and waste management	0.053	0.045	0.243

Construction	0.071	0.038	0.062
Trade and vehicle repair	0.259	0.033	0.000
Transport and storage	0.083	0.035	0.019
Hospitality and food services	0.356	0.033	0.000
Information and communication	-0.025	0.034	0.466
Finance and insurance	0.004	0.034	0.905
Real estate activities	0.128	0.043	0.003
Professional and scientific services	0.227	0.033	0.000
Administrative and support services	0.278	0.033	0.000
Public administration	0.078	0.033	0.021
Education	0.470	0.033	0.000
Health and social services	0.161	0.033	0.000
Arts and entertainment	0.404	0.034	0.000
Other services	0.353	0.035	0.000

Now we can observe how, according to our model, being a woman increases the probability of working part-time by 6.5 percentage points.

Also, if we look at the rest of the variables, we can observe that the effect of the variables age and, especially, experience, is considerably low (0.29 and 0.078 percentage points, respectively).

Finally, if we look at how the different academic paths and labor sectors increase or decrease the possibility of working part-time, we find out that: studying Humanities causes the biggest increase in the dependent variable with a positive effect of 10.9 percentage points, while studying engineering causes a decrease of 6.3 percentage points.

Then, if we look at the labor sectors, we see how some of them even manage to increase the dependent variable by 47.01 percentage points (Education). One interesting fact about the results is that among all the significant categories of the variable sector, just the Manufacturing sector causes a decrease in the possibility of working part-time (-6.34)

With all this information, knowing that 21 of the 25 variables that we chose to control any variability that we could not observe, so we could study the effect of being a women on the possibility of working part-time, are significant with a confidence level of 95% (including the gender one that we insist, is the main focus of this research), having run the model with robust standard errors, and knowing that all the model is also significant, we can conclude that being a women has a positive effect over the possibility of working part-time.

Despite that, something that we noticed is that we had a low Pseudo R2 (0.0926), which indicates to us that our model only increases the prediction of a model without explanatory variables by 9.26%. Although this is something normal in studies with dummy variables, we decided to run a link test so we can observe the specification of our model:

Table 13: Link test of the part-time model

Probit regression

Number of obs = 66,447
LR chi2(2) = 8456.45
Prob > chi2 = 0.0000

Log likelihood = -41289.684

Pseudo R2 = 0.0929

part_time	**Coef.**	**Std.Err.**	**z**	**P>z**	**[95%Conf.**	**Interval]**
_hat	0.961	0.014	68.820	0.000	0.933	0.988
_hatsq	-0.088	0.018	-5.010	0.000	-0.122	-0.054
_cons	0.013	0.006	2.200	0.028	0.001	0.025

In that test, we can observe how our model has a specification problem. Since the variable _hatsq, which should be insignificant (p-value>0.05), is significant.

Anyway, this problem is "normal" since we know that we are studying something too multifactorial and complex, which is influenced by too many different factors (individual, social, or structural), that cannot be observed, measured, and studied directly by the data that we own, since we would have to know things like the economic status of the family of each individual, or personal preferences as the literature that we presented in this study defence.

However, the model that we used presents consistent effects with the theory and previous literature, statistical robustness, and high levels of significance in all its sections. All this (including the problems of specification) is something that may happen equally in the next models that we will run.

4.4. Overeducation Model

The model will have the following form:

- $P(\text{Overeducation}=1 \mid X) = \Phi(\beta_0 + \beta_1 \text{ Gender} + \beta_2 \text{ Age} + \beta_3 \text{ Experience} + \beta_4 \text{ GPA} + \beta_5 \text{ edu} + \beta_6 \text{ Sector})$

And we will be using the next command in Stata to run it:

probit overeducation women age experience GPA i.edu i.sector if edu !=1 & aftergradu==1, vce(robust)

As you can observe, similar to the past model, we will omit directly the category edu 1 ***(if edu !=1)***, since it represents the category “None” and with the low quantity of observations that it has (7 of 151.283), it gives us some problems running it.

Also, we will use the variable aftergradu ***(aftergradu==1)***, which indicates 1 if the job position is after graduation and 0 if it is before, and we anticipated the possibility of heteroskedasticity with the robust standard errors ***(vce(robust))***.

Now that we have already run the previous model with part-time, we will run this model and comment on what we expect to happen. We will equally assume the presence of heteroskedasticity in the model and run it directly with standard robust errors:

Table 14: Overeducation model

overeducation	Coef.	St.Err.	t-value	Sig
women	.114	.015	7.73	***
age	-.014	.001	-9.90	
experience	.001	0	4.16	***
GPA	-.167	.01	-17.34	***
edu				
base Natural Sc~s	0	.	.	
Social Sciences	.572	.031	18.60	***
Engineering	-.333	.036	-9.20	***
Humanities	.679	.038	17.91	***
Health	.064	.033	1.92	*
sector				
base Primary Se~r	0	.	.	
Extractive industr~s	.438	.427	1.03	
Manufacturing	-.124	.107	-1.16	
Energy, gas, water~n	-.889	.139	-6.38	***
Construction	-.341	.119	-2.87	***
Trade and vehicle ~r	.255	.108	2.37	**
Transport and stor~e	.584	.119	4.90	***
Hospitality and fo~s	.243	.108	2.24	**
Information and co~n	-1.598	.112	-14.28	***
Finance and insura~e	-.162	.111	-1.46	
Real estate activi~s	-.06	.137	-0.43	
Professional and s~v	-.663	.107	-6.19	***
Administrative and~i	-.306	.108	-2.84	***
Public administrat~n	-1.616	.109	-14.84	***
Education	-2.183	.108	-20.26	***
Health and social ~s	-2.043	.107	-19.06	***
Arts and entertain~t	-.917	.108	-8.49	***
Other services	-1.326	.112	-11.86	***
Constant	1.83	.136	13.47	***

Mean dependent var	0.362	SD dependent var	0.481
Pseudo r-squared	0.370	Number of obs	66447
Chi-square	24288.646	Prob > chi2	0.000
Akaike crit. (AIC)	54826.343	Bayesian crit. (BIC)	55063.051

*** $p<.01$, ** $p<.05$, * $p<.1$

Table 15: Marginals of the overeducation model

	dy/dx	Std.Err.	P>z
women	0.026	0.003	0.000
age	-0.003	0.000	0.000
experience	0.000	0.000	0.000
GPA	-0.038	0.002	0.000
edu			
Social Sciences	0.135	0.007	0.000
Engineering	-0.070	0.008	0.000
Humanities	0.163	0.009	0.000
Health	0.014	0.007	0.053
sector			
Extractive industries	0.116	0.096	0.227
Manufacturing	-0.040	0.033	0.231
Energy, gas, water and waste management	-0.319	0.047	0.000
Construction	-0.115	0.038	0.003
Trade and vehicle repair	0.072	0.033	0.029
Transport and storage	0.145	0.034	0.000
Hospitality and food services	0.069	0.033	0.037
Information and communication	-0.547	0.034	0.000
Finance and insurance	-0.052	0.035	0.131
Real estate activities	-0.019	0.043	0.663
Professional and scientific services	-0.234	0.033	0.000
Administrative and support services	-0.102	0.033	0.002
Public administration	-0.552	0.033	0.000
Education	-0.661	0.033	0.000
Health and social services	-0.640	0.033	0.000
Arts and entertainment	-0.329	0.034	0.000
Other services	-0.469	0.034	0.000

With those results, we can observe how much each explanatory variable affects the dependent one, but we will first check the sign of the relationship and the level of significance of each one.

As we can see now, 20 of the 25 variables that we consider for our model are significant at a level of 95%. If we look at the significance of the model in general, we can see by the chi-squared statistic that the model is significant with a Prob > chi2 of 0.000.

Similarly to the part-time model, now being a woman increases the possibility of being overeducated (despite being just 2.6 percentage points, it is an increase anyway).

As we can observe, looking at the marginals of the simple variables, now the one that does not have the expected effect on the dependent variable is experience. Anyway, the value of that positive effect is considerably low, since the margins of the model indicate that it only increases by 0.024 percentage points.

Also, similarly to the part-time model, studying humanities (and now also social sciences), increases the dependent variable by 16.25 and 13.54 percentage points, studying engineering causes a decrease of 6.96 percentage points.

The biggest difference with the part-time model is that now we have more categories of the sector variable that have a negative impact on the dependent variable. Having sectors such as Education (with a completely different effect than in the part-time model) or health and social services, which causes a decrease in the possibility of being overeducated of more than 60.0 percentage points. Contrary, for example, the Transport and storage sector has a positive effect of 14.5 percentage points.

Now we do not have a big problem with the Pseudo R2, since now the model specifies 37.03% more than a model without variables. A level that in social sciences can be considered considerably high. Anyway, we also ran the link test. You could find the results in the appendix.

The important thing about this model is also that it shows significance both in general (model) and individual (variables), and pretty solid results comparing it to the literature.

4.5. Time to a high-qualified job Model

The model will have the following form:

- daystohighqcontr = β_0 + β_1 Gender + β_2 Age + β_3 Age2 + β_4 Experience + β_5 GPA + β_6 edu + β_7 Sector + εi

And we will be using the next command in Stata to run it:

reg daystohighqcontr women age age2 experience GPA i.edu i.sector if aftergradu==1, vce(robust)

We will not follow the same process as in the previous models. Starting with the variable edu, we have not omitted the category "None", since now, as we have valid observations, Stata can use it as a reference, and we can observe the effects of Natural Sciences also.

We will continue using the variable aftergradu ***(aftergradu==1)***, which indicates 1 if the job position is after graduation and 0 if it is before, and we anticipated the possibility of heteroskedasticity with the robust standard errors ***(vce(robust))***.

Also, we added the variable "age2" that represents the square of the age, since the age may not have a continuous effect on the dependent variable.

Finally, we will not follow the exact same process as in the other ones, since the model that we will be running is a linear regression model. We use that model instead of the probit one because it is the model that best works with continuous dependent variables:

Table 16: Days to a highly qualified job model

daystohighqcontr	Coef.	St.Err.	t-value	Sig
women	33.892	15.773	2.15	**
age	256.673	21.711	11.82	***
age2	-2.998	.362	-8.29	***
experience	-6.743	.851	-7.93	***
GPA	-35.873	10.247	-3.50	***
edu				
base None	0	.	.	
Natural Sciences	159.676	61.453	2.60	***
Social Sciences	147.733	59.565	2.48	**
Engineering	25.619	59.933	0.43	
Humanities	215.654	68.071	3.17	***
Health	243.489	61.583	3.95	***
sector				
base Primary Se~r	0	.	.	
Extractive industr~s	-168.688	161.765	-1.04	
Manufacturing	-65.01	162.626	-0.40	
Energy, gas, water~n	-65.724	188.461	-0.35	
Construction	-158.275	174.919	-0.90	
Trade and vehicle ~r	-159.909	163.296	-0.98	
Transport and stor~e	242.586	232.294	1.04	
Hospitality and fo~s	-273.895	166.967	-1.64	
Information and co~n	-118.717	162.77	-0.73	
Finance and insura~e	-137.205	182.299	-0.75	
Real estate activi~s	-192.987	199.367	-0.97	
Professional and s~v	-98.887	161.686	-0.61	
Administrative and~i	-92.269	164.63	-0.56	
Public administrat~n	-267.168	162.889	-1.64	
Education	-197.836	161.429	-1.23	
Health and social ~s	-292.059	161.535	-1.81	*
Arts and entertain~t	-234.842	162.336	-1.45	
Other services	-232.308	164.891	-1.41	
Constant	-3789.845	378.088	-10.02	***
Mean dependent var	458.363	SD dependent var	618.934	
R-squared	0.278	Number of obs	6955	
F-test	44.782	Prob > F	0.000	
Akaike crit. (AIC)	106941.255	Bayesian crit. (BIC)	107132.977	

*** *p<.01, ** p<.05, * p<.1*

Now, we previously ran a Breusch-Pagan test to check if we also needed to reject the homoskedasticity hypothesis and accept the heteroskedasticity in the model. The test gave us a chi-squared of 0.000. So, as in the previous cases, we must reject the homoskedasticity hypothesis. The results that we gave here are already with standard robust errors. You can find the test in the appendix.

In the case of a linear regression model, we can observe directly the effect of each explanatory variable on the dependent variable by looking at the coefficients of the model. If we look at our main variable (women), we can observe how being a woman increases by 33.89 days the time that the individual would have to spend before getting a highly qualified job on average.

If we look at the variable age we could think that increasing a year of age will increase the time that the individual would have to spend on more than 250 days, but as we added the variable age2 we can comprehend that the age of the individual does not have a continuous effect on the dependent variable.

Now, since age2 has a negative effect on the variables, all three basic variables have the expected sign over the dependent variable. While an extra job experience will reduce the days that the individual will have to spend on 6.74 days, an extra point on the GPA decreases it by 35.87 days.

In this case, just 9 of the 27 variables are significant at a level of 95%, since none of the categories of the variable sector are significant.

Anyway, the categories of the variable edu have a big impact. Studying health will add 243.5 days, and humanities 215.65. This time, despite having a pretty low effect compared to the others, studying engineering does not result in a significant improvement in the days that the individual would have to spend.

The F statistic indicates to us that the model in general is also significant (0.0000).

If we look at our model's R-squared, we can observe that it indicates a level of 0.2779. Meaning this that 27.79% of the variability of the dependent variable is explained by the model. This level is a little bit short but acceptable in the social sciences since they usually are multifactorial (as in the other models that we ran).

Chapter 5: Conclusions

The empirical part of this study aimed to analyse the effects of being a woman on 3 different situations that we can observe almost every day: The part-time work, the overeducation problem, and the time that a person must wait before getting a highly qualified job.

In our hypothesis, based on the literature that we read before, we expected that being a woman increases the time before getting a highly qualified job and the possibilities of working part-time and being overeducated. That is to say, since we were using a dummy variable for being a woman that indicated 1 if the individual was a woman and 0 if not, the variable woman had to have a positive relationship with all three dependent variables.

With this idea, we chose to run three different models that could help us explain which are the effects of gender on them, using variables such as age, chosen academic path, and labor experience, with the idea of isolating the effects of gender on the dependent variables.

Once we ran the models, we observed that our hypothesis was true in all three of the cases. Being a woman increased the possibilities of working part-time, being overeducated, and the time that individuals have to spend before getting a highly qualified job.

With those results, we accepted the H0 hypothesis in all three cases.

With the levels of robustness and consistency that our models showed, we can conclude that, as stated by the previous literature, being a woman increases the possibilities of working part-time and being overeducated, increasing also the time that the individuals have to spend before getting a highly qualified job. Showing a competitive disadvantage on the labor market.

All those results may make us think that the reason behind this competitive disadvantage of women on the labor market, which later creates the wage differences that the economists talked about, may have the explanation by the different personal traits that men and women have on average while facing the job search (which will end up making them finish with different results of the job search) and the academic path that they select (which later will affect them on the school-to-work transition).

6.1. Limitations and Future Research

Like all social-related problems, the gender differences in the labor market during the job search process are complex and multifactorial, influenced by a large number of individual and social factors, a lot of them even not observable.

With that, we are not trying to say that our study was not valid, but we encourage other researchers to replicate this model with other variables that we may have omitted.

References

Barone, C., & Ortiz, L. (2011). Overeducation among European University Graduates: A comparative analysis of its incidence and the importance of higher education differentiation. *Higher Education*, *61*(3). https://doi.org/10.1007/s10734-010-9380-0

Becker, G. S., Hubbard, W. H. J., & Murphy, K. M. (2010). *Explaining the Worldwide Boom in Higher Education of Women Higher Education of Women*. http://ssrn.com/abstract=1681606

Bertrand, M. (2011). New Perspectives on Gender. *Handbook of Labor Economics*, *4*, 1543–1590.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, & explanations. *Journal of Economic Literature 2017, 55(3), 789–865*, *55*(3), 789–865. https://doi.org/10.1257/jel.20160995

Boto-García, D., & Escalonilla, M. (2022). University education, mismatched jobs: are there gender differences in the drivers of overeducation? *Economia Politica*, *39*(3), 861–902. https://doi.org/10.1007/s40888-022-00270-y

Caliendo, M., Cobb-Clark, D., & Uhlendorff, A. (2010). *Locus of Control and Job Search Strategies*.

Caroleo, F. E., & Pastore, F. (2018). Overeducation at a Glance. Determinants and Wage Effects of the Educational Mismatch Based on AlmaLaurea Data. *Social Indicators Research*, *137*(3). https://doi.org/10.1007/s11205-017-1641-1

Charalambidou, C., & McIntosh, S. (2021). Over-education in Cyprus: Micro and macro determinants, persistence and state dependence. A dynamic panel analysis. *Manchester School*, *89*(2). https://doi.org/10.1111/manc.12357

Cortés, P., Pan, J., Pilossoph, L., Reuben, E., & Zafar, B. (2023). Gender Differences in Job Search and the Earnings Gap: Evidence from the Field and Lab. *Quarterly Journal of Economics*, *138*(4), 2069–2126. https://doi.org/10.1093/qje/qjad017

Cortés, P., French, J., Pan, J., & Zafar, B. (2024). *Gender Differences in Negotiations and Labor Market Outcomes: Evidence from an Information Intervention with College Students **.

Devillanova, C. (2013). Over-education and spatial flexibility: New evidence from Italian survey data. *Papers in Regional Science*, *92*(3). https://doi.org/10.1111/j.1435-5957.2012.00434.x

Farré, L., Jofre-Monseny, J., & Torrecillas, J. (2022). *COMMUTING TIME AND THE GENDER GAP IN LABOR MARKET PARTICIPATION*. http://www.ieb.ub.edu

Ferri, V., García-Pereiro, T., & Pace, R. (2023). Gender pay-gap: exploring the school-to-work transition of graduates in Italy. *International Journal of Manpower*, *44*(6), 1143–1167. https://doi.org/10.1108/IJM-06-2021-0363

Flinn, C., Todd, P., & Zhang, W. (2019). *Personality Traits, Job Search and the Gender Wage Gap*.

Francesconi, M. (2003). A joint dynamic model of fertility and work of married women. In *Journal of Labor Economics* (Vol. 20, Issue 2). https://doi.org/10.1086/338220

Frank, R. H. (1978). Why women earn less: The theory and estimation of differential over qualification. *American Economic Review*, *68*(3).

Gimenez-Nadal, J. I., & Molina, J. A. (2014). Commuting time and labour supply in the Netherlands: A time use study. *Journal of Transport Economics and Policy*, *48*.

Graham, M., & Gerhart, B. (2000). *Discrimination by Parts: A Fixed-Effects Analysis of Starting Pay Differences across Gender*. https://www.researchgate.net/publication/5220668

Hansen, B., Mcnichols, D., Agan, A., Black, S., Cowgill, B., Davis, J., Doleac, J., Exley, C., Goldin, C., Horton, J., Kuhn, M., Minard, S., Niederle, M., Nix, E., Rubin, E., Sabia, J., Sarsons, H., Stoddard, O., & Waddell, G. (2020). *We are grateful to*. http://www.nber.org/papers/w27054

Hirsch, B. T. (2005). Why do part-time workers earn less? The role of worker and job skills. *Industrial and Labor Relations Review*, *58*(4). https://doi.org/10.1177/001979390505800401

Kleven, H., Landais, C., Posch, J., Steinhauer, A., & Zweimüller, J. (2019). *Countries: Evidence and Explanations*. http://www.nber.org/papers/w25524

Le Barbanchon, T., Rathelot, R., & Roulet, A. (2021). Gender Differences in Job Search: Trading off Commute against Wage. In *Quarterly Journal of Economics* (Vol. 136, Issue 1, pp. 381–426). Oxford University Press. https://doi.org/10.1093/qje/qjaa033

Mcgoldrick, K., & Robst, J. (1996). Gender Differences in Overeducation: A Test of the Theory of Differential Overqualification. *American Economic Review*, *86*(2).

Olivetti, C., Pan, J., & Petrongolo, B. (2024). *The Evolution of Gender in the Labor Market*. http://www.nber.org/papers/w33153

Rachel, C., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature 2009, 47:2, 1–27*.

Rudakov, V., Kiryushina, M., Figueiredo, H., & Teixeira, P. N. (2023). Early career gender wage gaps among university graduates in Russia. *International Journal of Manpower*, *44*(6), 1046–1070. https://doi.org/10.1108/IJM-03-2021-0206

Sandner, M. ;, & Yükselen, I. (2024). *Unraveling the Gender Wage Gap: Exploring Early Career Patterns among University Graduates*. https://hdl.handle.net/10419/305735

Zafar, B., Bernal, R., Brown, M., Delavande, A., Hinds, M., Jones, B., Lieb, H., Linsenmeier, J., Madeira, C., Malamud, O., Ross, S., Trapeznikova, I., Urzua, S., &

van der Klaauw, W. (2009). *Federal Reserve Bank of New York Staff Reports College Major Choice and the Gender Gap*.

Appendix

Tests of the Part-time model:

Table 1: Part-time model considering part-time less than 30 hours worked

```
Probit regression                               Number of obs     =     66,447
                                                Wald chi2(25)     =    7794.71
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -40176.007               Pseudo R2         =     0.0958
```

part_time2	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
women	.1840848	.0127771	14.41	0.000	.1590422	.2091273
age	-.0068041	.0011101	-6.13	0.000	-.0089797	-.0046284
experience	-.0025096	.0002293	-10.94	0.000	-.002959	-.0020602
GPA	.0454807	.0080268	5.67	0.000	.0297484	.0612129
edu						
Social Sciences	.2018891	.029204	6.91	0.000	.1446504	.2591278
Engineering	-.1615505	.0357185	-4.52	0.000	-.2315575	-.0915436
Humanities	.3106473	.0354446	8.76	0.000	.2411771	.3801174
Health	.230903	.0309584	7.46	0.000	.1702256	.2915803
sector						
Extractive industries	.9332972	.4087993	2.28	0.022	.1320653	1.734529
Manufacturing	-.2380119	.1220216	-1.95	0.051	-.4771698	.0011461
Energy, gas, water and waste managem..	.1170483	.1584784	0.74	0.460	-.1935637	.4276604
Construction	.2518345	.1335018	1.89	0.059	-.0098242	.5134932
Trade and vehicle repair	.6606869	.1204264	5.49	0.000	.4246555	.8967183
Transport and storage	.2790698	.1263547	2.21	0.027	.0314192	.5267204
Hospitality and food services	.9654659	.1207481	8.00	0.000	.728804	1.202128
Information and communication	-.0594153	.1250649	-0.48	0.635	-.3045379	.1857073
Finance and insurance	-.2836523	.1274279	-2.23	0.026	-.5334063	-.0338982
Real estate activities	.4202517	.1433548	2.93	0.003	.1392814	.7012219
Professional and scientific services	.6514775	.1204766	5.41	0.000	.4153478	.8876072
Administrative and support services	.7486066	.1207922	6.20	0.000	.5118583	.9853548
Public administration	.2824979	.1214501	2.33	0.020	.0444601	.5205357
Education	1.329545	.1200065	11.08	0.000	1.094336	1.564753
Health and social services	.4486012	.1199548	3.74	0.000	.213494	.6837083
Arts and entertainment	1.095217	.1209661	9.05	0.000	.8581278	1.332306
Other services	.9524969	.1234907	7.71	0.000	.7104596	1.194534
_cons	-1.331658	.1378785	-9.66	0.000	-1.601895	-1.061421

Table 2: Part-time model considering part-time less than 20 hours worked

```
Probit regression                                   Number of obs     =     66,436
                                                    Wald chi2(24)     =    7509.77
                                                    Prob > chi2       =     0.0000
Log pseudolikelihood = -31422.662                   Pseudo R2         =     0.1158
```

part_time5	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
women	.1205286	.0143716	8.39	0.000	.0923608	.1486965
age	-.0026828	.0012477	-2.15	0.032	-.0051282	-.0002373
experience	-.0024909	.0002729	-9.13	0.000	-.0030259	-.001956
GPA	.0468818	.0088657	5.29	0.000	.0295054	.0642581
edu						
Social Sciences	.1060254	.0331392	3.20	0.001	.0410738	.1709769
Engineering	-.118308	.0411823	-2.87	0.004	-.1990237	-.0375922
Humanities	.2350854	.0391149	6.01	0.000	.1584216	.3117492
Health	.0979422	.0351812	2.78	0.005	.0289883	.1668961
sector						
Extractive industries	0	(empty)				
Manufacturing	-.0474233	.1775634	-0.27	0.789	-.3954412	.3005946
Energy, gas, water and waste managem..	-.0004894	.2354075	-0.00	0.998	-.4618796	.4609008
Construction	.104251	.1955517	0.53	0.594	-.2790233	.4875254
Trade and vehicle repair	.6661892	.1751217	3.80	0.000	.3229569	1.009422
Transport and storage	.3566339	.1817413	1.96	0.050	.0004274	.7128403
Hospitality and food services	1.206649	.1750751	6.89	0.000	.8635083	1.54979
Information and communication	.1528867	.1805737	0.85	0.397	-.2010311	.5068046
Finance and insurance	-.0418946	.1844705	-0.23	0.820	-.4034501	.3196609
Real estate activities	.6354609	.1971601	3.22	0.001	.2490342	1.021888
Professional and scientific services	.8899105	.1749465	5.09	0.000	.5470216	1.232799
Administrative and support services	.7343636	.175451	4.19	0.000	.390486	1.078241
Public administration	.389338	.1764041	2.21	0.027	.0435924	.7350837
Education	1.665016	.1744465	9.54	0.000	1.323107	2.006925
Health and social services	.7160296	.1746318	4.10	0.000	.3737576	1.058302
Arts and entertainment	1.455758	.1751398	8.31	0.000	1.11249	1.799025
Other services	1.244434	.1770526	7.03	0.000	.8974176	1.591451
_cons	-2.061525	.1897456	-10.86	0.000	-2.433419	-1.68963

Table 3: Part-time model considering part-time less than 20 hours worked

```
Probit regression                               Number of obs     =     66,436
                                                Wald chi2(24)     =    4762.25
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -19329.384               Pseudo R2         =     0.1196
```

part_time4	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
women	.0440951	.0176963	2.49	0.013	.009411	.0787793
age	.0040993	.0015093	2.72	0.007	.0011412	.0070574
experience	-.0016114	.0003608	-4.47	0.000	-.0023184	-.0009043
GPA	-.0069312	.0108177	-0.64	0.522	-.0281335	.014271
edu						
Social Sciences	.0752701	.0399377	1.88	0.059	-.0030064	.1535466
Engineering	-.0184595	.0496729	-0.37	0.710	-.1158166	.0788976
Humanities	.2561115	.0461917	5.54	0.000	.1655774	.3466455
Health	.1506229	.0425757	3.54	0.000	.0671761	.2340697
sector						
Extractive industries	0	(empty)				
Manufacturing	.03681	.2767671	0.13	0.894	-.5056435	.5792635
Energy, gas, water and waste managem..	.0527501	.3512484	0.15	0.881	-.635684	.7411843
Construction	.3100077	.2929599	1.06	0.290	-.2641832	.8841986
Trade and vehicle repair	.6159747	.273063	2.26	0.024	.0807811	1.151168
Transport and storage	.5883546	.2787949	2.11	0.035	.0419266	1.134783
Hospitality and food services	1.127801	.2727054	4.14	0.000	.5933083	1.662294
Information and communication	.2028212	.2802286	0.72	0.469	-.3464168	.7520593
Finance and insurance	-.243933	.2966483	-0.82	0.411	-.825353	.3374869
Real estate activities	.8739942	.2915556	3.00	0.003	.3025557	1.445433
Professional and scientific services	.9665419	.2725448	3.55	0.000	.4323639	1.50072
Administrative and support services	.7954768	.2731228	2.91	0.004	.2601659	1.330788
Public administration	.3743967	.2748135	1.36	0.173	-.1642278	.9130213
Education	1.607237	.272051	5.91	0.000	1.074027	2.140447
Health and social services	.5761437	.2724012	2.12	0.034	.0422471	1.11004
Arts and entertainment	1.551285	.2725669	5.69	0.000	1.017063	2.085506
Other services	1.405064	.274022	5.13	0.000	.8679907	1.942137
_cons	-2.379884	.2872188	-8.29	0.000	-2.942822	-1.816945

Tests of the overeducation model

Table 4: Overeducation model link test

```
Probit regression                               Number of obs     =     66,447
                                                LR chi2(2)        =   32216.94
                                                Prob > chi2       =     0.0000
Log likelihood = -27384.065                     Pseudo R2         =     0.3704
```

overeducation	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_hat	1.01062	.0077194	130.92	0.000	.9954903	1.02575
_hatsq	.0189814	.0076157	2.49	0.013	.0040548	.033908
_cons	-.015521	.0088688	-1.75	0.080	-.0329036	.0018616

Tests of the days to a highly qualified job model

Table 5: Breusch-Pagan test

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of daystohighqcontr

         chi2(1)      =  5054.99
         Prob > chi2  =   0.0000
```