UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE AUGUSTIN COURNOT

Bureau d'Économie Théorique et Appliquée (UMR 7522)

THÈSE

présentée par :

Bruno André RODRIGUES COELHO

soutenue le : **12 décembre 2018**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/ Spécialité : Sciences-Économiques

Essais sur l'offre de travail et la fécondité des femmes

THÈSE dirigée par :

Monsieur KOEBEL Bertrand	Professeur, Université de Strasbourg
Monsieur LAISNEY François	Professeur, Université de Strasbourg

RAPPORTEURS :

Madame MEURS Dominique	Professeur, Université Paris - Nanterre
Monsieur LANOT Gauthier	Professeur, Umea School of Business, Umea University

AUTRES MEMBRES DU JURY :

Monsieur LEFEBVRE Mathieu	Professeur, Université de Strasbourg

Contents

List of Figures

List of Tables

Remerciements

Cette thèse est le fruit de la contribution de plusieurs personnes.

Pour commencer, je tiens à remercier Bertrand Koebel et François Laisney qui ont accepté d'encadrer cette thèse. J'ai énormément appris avec eux, aussi bien techniquement que humainement. Leur porte était toujours ouverte et eux toujours disponibles pour répondre à mes questions. Je suis maintenant dans une position où l'on frappe à ma porte et c'est à mon tour de donner de mon temps. Je pense que si j'avais été encadré par d'autres personnes, moins disponibles, je ferais moi-même moins d'efforts pour être disponible à mon tour.
Durant la rédaction des différents chapitres qui composent cette thèse, j'ai aussi été amené à rencontrer et travailler avec d'autres brillants chercheurs. Je remercie Hans-Martin von Gaudecker et Holger Stichnoth, avec qui j'ai pu travailler et échanger, mais aussi Mathieu Lefebvre et Ralf Wilke pour leurs commentaires qui ont permis de grandement améliorer les travaux qui constituent cette thèse.
Je serai toujours reconnaissant à Chiara Peroni et aux collègues de STATEC Research, pour l'opportunité qu'ils m'ont accordée de pouvoir travailler avec eux. Ce fût ma première expérience professionnelle après l'Université et j'en garde un très bon souvenir.
Je remercie aussi les chercheurs du LIS Cross-National Data Center au Luxembourg pour leur aide précieuse quand je travaillais avec leurs données lorsque j'étais assistant de recherche à STATEC Research.

Je remercie aussi Dominique Meurs et Gauthier Lanot d'avoir accepté d'évaluer cette thèse.

Je voudrais aussi remercier tous les contributeurs aux logiciels libres qui ont rendu possible les travaux qui constituent cette thèse. Sans leur travail bénévole pour faire avancer la science, rien de tout cela n'aurait été possible.

Une des autres sorties de cette thèse sont toutes les amitiés que j'ai pu tisser durant ces années passées à travailler aux côtés des autres thésards du BETA Strasbourg. J'ai la chance de pouvoir compter parmi mes amis Frédéric Olland, Audrey-Rose Menard, Lionel Rischmann, Emre Özel, Frédo Jean-Baptiste, Nguyen To The, Thomas Coudert, Guillaume Schonenberger, Marlène Le Kim, Cyrielle Poiraud, Benjamin Ouvrard, Nicolas

Lampach, Rémy Guichardaz... et aussi Imad Ismail, avec qui j'ai passé beaucoup de soirées à discuter recherches et autres sujets moins importants. Je suis très heureux de pouvoir compter Imad et Sophie parmi mes amis les plus proches.
Je suis aussi très chanceux d'avoir pu co-écrire avec mon ami Vincent Vergnat, et je suis heureux qu'il continue sa carrière de chercheur au Luxembourg.

Ensuite, il y a les membres de ma famille qui n'ont jamais cessé de m'encourager à poursuivre mes études. Mes parents, mes frères, m'ont toujours soutenu et sans l'éducation de mes parents je ne serai pas là où je suis aujourd'hui. Je pense aussi à mon grand-père qui n'a jamais cessé d'être curieux tout au long de sa vie et qui malgré sa vie difficile ne s'est jamais plaint de rien et a toujours été reconnaissant de ce qu'il avait.

Finalement, je remercie ma tendre épouse Stéphanie pour son soutien inconditionnel au quotidien.

Chapter 1

Introduction

1.1 Short description of the different chapters.

This thesis deals with labour supply and childbearing decisions of women in France. Each chapter corresponds to a research paper; each of these research papers contains a detailed literature review. This introduction will discuss the different data sources that were used in detail to make this thesis possible. Indeed, a very important element of empirical work is obviously data availability. The previous sentence seems self-evident, however, access to data is still a major hurdle that empiricists have to overcome. This introductory chapter will discuss systems that allow researchers to access data and then present the datasets used in most of the papers in greater detail. The last paper of this thesis discusses the notion of version control by reviewing and presenting Git. In the last part of the introduction, I will explain what this tool called Git is and why I discuss this topic in this thesis.

1.2 Access to data for empirical studies in economics

Economists, and other social scientists, face a major hurdle not encountered in other disciplines such as the life sciences: access to high quality data to conduct empirical studies. In life sciences and other disciplines, such as machine learning, finding data can be a complex task, but often the researchers are not confronted to administrative hurdles. For example, a data scientist interested in sentiment analysis after a political debate, can simply scrape data from Twitter, or from any other social network. A physicist interested in the behaviour of a certain particle would *only* need to observe and use tools to measure that particle. The precendent sentence is an oversimplification, of course,

as there certainly are also legal and administrative hurdles. Economists however, and mostly micro-economists, face a very practical problem concerning data: privacy issues. A labour economist interested in female labour supply for instance, would have to have data on thousands of women's wages, education levels, fertility, and other such variables, most of which might be sensitive. An economist interested in industrial organization would need to have access to accounting data on thousands (or much more) of firms. This is usually very sensitive data that people and organizations are very reluctant to hand over. This means that access to this type of data is in general regulated and only possible within a very specific research project. To further protect individuals, a new European Regulation, the *General Data Protection Regulation* which came into force in May 2018 gives much more control to individuals over their personal data. Individuals will also be able to give or deny consent to a data collector to further process their data. Collection of sensitive individual characteristics, such as those listed in Article 9 of the Regulation (*[...] racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation*) will be prohibited. This should not be a problem for research, however, as Article 9 further states that such processing of information is lawful if done *for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.*

Economic research is arguably in the public interest. The next subsections will present the data I have used for this thesis in general terms. More details are then given in each chapter.

1.2.1 The LIS data

In the first chapter I estimate a hierarchical model using French and German micro-data. Comparability issues may arise when using data, especially micro-data, from different countries. To solve this problem, I use data from the LIS[1] Cross-National Data Center located in Esch-Belval in Luxembourg. Access to the data is done remotely. Remote access is becoming more widespread, and several models have emerged. In this section I will describe the model that the LIS data center is using. In the next section I will describe the model of the CASD[2] which is located in France. The CASD makes remote access to data from the INSEE and other institutes possible. Any researcher can ask for

[1] In the past, LIS stood for Luxembourg Income Study, which is not the case anymore.
[2] CASD: *Centre d'Accès Sécurisé aux Données*, Secure Data Access Center

access to the data without needing to specify a project beforehand. There is no need to define a project beforehand because researchers that get access to the data will never actually see it. Once the researcher's demand for access is granted (which happens very fast), the researcher receives a login as well as a password. With this, the researcher can login to LISSY; LISSY is a piece of software written in JAVA through which it is possible to send computer code written in either R, STATA or SAS. The code gets executed over at the LIS' servers, and the results are sent back to the researcher, in almost real time. This is a lot faster than sending the code via email, which is still a possibility. Compared to a model such as the one of the CASD, mere days go back from the date access gets asked to when it is granted. However, once the researcher can login through the CASD, working on the data is just like if it was happening on a regular computer. The researcher can look at the data line by line, which cannot be done with LISSY. For example, the following R command would tipycall print the first six lines of the dataset:

```
head(lis_data)
```

Calling this command in LISSY, however, produces nothing. The consequence of the researcher never looking at the data makes it thus possible to avoid long and tedious procedures while protecting the privacy of citizens. The way of accessing data is not the only difference though; the LIS offers access to datasets to study the income of people in various countries, while the data accessible through the CASD is only French data.[3] The LIS team spends a lot of time and effort to make sure that the datasets are comparable, as well as imputing missing values. LIS data is thus very well suited for cross national comparisons, which is the justification of why I use it for the first chapter of this thesis.

1.2.2 The DADS-EDP data

The DADS-EDP[4] data is data on French citizens and firms that is accessible through the CASD, a remote access system that I have briefly discussed in the previous section. Access to this data requires more patience than for the LIS, but this is explained by

[3] There is however the possibility to access German data from the IAB (Institute for Employment Research) at an access point installed in the CASD premises in Palaiseau. In Nüremberg, German researchers can access the French data available on the CASD platform from a secure access point installed in the premises of the IAB.

[4] DADS stands for *Déclaration Annuelle des Données Sociales*, or Annual Declaration of Social data. EDP stands for *Échantillon Démographique Permanent*, or Permanant Demographic sample.

the fact that in the case of CASD data, the researcher has access to it almost as if the data was stored locally. This is not the case however, the data never exits the servers of the CASD, but working with data through the CASD is very comfortable. Researchers that wish to access the data need to submit a proposal that has to be approved by a committee. Once the project is approved, the researchers need to go to Paris to sign some more papers, and to have one of their fingerprint registered. The fingerprint is needed to log-in to the system. The researchers also receive a crash course in statistical disclosure control. Once this is done, researchers receive a special keyboard at their institutes. Once the keyboard is plugged to the internet and to a computer screen, the researchers can login to the system by using a keycard received at the training session and by presenting their fingerprint to the fingerprint reader. Once this is done, the researcher is logged into a Microsoft Windows virtual machine that contains the data the researcher asked access for as well as standard statistical software, such as GNU R, SAS and STATA. Data for Chapter 2 and 3 were accessed this way. This procedure might seem tedious, but is necessary to protect the privacy of citizens; the LIS Data Center opted for an easier form of remote access.

1.2.3 Version control to facilitate research

Chapter 5 is somewhat special as it does not deal with labour supply nor fertility nor economics at all for that matter. It discusses a tool called Git that was of vital importance to realise this thesis. This chapter is published as Rodrigues (2016).

Git is a tool used for software carpentry or software development. In the last chapter I discuss how research resembles Open Source software development and thus why using software development tools is useful. Git allows to track changes of the source code of a piece of software, or, in the case of a scientific article, changes made to the document. It is of course also possible to track changes to the computer code that does the analysis. Git makes collaboration extremely streamlined, and two of the papers that constitute this thesis have been co-authored with Git. The source code to this thesis is also tracked using Git and the source code is available on the following repository: `https://bitbucket.org/b-rodrigues/thesis2016/src/master/`. The source code to the papers upon which Chapters 3 and 4 are based is also available.[5]

[5]Chapter 3 paper: `https://bitbucket.org/b-rodrigues/diff-in-diff/src`, Chapter 4 paper: `https://bitbucket.org/b-rodrigues/maternity_duration/src`. Note that there might be slight differences in presentation between the these versions and the chapters of the thesis.

1.2.4 Questions that this thesis aims to answer

This thesis studies labour supply and fertility decisions of women. Each chapter deals with this topic, but from a different perspective. Chapter 2 aims to see if there are differences between French and German regions regarding total number of children women decide to have. These two countries were chosen because they are similar in many ways, but also differ greatly when in comes to women's participation to the labour market and the fertility decisions. Chapter 2 goes deeper in the analysis and considers the regional level instead of the national level. Then, Chapter 3 studies the impact of a first, second and third birth on the labour supply and wages of both women and men. The question here is to see to which extend these births impact negatively, or perhaps in the case of men even positively, labour market outcomes. Chapter 4 studies the duration of the maternity leave and how mothers decide to come back to the labour market. How long are the breaks the mothers take due to childbearing? And how do they come back? Do most women come back to full time work, or rather part time work? Chapter 3 and Chapter 4 focus on France.

Bibliography

Bruno Rodrigues. Version control systems to facilitate research collaboration in economics. *Computational Economics*, 48(3):547–553, 2016.

Chapter 2

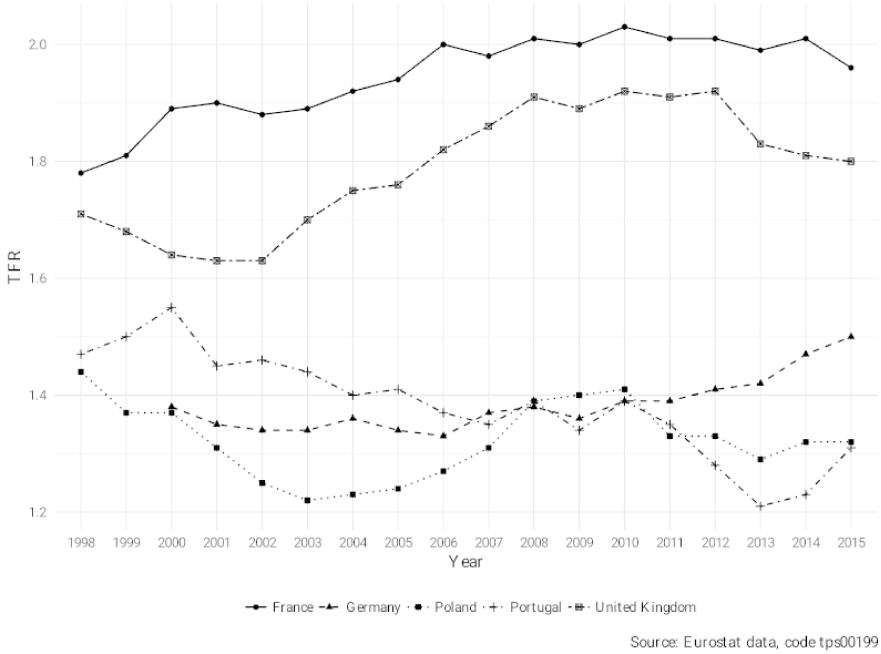
Childbearing differences between French and German women

2.1 Introduction and literature review

This chapter investigates total fertility rate (henceforth TFR) differences between European countries. Figure 2.1 shows the TFR through time for some selected European countries. This figure suggests the existence of two groups of countries regarding TFR.

I will focus on the case of France and Germany, which are the typical high and low TFR western European countries. However, I will not only look at the aggregated level of TFR, but will look at individual households (more precisely the women living in these households) within French and German regions. The goal here is to assess to what extent French and German regions differ from each other, when controlling for observable individual variables.

FIGURE 2.1: Total fertility rate in selected European countries



Goldstein et al. (2003) showed using data from the 2001 Eurobarometer survey that residents from Western and Eastern Germany, as well as from Austria, had a lower ideal family size[1] (1.7, which is below the replacement fertility rate) than France (2.5, which is above the replacement fertility rate). For both countries the actual average family size was in fact lower than the ideal family size. According to Goldstein et al. (2003), below replacement fertility rates in German speaking nations is a cultural phenomenon. This cultural phenomenon could also vary from region to region. Some regions might have "a tradition" of large families, while others do not. *Regions* are the NUTS 1 classification level; in the case of Germany, the NUTS 1 classification corresponds to the German *Bundesländer* (States) constituting the German Federation. In the case of France, the NUTS 1 classification corresponds to an arbitrary aggregation of regions. For example, the *Mediterranean* NUTS 1 unit is the aggregation of the three following regions: *Languedoc-Roussillon, Provence-Alpes-Côte d'Azur, Corsica.*[2] People living in different regions might be affected differently by macro-economic variables or by policies decided at the national level. Looking at regions is very important, as there is evidence of regional differences, for example between Eastern and Western Germany (Goldstein et al., 2012). In order to take individuals living in regions into account, I estimate

[1] By ideal family size, it is meant the family size that people wished they had.

[2] In France, these territorial units are called ZEAT, for *Zone d'études et d'aménagement du territoire* or Zone of study and territorial planning.

a hierarchical Poisson model with varying intercept at the NUTS 1 level. I choose a Poisson model because the dependent variable I study is a count variable, namely the number of children living in the household at the time of data collection.

What drives fertility? According to the literature, there are three drivers (if disregarding preferences of the couple for children): unemployment (or the larger economic environment), the education level (or wages) of mothers, and child care availability. In this chapter, I will focus on unemployment (measured at the NUTS 1 level) and education level/wages (variable measured at the individual level).

At the macroeconomic level, there is an extensive literature on how total fertility rates react to different macro-economic variables, such as unemployment, economic uncertainty or the business cycle (Butz and Ward, 1979). Adsera (2005) discusses TFR differentials in developed OECD countries and shows that contrary to microeconomic theory, which states that long term unemployment is the right moment for childbearing, due to a lower opportunity cost, childbearing is postponed until later. A similar result has been found by Comolli (2017), where in 22 out of a panel of 32 Western countries, female unemployment significantly reduced fertility rates by an average of 3%. Using US data for the period 1975-2010, Currie and Schwandt (2014) show that short-term and long-term fertility is negatively impacted by an unemployment spell experienced at ages 20-24. Sobotka et al. (2011)'s literature review gives a very good overview of the existing evidence of the effect of unemployment on fertility, as well as partnership formation. Sobotka et al. (2011) highlight the fact that fertility decreases in economic downturns, but this effect might very well depend on the women's education. For highly educated women, having a child during a recession might have a very important impact on their careers. However, for lower educated women, who struggle to find a job during a recession, parenthood is an affordable option in terms of opportunity costs. The authors also provide preliminary evidence of the effects of the Great Recession on fertility using data from EUROSTAT from the year 2010. In 2009, fertility rates decreased in 15 EU countries. Looking closer, inside a country, there is also evidence that different parts of a country might be affected differently by a recession or a sluggish economy. Vitali and Billari (2017) focus on Italian provinces, regressing the provincial TFR on control variables also measured at the regional level. Their results show that provincial fertility is negatively associated with GDP in central Italy. Moving North, this association turns positive, and moving South, this association becomes not statistically significant anymore.

There are also studies using microeconomic data sources. Ermisch (1989) estimates a microeconomic model on British data and finds that higher husband earnings increase the couple's fertility and reduces the wife's labour supply. Baizán (2009) studies the impact of accessibility to childcare of fertility in Spanish regions. In terms of data used, Baizán article is the closest to this one; the author uses the European Community Household Panel (ECHP). The author shows that an increase of childcare coverage has a positive effect on fertility. Rondinelli et al. (2010), using the Labour Force Surey (LFS) for Italy, study the effect of predicted wages on the postponement of motherhood, and find that higher predicted wages delay the first birth strongly, but less so for higher order births. The authors also find differences between Northern and Southern Italian regions, a fact they explain by better child care access in the North. Goldstein et al. (2013) use data from the Human Fertility Database (HFD)[3] as well as from the OECD and show that countries that were hit harder by the Great Recession than other countries had a stronger decrease in fertility. It is notably younger people in Southern European countries that postpone a first childbirth more than older people. A similar result was found for Central and Eastern Europe. This paper also studies the effect of the Great Recession by controlling for the growth of the unemployment rates from 2008 to 2009, but the object under study are not countries, but individual women in the NUTS 1 regions of France and Germany. Other papers have taken advantage of nested structures in data; Hank and Kreyenfeld (2003) estimate a hierarchical discrete-time logit model using individual data[4], but also using data of public day care slots in Western German districts. They do not find any significant local effect; only the socio-demographic variables such as education were significant, for example lower educated women have lower first birth risks. Klüsener et al. (2013) uses a natural experiment to disentangle the is a most important driver of childbearing decisions; institutional settings or culture? For this, the authors use the fact that the German speaking region of Belgium is culturally very similar to Germany, but the institutional setting is the Belgian one. The authors show evidence that the German-speaking minority of Belgium retained German as their mother tongue and that they are also much more exposed to German culture and media. Klüsener et al. (2013) consider the family policies that the German-speaking minority has been exposed to as being the treatment. The authors find that culture does not play a major role, but indeed, it is the institutional setting and the family policies that matter. The present study aims at contributing to this literature by considering two countries, France and Germany, as being the representatives, respectively, of high and low TFR countries.

[3] Provided by Eurostat
[4] From the German Socio-Economic Panel

TABLE 2.1: Current main activity status

Country	Status	N	Frequency
France	Mainly employed	13623	51.04
	Not mainly employed	689	2.58
	Retired	6677	25.02
	Homemaker/Care for children	2561	9.59
	Unemployment	3141	11.77
Germany	Mainly employed	10224	55.12
	Not mainly employed	106	0.57
	Retired	6237	33.62
	Homemaker/Care for children	1122	6.05
	Unemployment	820	4.42

The harmonized micro-data used makes a comparative study possible, and will help shed light on which variables correlate most with the number of children.

2.2 The LIS data: descriptive statistics and visualizations

The data used in this paper is supplied by the LIS data center in Luxembourg. LIS collects and prepares data from several countries that participate in the program. The data is then harmonized, and can thus be used for cross-national comparisons, such as the one in this paper. All the results shown are computed from the latest wave available for both France and Germany, 2010.

The French source for the harmonized data available in LIS is the household budget survey[5], containing 41285 individuals distributed among 15797 households. The German source is the German Socio-Economic Panel, with 26952 individuals distributed among 12146 households. Regional TFR rates were computed over all the female individuals, but for the purposes of this study, women enrolled in education were removed from the analysis (retired women were removed for the regression).

The following tables show the frequency of the categories of the following variables; *Current Main Activity Status* (`cmas`), *Citizenship* (`citizenship`), *Education level* (`education`), *If the person has ever worked* (`everwork`), and *Marital Status* (`marital status`).

`cmas` is a self-assessed variable. This variable gives the activity at which the surveyed spend the most time in the year. For example, someone who had been sick for some

[5] *Enquête Budge de Famille*

TABLE 2.2: Citizenship status

Country	Citizenship category	N	Frequency
France	Native	23047	86.35
	Naturalized	1159	4.34
	EU citizen	429	1.61
	African citizen	1243	4.66
	Other	813	3.05
Germany	Native	16857	90.87
	Naturalized	783	4.22
	EU citizen	491	2.65
	African citizen	17	0.09
	Other	402	2.17

TABLE 2.3: Education level

Country	Level	N	Frequency
France	Low	11475	42.99
	Medium	9604	35.98
	High	5612	21.03
Germany	Low	2689	14.50
	Medium	10244	55.22
	High	5617	30.28

months while answering the survey, but who works otherwise, would have answered `Mainly employed`. In other surveys, such as the LFS, the employment status refers to a defined period before the survey, so in our example, that person would have answered `Not working`. In Table 2.1, we see that around 51% of women in France and 55% of women in Germany are mainly employed. The next categories with the largest share is retired, with 25% of respondents in France and 34% in Germany.

Table 2.2 shows the share of "natives", naturalized nationals and immigrants from different origins. France has a higher share of immigrants, mostly from Africa, while Germany has a higher share of immigrants from the European Union. This variable is later collapsed to a new variable with two categories, `Immigrant` and `Non-immigrant` for the regression model.

In Table 2.3, the education level of French and German women is reported in three levels. *Low level* corresponds to the International Standard Classification of Education (ISCED) levels 0, 1 and 2 which correspond having achieved less than secondary education, *Medium level* to ISCED 4 or 5 for secondary education completed and *High*

TABLE 2.4: Ever worked in job or business

Country	Has ever worked?	N	Frequency
France	Has never worked	2780	10.42
	Has worked	23911	89.58
Germany	Has never worked	201	1.08
	Has worked	18349	98.92

TABLE 2.5: Marital status

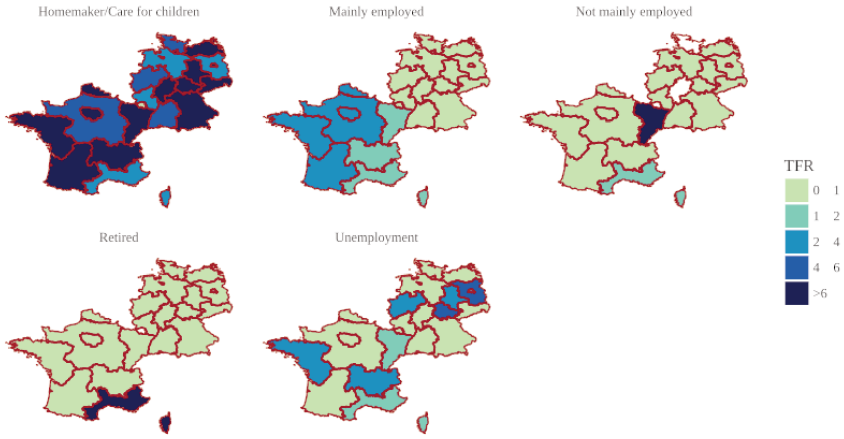
Country	Level	N	Frequency
France	Married	13165	49.32
	Never married/not in union	9551	35.78
	Separated/divorced	2251	8.43
	Widowed	1724	6.46
Germany	Married	11977	68.39
	Never married/not in union	2886	16.48
	Separated/divorced	1290	7.37
	Widowed	1359	7.76

level to ISCED 5 or 6, for high to tertiary education completed. These 3 categories are unfortunately very broad, especially when one considers that in the national surveys the education variable is much more detailed. This variable is still available in the LIS data as `educ_c`, but then comparison between the German and French `educ_c` variables is not necessarily meaningful.

Table 2.4 shows a dummy variable that reports the existence of any work experience at all during the entire life. About only 1% of German women have never worked, while this number climbs to 10% for France. Because this variable is very unbalanced for Germany, I drop it from the regression model. Finally Table 2.5 reports the share of married, non-married, separated or divorced women in France and Germany. Germany has a higher share of married women than France.

Next, we visualize these tables on maps, and show differences between French and German regions. In Figure 2.2, we can see the average fertility rates of French and German regions, by different *current main activity status*.

FIGURE 2.2: Total fertility rate by current main activity



In almost every region for every status, the fertility rate is very low for Germany. In most cases the TFR is between 0 and 1. For France, women in the status *Mainly employed* have in most regions a TFR between 2 and 4.

Figure 2.3 shows the TFR by education level. One would expect the TFR to decrease with increasing education level of the mother. Overall, this seems to be the case, but in some regions, the pattern is reversed. This is especially the case in Germany, for example in Schleswig-Holstein (the northernmost region) or in Sachsen (the easternmost).

FIGURE 2.3: Total fertility rate by education level

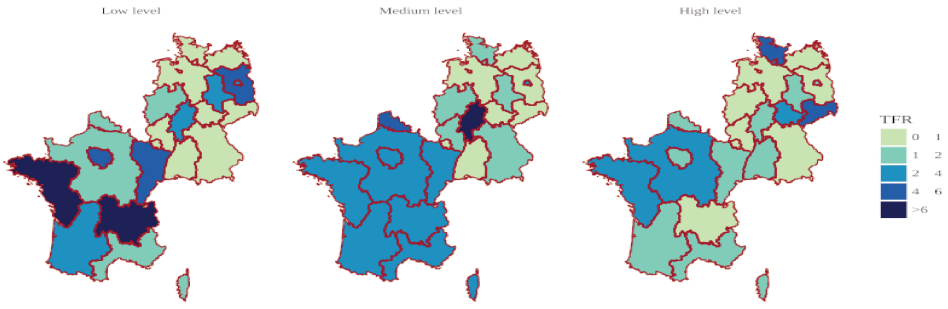
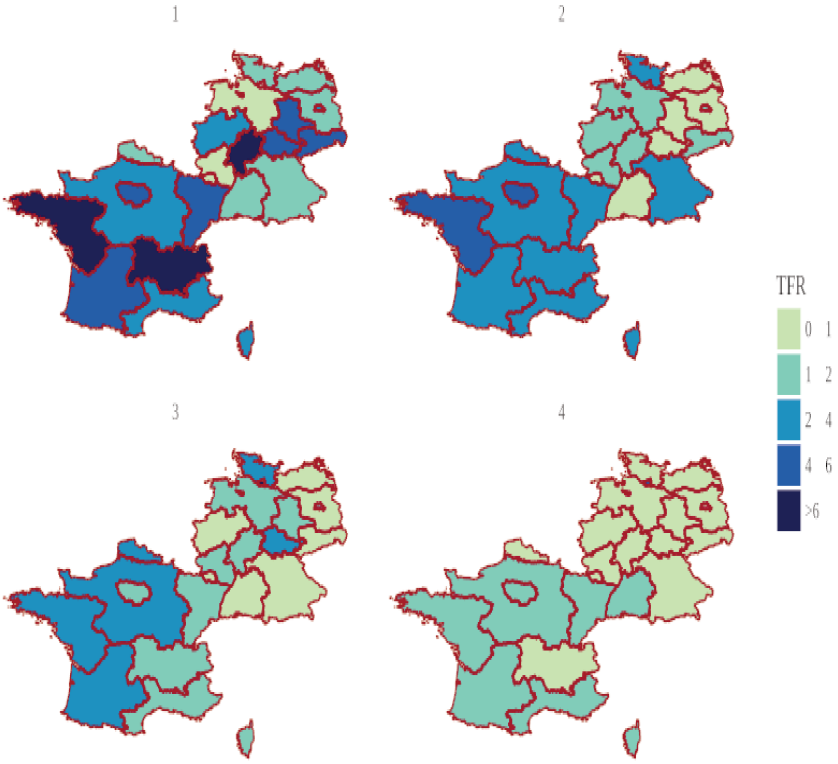


Figure 2.4 shows the TFR by quartiles of personal income. As personal income increases, TFR decreases.

FIGURE 2.4: Total fertility rate by personal labour income quartile



Figures 2.5 and 2.6 show the TFR by total household labour income quartile and total household capital income quartile respectively (the income quartiles were computed for France and Germany separately). In both cases, higher income, regardless of its source, seems to be associated with lower TFR in Germany; not so much in France, which contrasts with Figure 2.4.

FIGURE 2.5: Total fertility rate by household labour income quartile

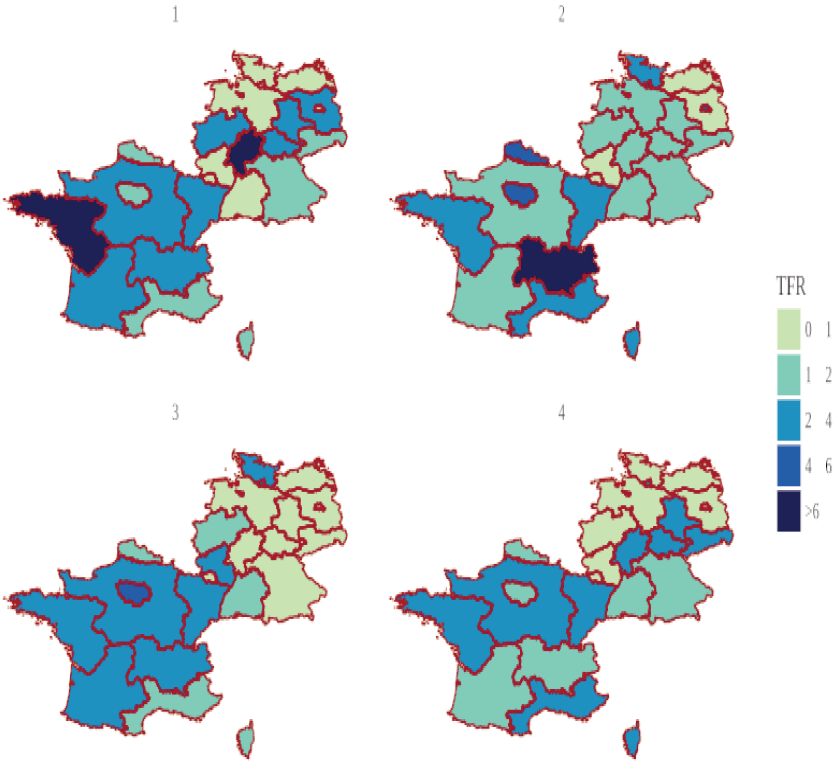


FIGURE 2.6: Total fertility rate by household capital income quartile

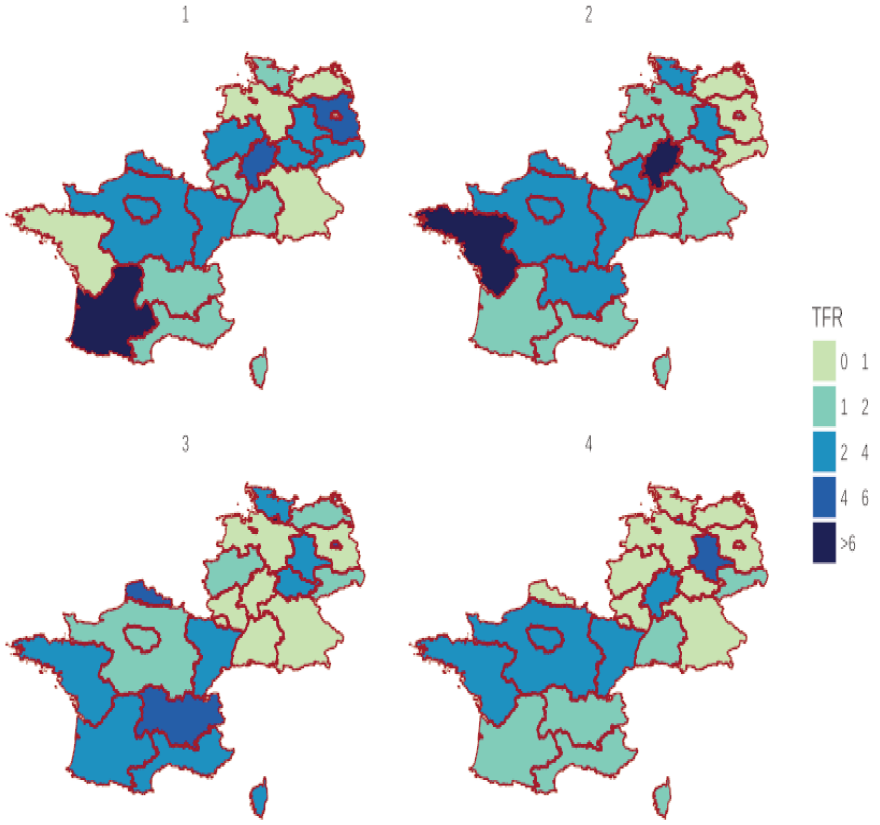
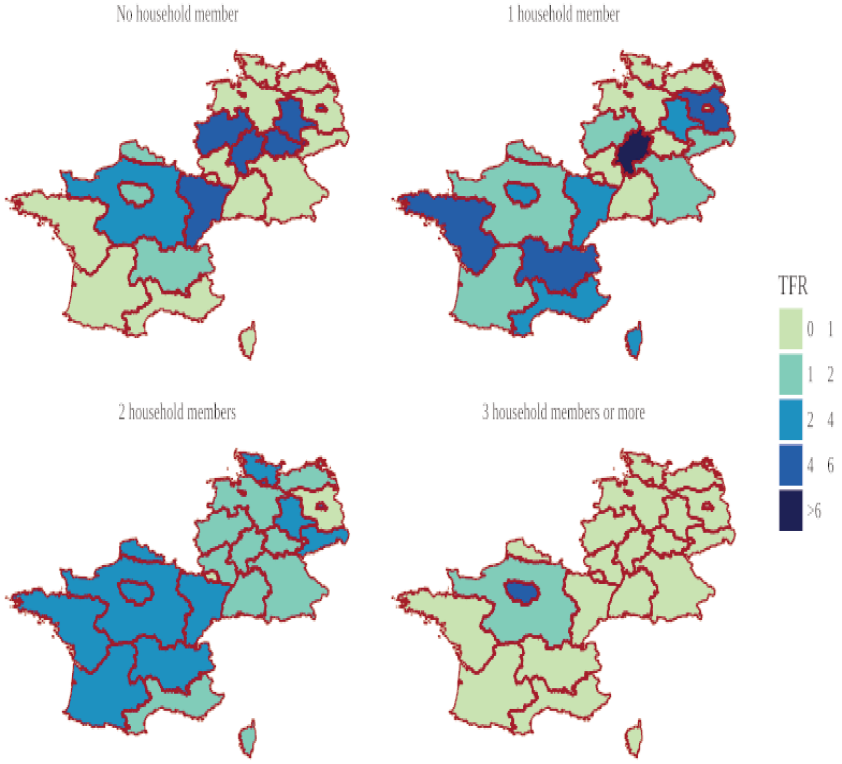


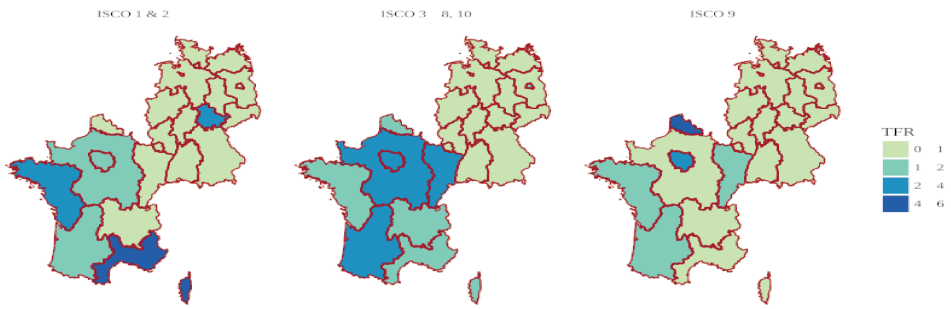
Figure 2.7 shows the TFR by number of household members with a labour income. Maybe surprisingly, TFR is quite high in certain regions when no household members have an income, while very low when 3 household members or more have an income.

FIGURE 2.7: Total fertility rate by number of household members with a labour income



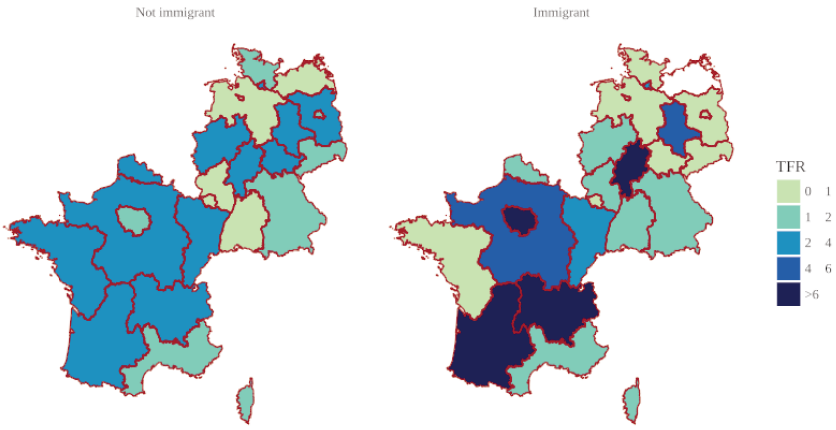
The occupation of the mother does not seem to play a very large role, as can be seen in Figure 2.8, especially in Germany. But one has to keep in mind that this map shows the TFR for mothers whose variable *current main activity* is not `Homemaker/Care for Children`, and thus it is not surprising that TFR is so low, regardless of occupation.

FIGURE 2.8: Total fertility rate by occupation



Immigrant status is associated with higher TFR, as can be seen in Figure 2.9. Figure 2.10 goes into more detail, and shows TFR by citizenship (*other* includes stateless persons as well).

FIGURE 2.9: Total fertility rate by immigrant status



Due to data limitations, it was not possible to look into more detail than these levels. However, one can see differences between France and Germany; there are not a lot of African citizens in Germany, while EU citizens in Germany have a lower TFR than those in France. This is also the case for naturalized citizens.

FIGURE 2.10: Total fertility rate by citizenship

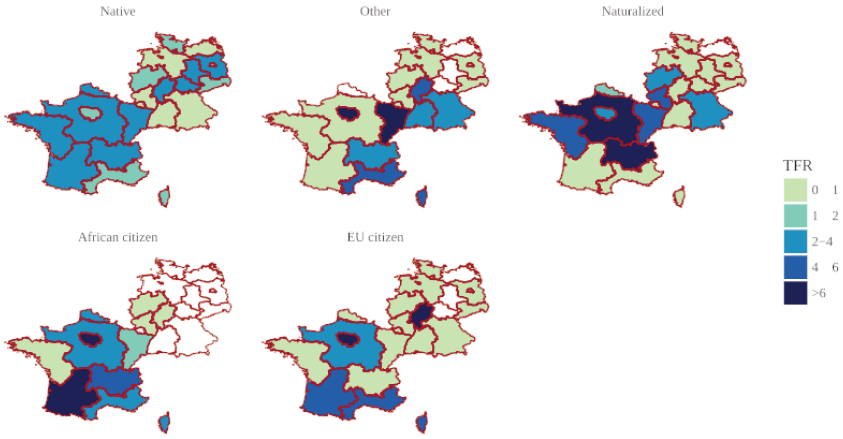


Figure 2.11 shows the difference in TFR between rural and non-rural households, but there does not seem to be any striking differences.

FIGURE 2.11: Total fertility rate by rural/non-rural household

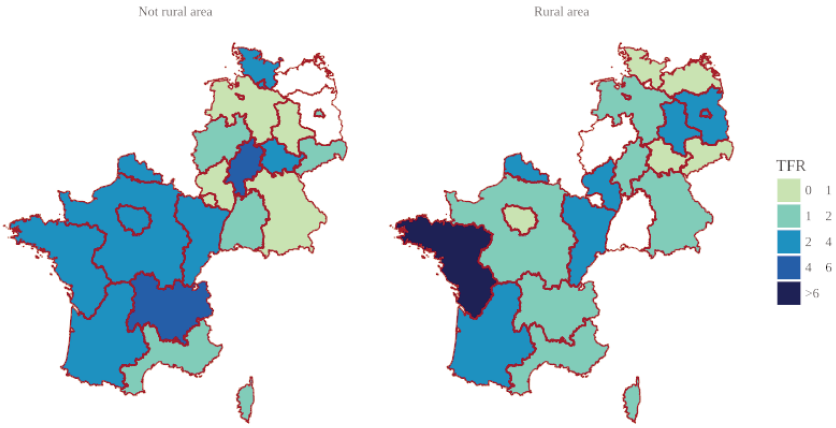
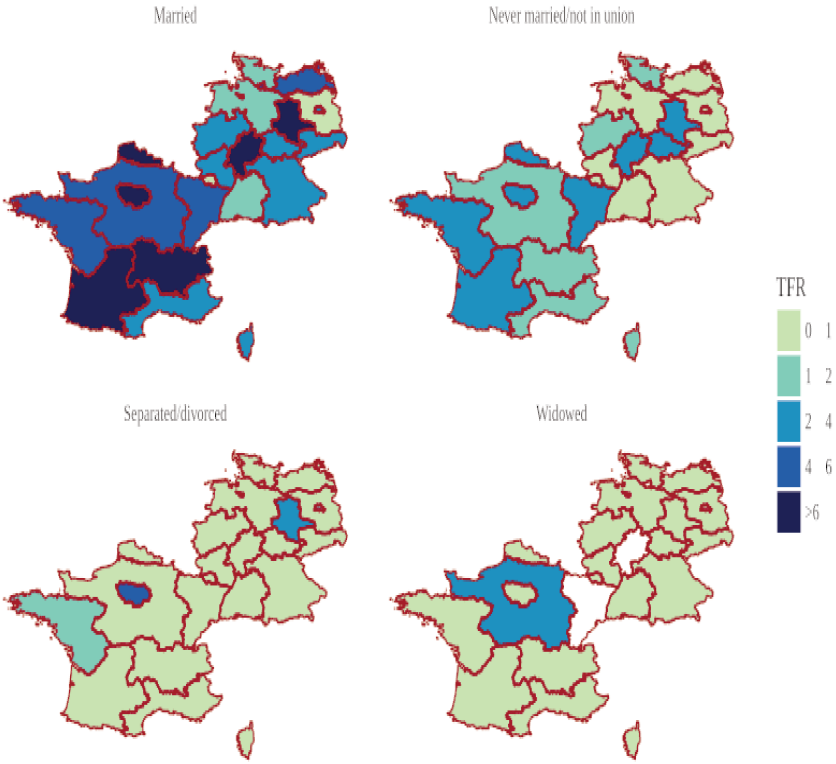


FIGURE 2.12: Total fertility rate by marital status

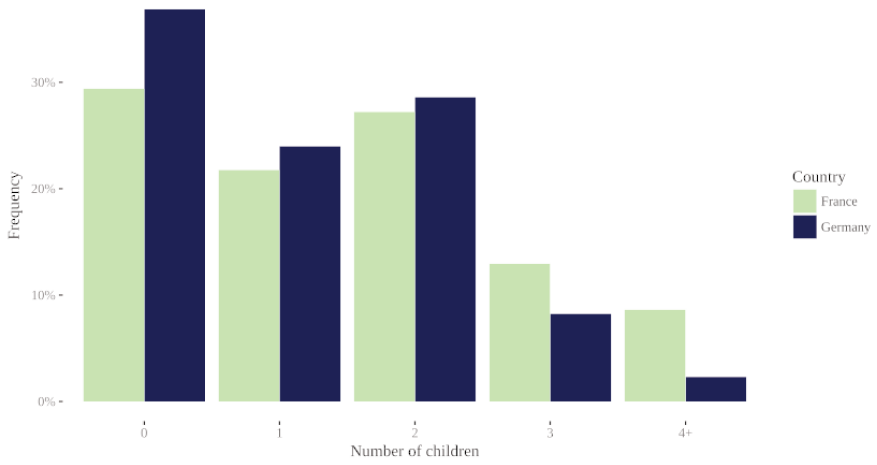


Looking at these maps, it would seem that higher TFR rates are associated with lower income, immigrant and less educated households. In the next section, I will estimate a hierarchichal Poisson model that will enable looking more closely into these correlations.

2.3 Results from the hierarchical model

In this section, I present the results of the estimation of the hierarchical model. A hierarchical model takes the nested structure of the data into account. Estimation of this model will shed some light on variables that might be important in explaining why women have the number of children they have. The model I estimate is a random-intercept model. To make things clear, the jargon used in this section is the one from the statistical literature, not the econometric literature. I draw the reader's attention to this, as the statistical and econometric literature use the same terms to describe different models. The random intercept varies across regions. To summarize; I estimate a mixed-effects model with a random intercept, one per region. Figure 2.13 shows the distribution of the number of own children living in French and German households.

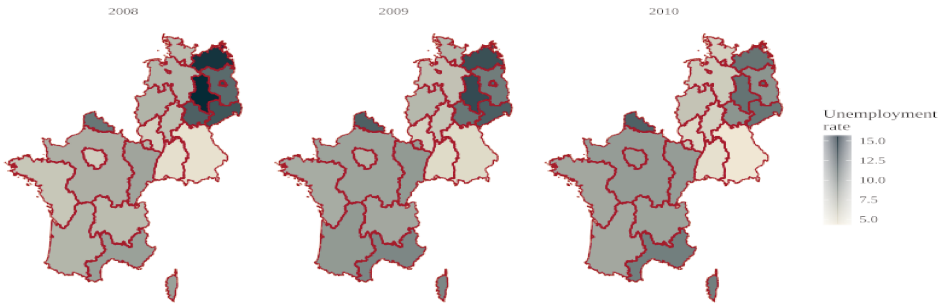
FIGURE 2.13: Frequency of the number of children living in French and German households



This variable is the dependent variable of the model; this is one major shortcoming of the data, as the number of children living in the household does not necessarily correspond to the number of children a women has. This number might underestimate the real number of children a woman has because (some) of her children might have left the household for various reasons. It might also overestimate it, as some of the children present in the household might be only the father's. However, there is no reason to believe that this underestimation and overestimations is any different between France and Germany.

Individuals are nested inside NUTS 1 regions. The unemployment rates are measured at the NUTS 1 level, while all the other variables are measured at the individual level. The individual variables are all measured in 2010 (the last year of availability for France in LIS data), but the unemployment rate is measured in 2008 and 2009. Unemployment rates are different from region to region, as shown in Figure 2.14:

FIGURE 2.14: Unemployment rate in 2008 to 2010 for France and German NUTS 1 regions



Unemployment can be seen as a proxy for the overall economic climate and unemployment spells have been shown to have an impact on the fertility and conception decisions of women, as discussed in the literature review. The unemployment rate is included as an explanatory variable in the model as the growth rate between 2008 and 2009. It would be possible to further nest the NUTS 1 regions inside countries, but in order to keep the model simple this was not done.

The number of children is an integer-valued variable, so I use a hierarchical Poisson model. Because younger women had less time to become pregnant than older women, I add an exposure variable called *fertile years* to account for different exposure times. Fertile years are equal to the difference between a woman's age and 15, with a maximum of 34 years. As stated above, the model is a hierarchical poisson model with a random intercept: the intercept is allowed to vary between regions. The model has the following form:

- i indexes households and j regions.
- n

TABLE 2.6: Covariates used in the model

Variable name	Description	Effect
(Intercept)	Intercept	Fixed and random (regional level)
immigr	Immigration status Non-immigrant, Immigrant	Fixed
cmas	Current main activity status Mainly Employed, Unemployed, Care for children	Fixed
educ	Education level Low, middle, and high education level	Fixed
marital	Marital status Married, single person, separated/divorced	Fixed
status	Employment status Employee, Temporary worker/apprentice, Self-Employed	Fixed
tx_rate	Regional unemployment rate growth between in 2008 and 2009	Fixed
hil	Household income (in millions)	Fixed
hil^2	Household income (in millions), squared	Fixed

I estimate five versions of the model, adding more and more interactions terms to each. Each of these five models is estimated twice, once without an overdispersion parameter, and then with an overdispersion parameter. Estimation of the model is done using the GNU R programming language (R Core Team, 2014) with the `lme4` package (Bates et al., 2015b). Plots were made with the `sjPlot` package (Lüdecke, 2017). All the code was run using `LISSY`, the remote platform of the LIS Data Center located in Luxembourg.

Results are presented in Table 2.7. Controlling for over-dispersion does not improve the fit, so only one table is presented (the one where I controlled for over-dispersion). Also, for most variables, adding further controls does not change the sign or size of the non-interacted variables such as education level and current main activity status.

I will only discuss the last model, containing all the explanatory variables and also with lowest AIC. This model also includes the most interactions between different explanatory variables. In order to facilitate interpretation of the results, I will show and discuss plots of incidence rate ratios and marginal effects, as reading Table 2.7 is not very practical.

TABLE 2.7: Results from the hierarchical model

	(1)	(2)	(3)	(4)	(5)
(Intercept)	**-2.77 *** (0.04)**	**-2.84 *** (0.04)**	**-2.85 *** (0.05)**	**-3.80 *** (0.18)**	**-3.28 *** (0.26)**
Is immigrant	**0.12 *** (0.02)**	**0.33 *** (0.04)**	**0.33 *** (0.04)**	**1.35 *** (0.20)**	**1.21 *** (0.20)**
Is unemployed	-0.02 (0.03)	-0.01 (0.03)	-0.04 (0.05)	-0.04 (0.05)	**-0.09 . (0.05)**
Is homemaker	**0.49 *** (0.02)**	**0.54 *** (0.03)**	**0.65 *** (0.04)**	**0.65 *** (0.04)**	**0.62 *** (0.05)**
Medium education level	-0.02 (0.02)	0.03 (0.02)	0.04 (0.02)	**0.05 . (0.03)**	-0.31 (0.22)
High education level	0.01 (0.02)	**0.11 *** (0.03)**	**0.12 *** (0.03)**	**0.12 *** (0.03)**	**-0.70 ** (0.24)**
Is single person	**-0.82 *** (0.02)**	**-0.82 *** (0.02)**	**-0.76 *** (0.04)**	**-2.04 *** (0.20)**	**-2.10 *** (0.20)**
Is separated/divorced	**-0.54 *** (0.03)**	**-0.54 *** (0.03)**	**-0.50 *** (0.05)**	-0.36 (0.28)	**-0.50 . (0.28)**
Household income	**-1.49 *** (0.31)**	**-1.56 *** (0.31)**	**-1.87 *** (0.48)**	**-2.01 *** (0.48)**	**-4.94 ** (1.79)**
Household income^2	**2.15 *** (0.56)**	**2.13 *** (0.56)**	**6.45 *** (1.60)**	**6.73 *** (1.59)**	**17.4 *** (8.74)
Is immigrant $\times \cdots$		-0.03	-0.02	0.01	-0.001

$\cdots$ household income^2	(5.55)	(6.16)	(15.58)
Is homemaker $\times$ $\cdots$	-2.46	-2.43	1.57
$\cdots$ household income^2	(1.74)	(1.84)	(2.51)
Medium education level$\times \cdots$			0.90
$\cdots$ household income			(1.85)
High education level$\times \cdots$			**5.53 ****
$\cdots$ household income			**(1.71)**
Medium education level$\times \cdots$			-7.90
$\cdots$ household income^2			(8.82)
High education level$\times \cdots$			**-17.15.**
$\cdots$ household income^2			**(8.81)**
Is single person $\times \cdots$	**-1.79 ***	-1.07	-1.18
$\cdots$ household income	**(0.90)**	(0.98)	(0.95)
Is separated/divorced $\times \cdots$	-0.63	-0.71	0.43
$\cdots$ household income	(1.37)	(2.26)	(1.69)
Is single person $\times$ $\cdots$	-0.63	-1.84	0.27
$\cdots$ household income^2	(1.36)	(3.04)	(3.18)
Is separated/divorced $\times$ $\cdots$	**-15.25 ***	**-15.11 .**	-21.26
$\cdots$ household income^2	**(7.66)**	**(8.34)**	(19.10)
Unemployment rate growth		**0.93 *****	**0.54 ***
		(0.16)	**(0.23)**
Is immigrant $\times$		**-0.91 *****	**-0.80 *****
$\cdots$ unemployment rate growth		**(0.17)**	**(0.18)**
Is single person $\cdots$		**1.13 *****	**1.17 *****
$\cdots$ unemployment rate growth		**(0.17)**	**(0.17)**
Is separated $\times$			

The dependent variable is the number of children living in the household at the time of the survey.

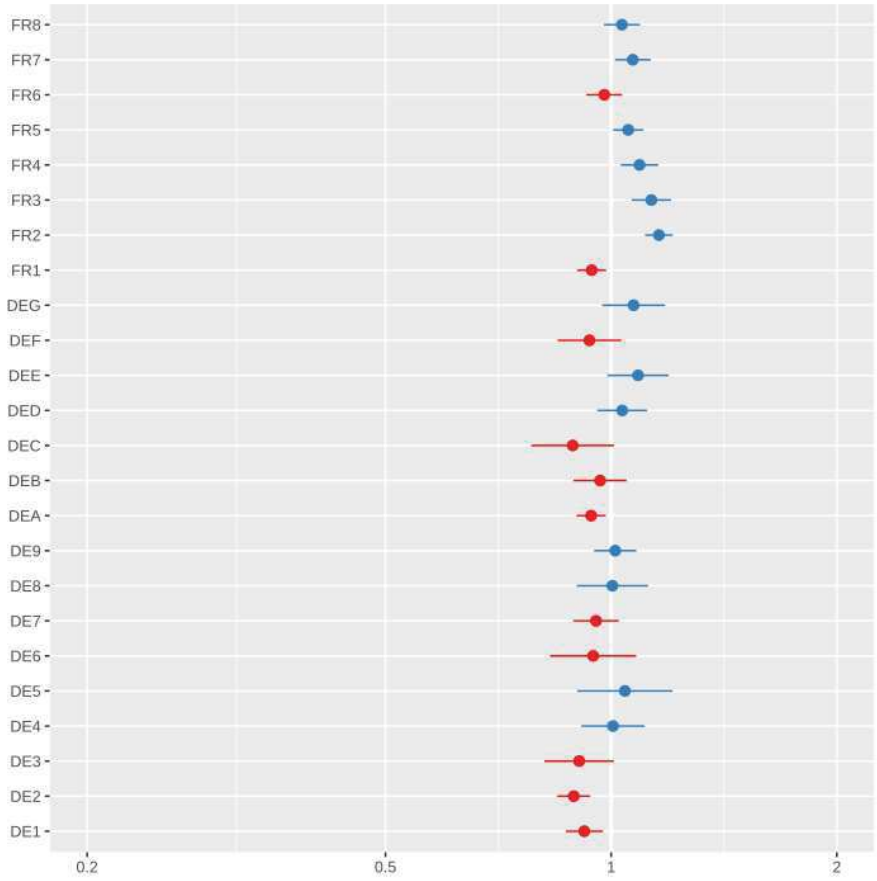
The most important insight of this model is Figure 2.15, which shows the incidence rates of each of the random intercepts. For most French regions[6], this incidence rate is positive. Only Paris is significantly less than 1. One possible explanation is that Paris attracts a lot of young people that might stay there for 2 or 3 years for a job. These young people are usually not in a point in their lives where they have children. In German regions[7] however, the incidence rates are either significantly less than 1, or not significant. This plot is evidence that, even after controlling for personal characteristics of women, there are differences in the number of children between French and German women that are captured by a region specific intercept. This result indicates that differences in TFR can most likely be explained by differences in the institutional settings of France and Germany, as well as culture, at least to some extent. In Figure 2.15, I show the estimated deviation between each region from the global average (Gelman and Hill, 2006).[8] Another point that one might notice, is that σ^2 increases from 0.03 to more than 7000. This is very likely due to numerical and or convergence issues. Despite that, since the coefficients do not vary much, this numerical imprecision is minor.

[6] FR1: Île de France, FR2: Champagne-Ardenne, Picardy, Upper Normandy, Centre, Lower Normandy, Burgundy, FR3: Nord-Pas-de-Calais, FR4: Lorraine, Alsace, Franche-Comté, FR5: Pays de la Loire, Brittany, Poitou-Charentes, FR6: Aquitaine, Midi-Pyrénées, Limousin, FR7: Rhône-Alpes, Auvergne, FR8 Languedoc-Roussillon, Provence-Alpes-Côte d'Azur, Corsica

[7] DE1: Baden-Württemberg, DE2: Bayern, DE3: Berlin, DE4: Brandenburg, DE5: Bremen, DE6: Hamburg, DE7: Hessen, DE8: Mecklenburg-Vorpommern, DE9: Niedersachsen, DEA: Nordrhein-Westfalen, DEB: Rheinland-Pfalz, DEC: Saarland, DED: Sachsen, DEE: Sachsen-Anhalt, DEF: Schleswig-Holstein, DEG: Thüringen

[8] This is the difference between the global average predicted response and the response for a region, exponentiated. A technical description can be found in Bates et al. (2015a).

FIGURE 2.15: Regional incidence rate



In the next sections, I present plots that will show the effect of the different variables on the number of children, the so-called fixed effects in the hierarchical models jargon. I will focus on household income and unemployment rate growth, which were identified as being two of the major economic drivers of fertility decisions.

2.3.1 The effect of household income on the number of children

FIGURE 2.16: Marginal effects of household income

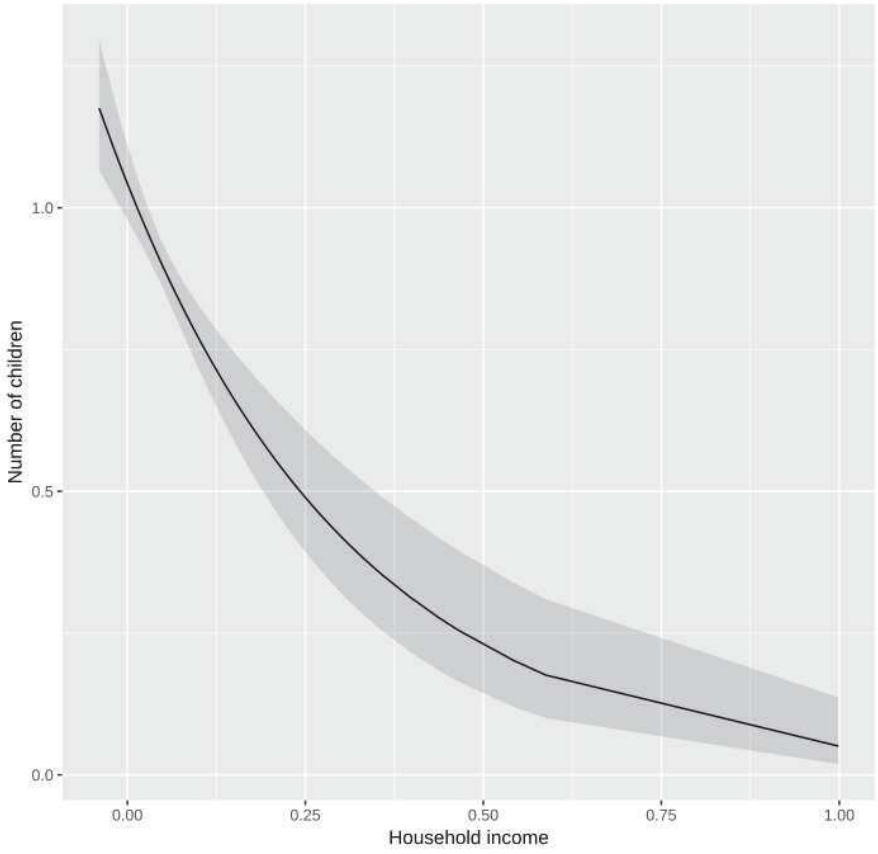


Figure 2.16 shows the marginal effect of household income on the number of children. In Figure 2.17, I show the marginal effect of income on the number of children, by immigrant status, and as can be seen, there is not any difference between the two curves. Figure 2.9 suggested that immigrants have more children on average than natives, but the effect of household income seems to work the same way for these two groups.

FIGURE 2.17: Marginal effects household income, by immigrant status

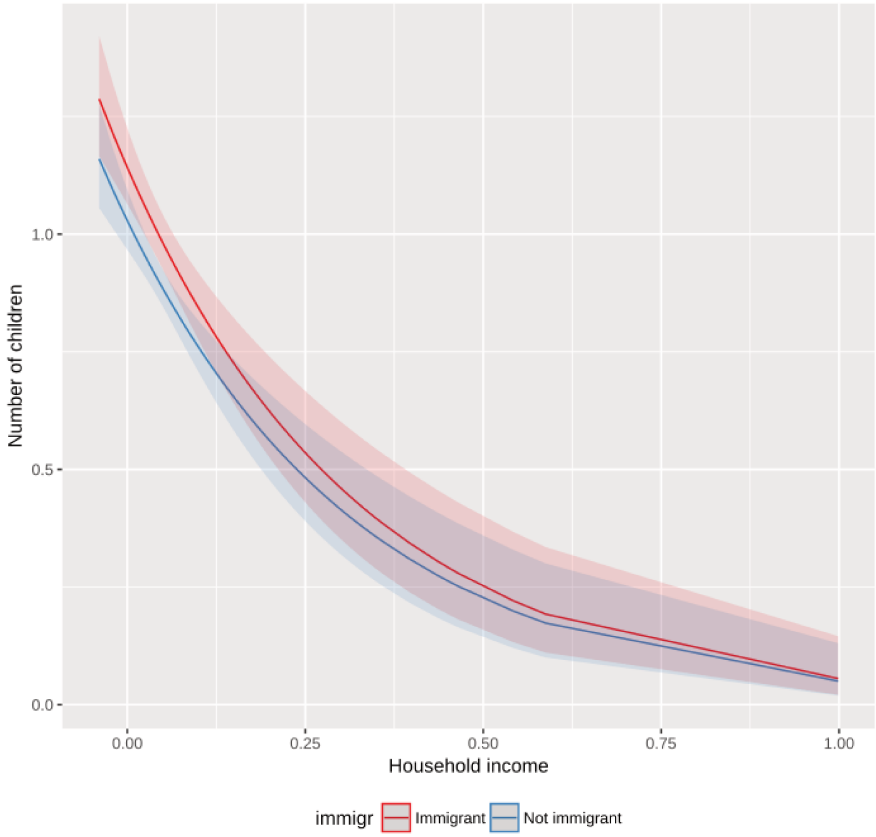


Figure 2.18 shows the effect of household income by current main activity status of the mother. The confidence intervals for the unemployed category become very wide as household income increases, which is not surprising. Indeed, there are not a lot of households where the mother is unemployed and with very large incomes.

FIGURE 2.18: Marginal effects household income, by current main activity status

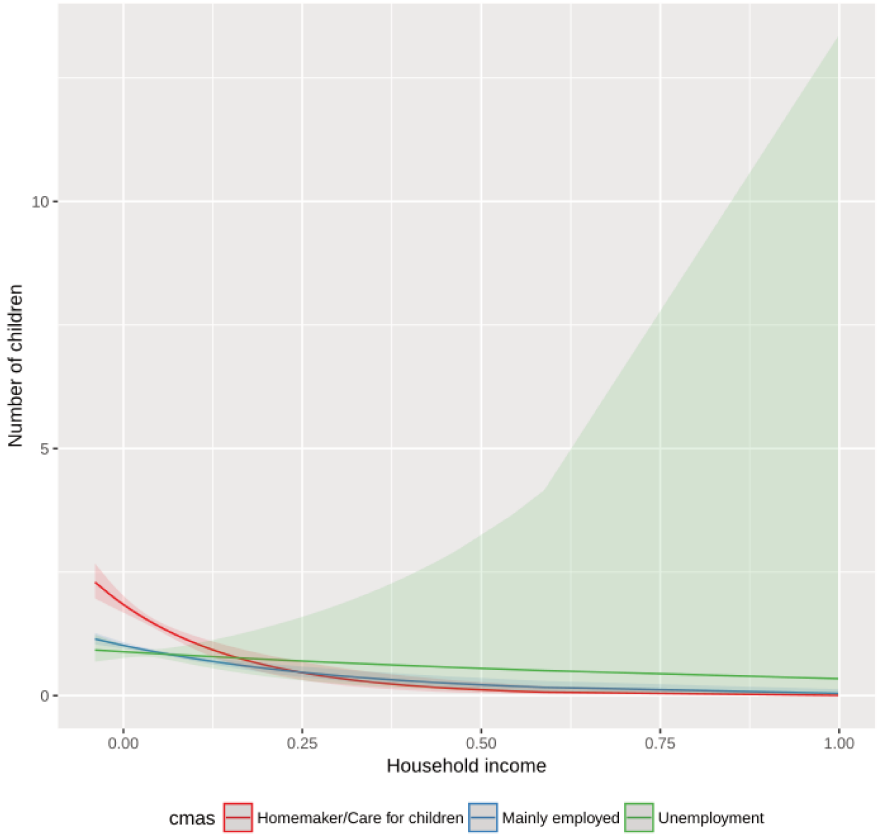
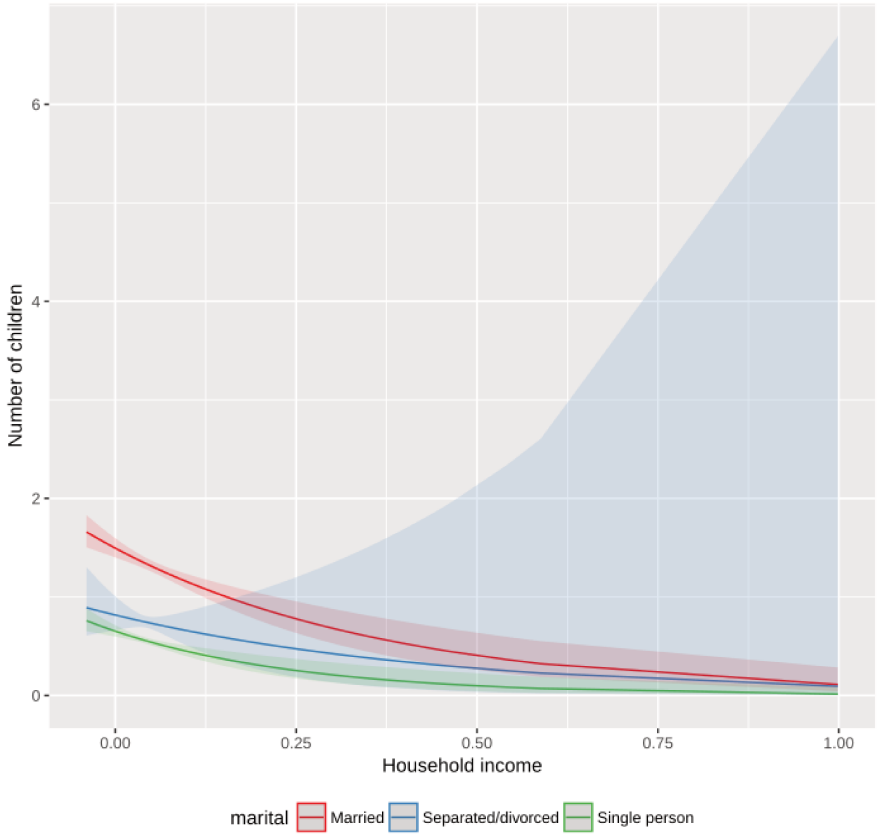


Figure 2.19 shows the marginal effect of household income by marital status. The `Separated/divorced` category has very large confidence bands, but when focusing on only the `Single person` and the `Married` categories, we see that the effect of household income on the number of children is very similar. The difference between the two is that the curve for married women is shifted higher.

FIGURE 2.19: Marginal effects of household income, by marital status



The last split I consider is the one by education status. For the low and medium education level categories the effect is the same, but is quite different for highly educated women.

FIGURE 2.20: Marginal effects of household income, by education level

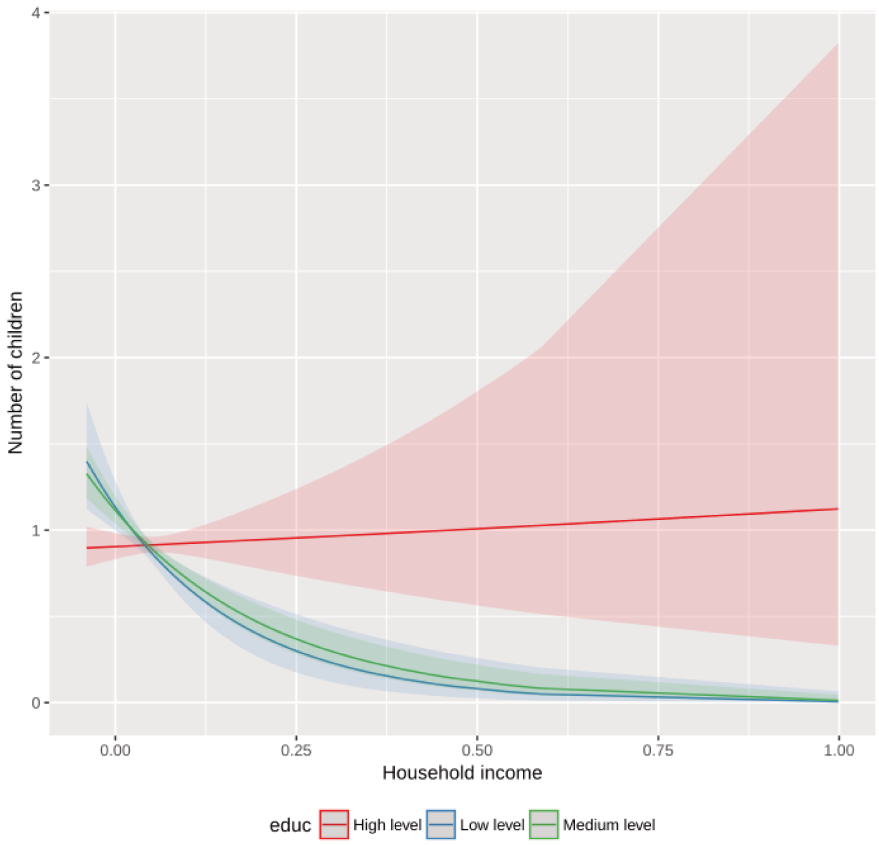
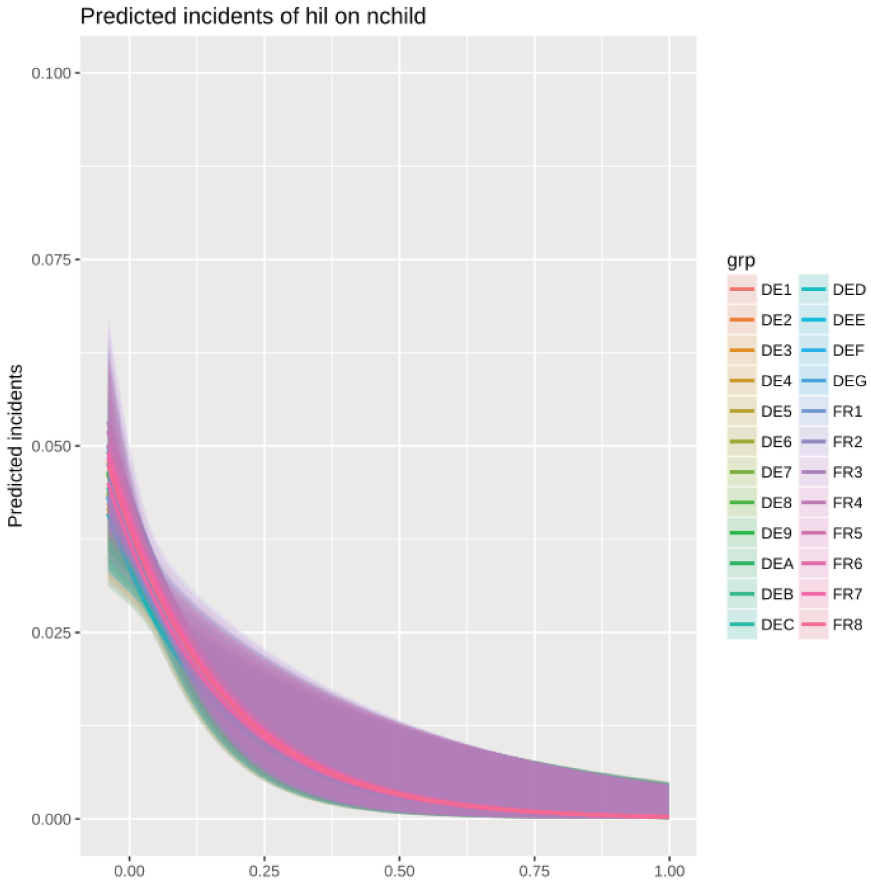


Figure 2.21 shows the predicted incidences[9] of household income on number of children by region. The confidence bands seem to all overlap, which suggests no major differences of household income on predicted incidences of number of children between the French and German regions.

[9] From sjPlot's documentation: *The predicted values are based on the fixed effects intercept, plus each random intercept and each specific fixed term's estimate. All other fixed effects are set to zero (i.e. ignored).*

FIGURE 2.21: Predicted incidences of household income on number of children, by region



2.3.2 The effect of unemployment growth

Figure 2.22 shows the marginal effect of unemployment growth on the number of children is positive, and practically linear. Figure 2.23 splits the effect by immigrant status. The effect is positive for both groups, but the curve is steeper for non immigrants. One possible explanation is that immigrants already have more children (as suggested by the positive and significant coefficient of Table 2.7) so the increase caused by unemployment rate growth is less important.

FIGURE 2.22: Marginal effects unemployment rate growth

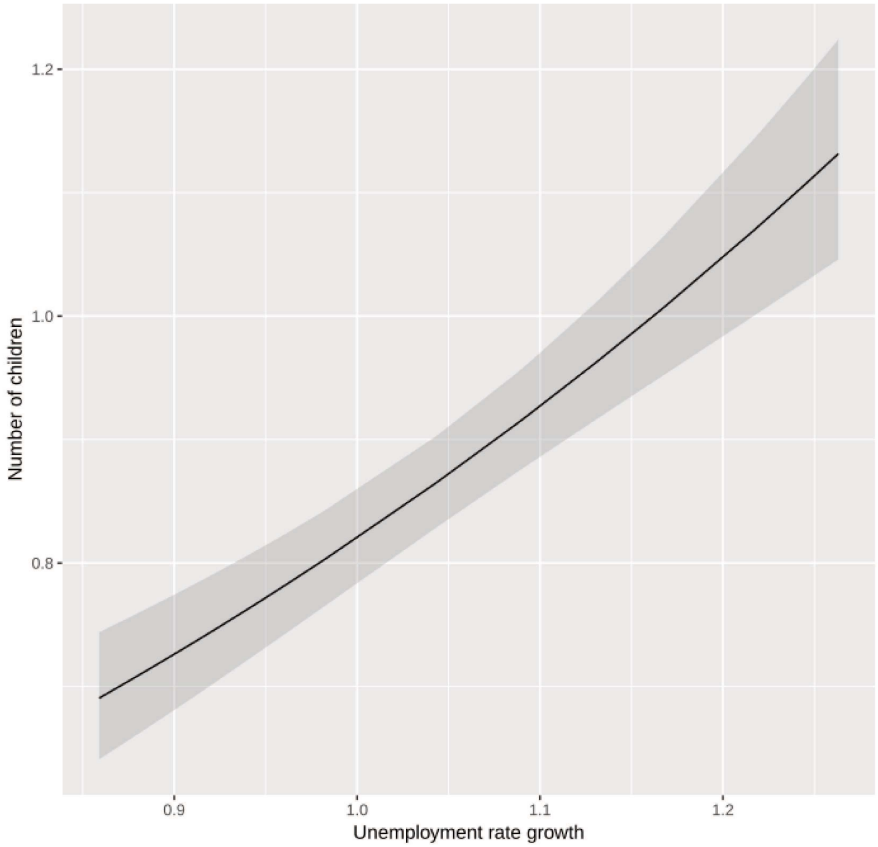


FIGURE 2.23: Marginal effects unemployment rate growth, by immigrant status

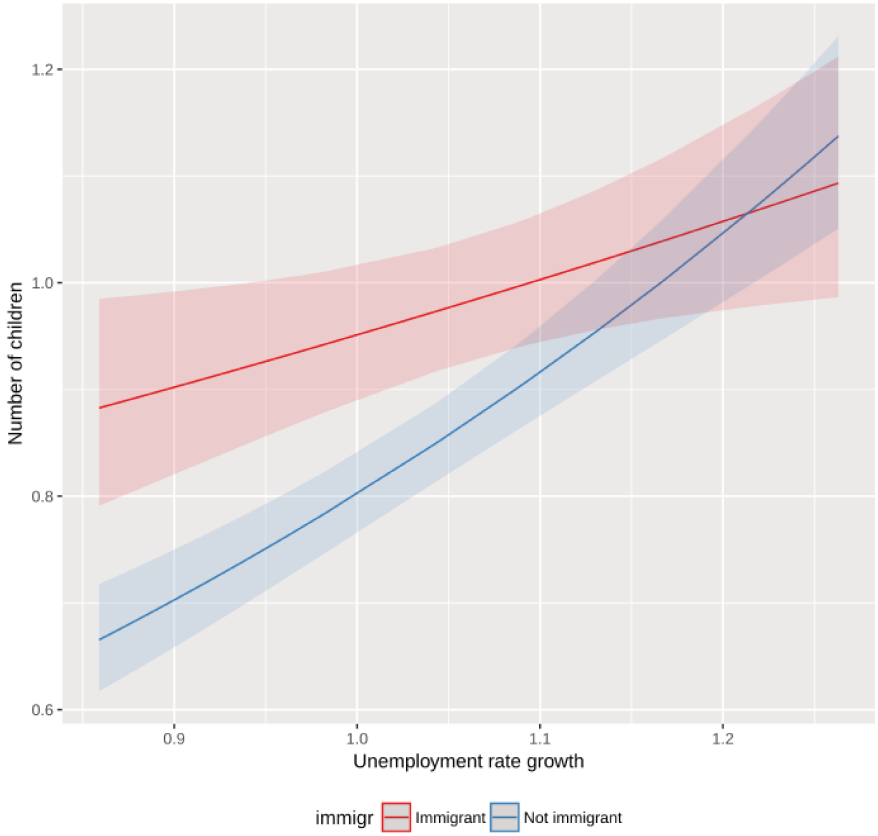
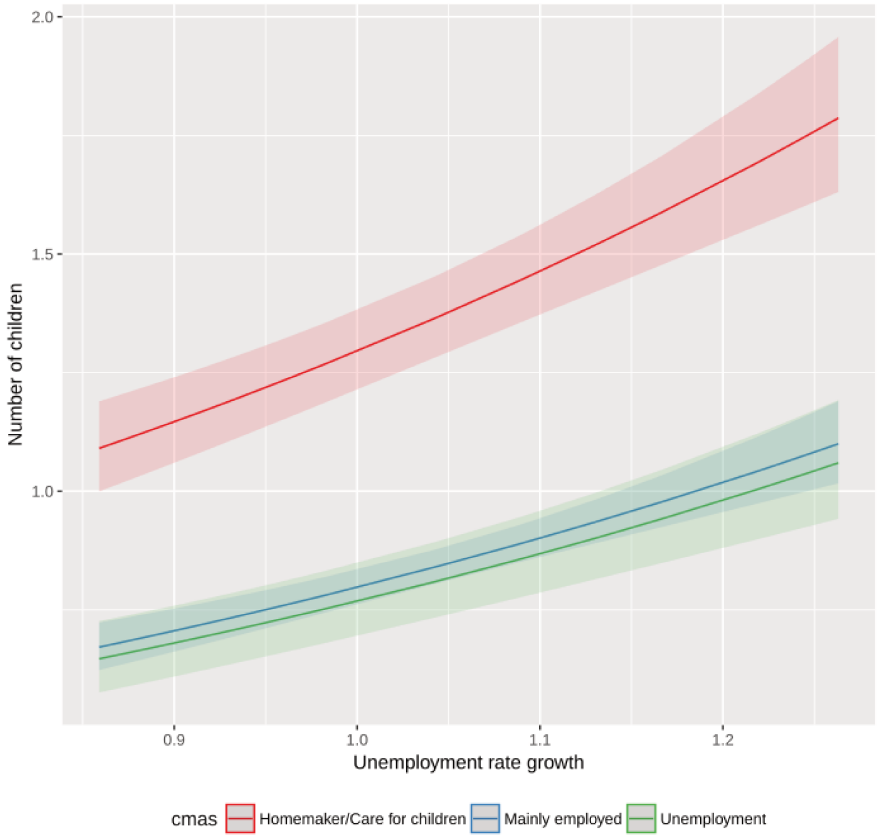


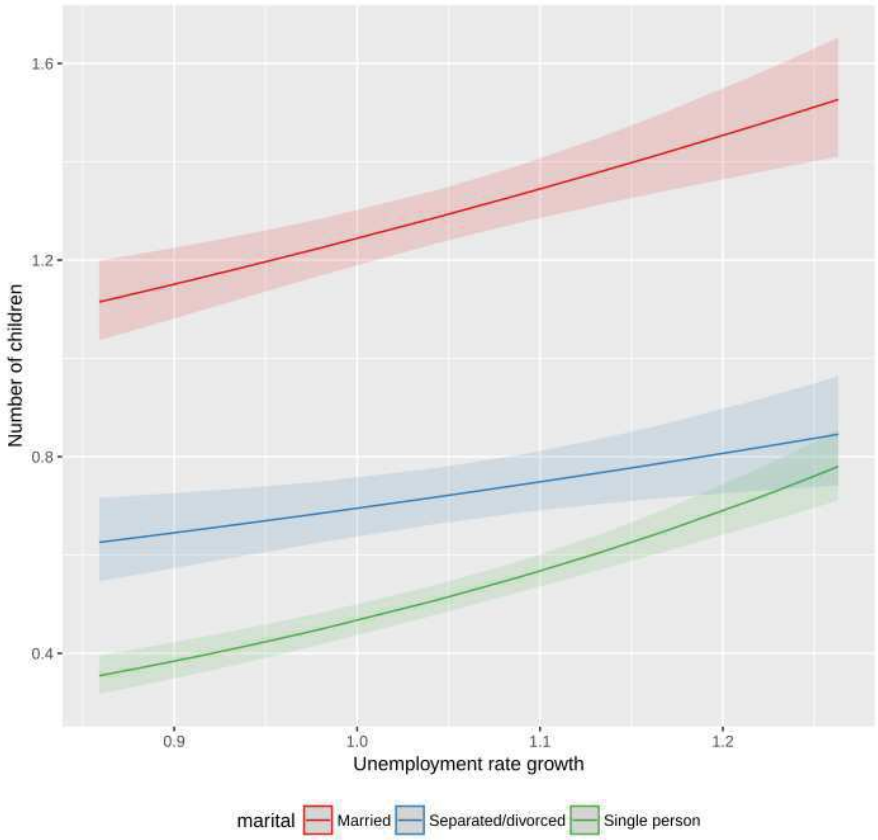
Figure 2.24 shows the effect by current main activity status. For all three categories, the effect is practically linear, and is the same for mainly employed and unemployed women.

FIGURE 2.24: Marginal effects unemployment rate growth, by current main activity status



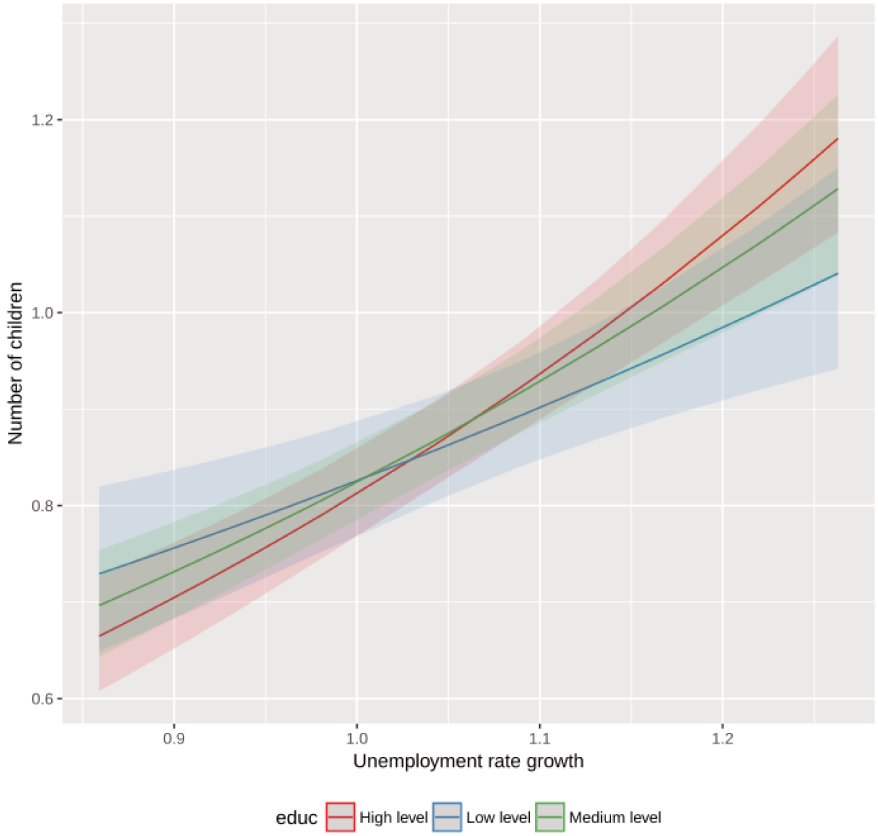
Splitting the effect by marital status shows three different curves, but all increasing. However for the `separated/divorced` category, the effect is not as pronounced as for the other two groups.

FIGURE 2.25: Marginal effects of unemployment rate growth, by marital status



As for education level, the curves are all super-imposed and the curves are quite steep.

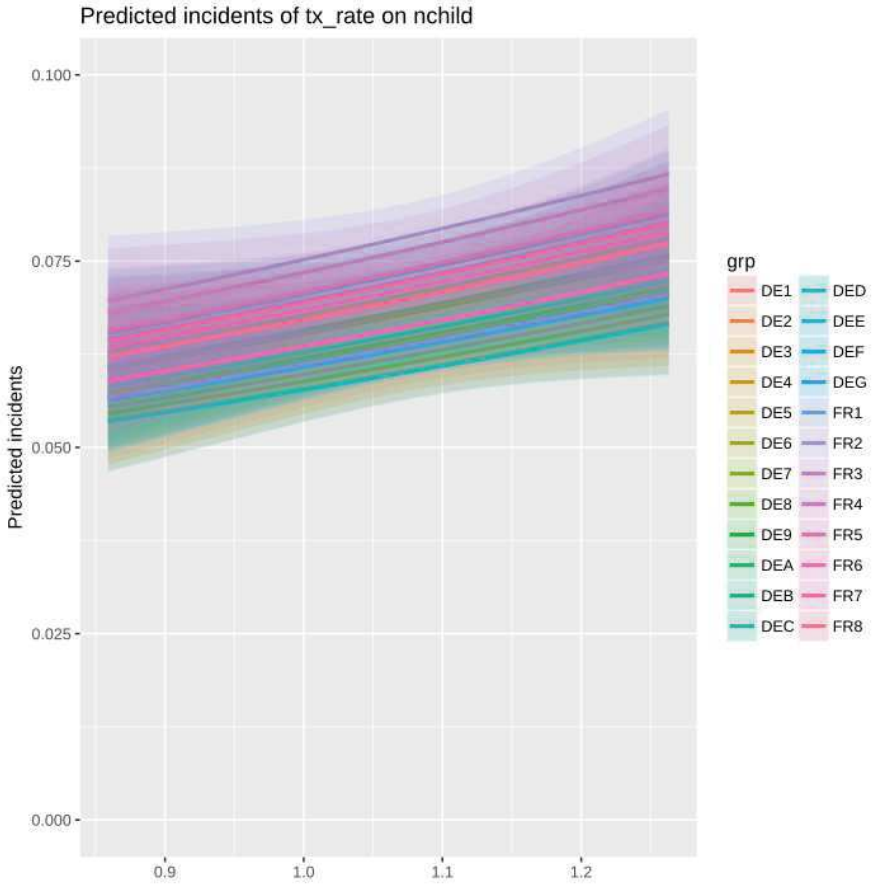
FIGURE 2.26: Marginal effects of unemployment rate growth, by education level



In the literature review I presented papers that argued that economic downturn correlated with lower fertility rates, which is not a result that I find here.

Figure 2.27 shows the predicted incidence of unemployment rate growth on number of children. Even though it is not easy to clearly see which region is above which, it is clear that the French regions are above the German ones. Figures 2.21 and 2.27 are not easy to decipher, but are included here because they show that the effect of these variables is roughly the same between all the French and German regions. This points to the national policies and perhaps also culture as the potential sources of differences in fertility.

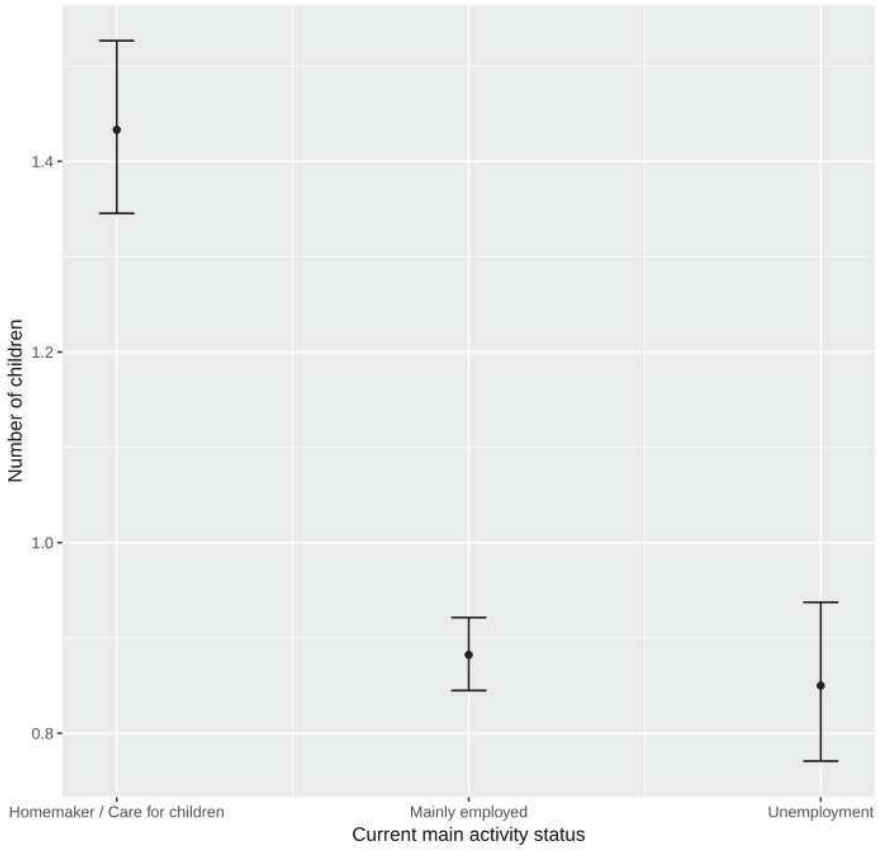
FIGURE 2.27: Predicted incidences of unemployment growth on number of children, by region



2.3.3 Other results

The coefficient of the immigrant status dummy is positive, indicating a positive relationship between immigrant status and number of children. Interacting immigrant status with current main activity status yields a negative effect when the immigrant is a homemaker. Interacting with education status also yields a negative coefficient; higher educated immigrants have fewer children than nationals. Unemployment rate growth further reduces the number of children for immigrant families than for nationals.

FIGURE 2.28: Incremental effect of current main activity status



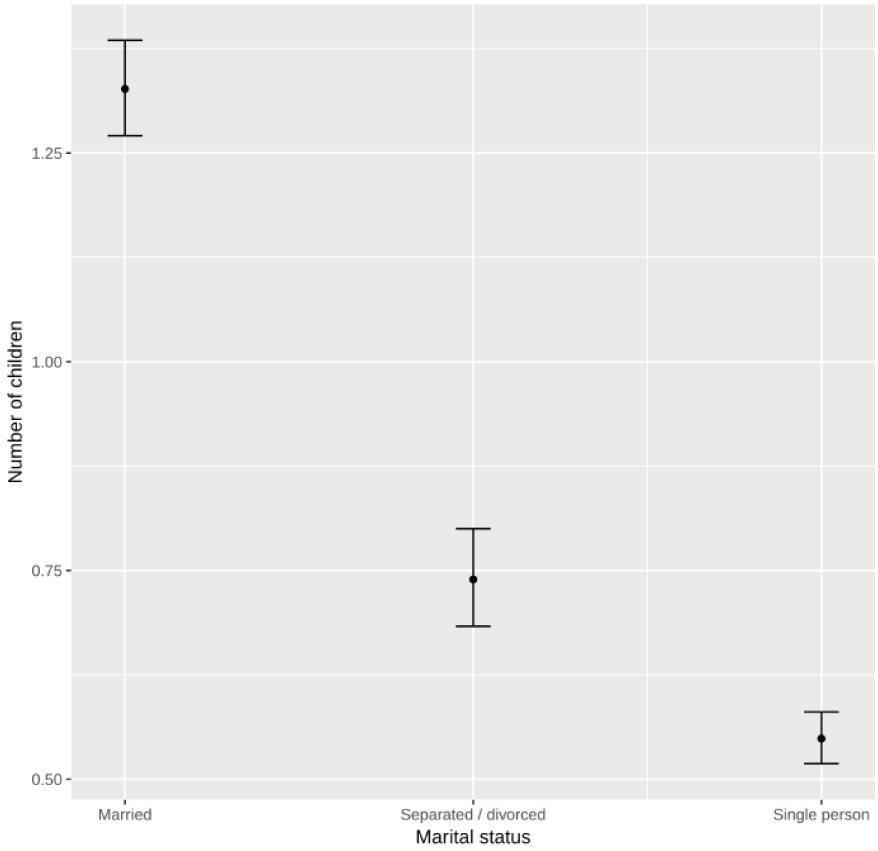
As hinted by Figure 2.2, current main activity status correlates with TFR in an expected manner; homemakers and women that care for children have more children than women that are mainly employed. Current main activity status is endogenous and from the present model it is not possible to derive any causal effects.

Interactions are not statistically significant, apart when interacting homemaker status with household income. For given household income, women that are homemakers have less children than employed women. This result can seem surprising. A possible explanation could be that stay-at-home mothers prefer investing in quality instead of quantity of children. Figure 2.28 shows the incremental effect of the different categories on number of children.

A high education level (corresponding to tertiary education) correlates with a lower number of children. Interacting education level with immigrant status further decreases number of children. Interacting with household income yields a positive effect for highly

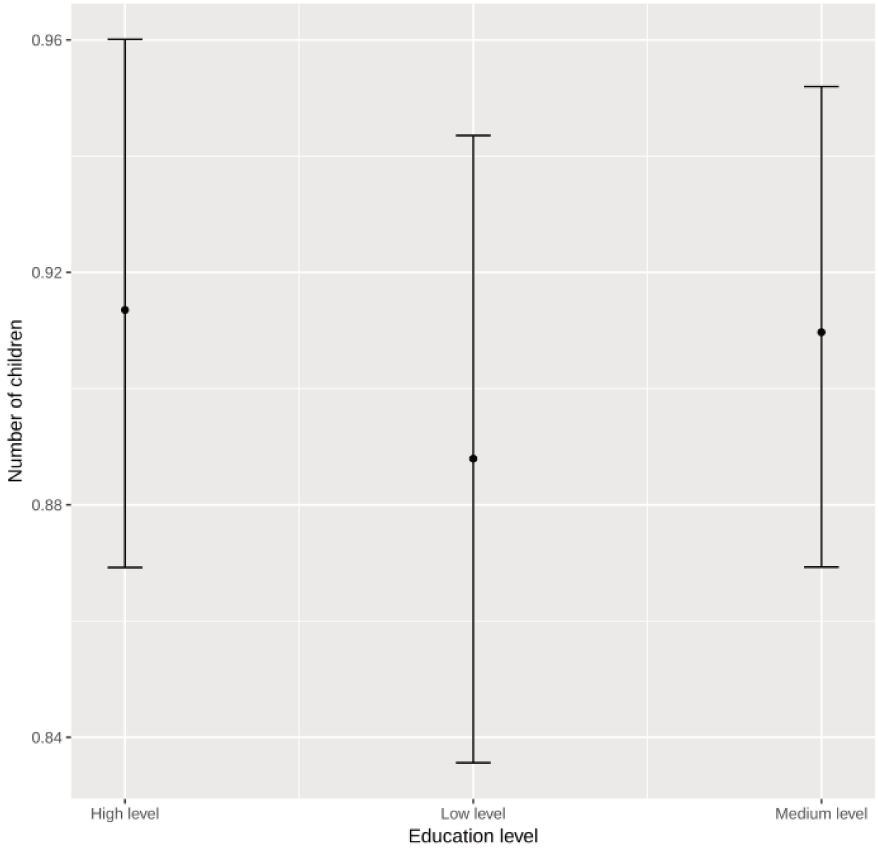
educated women and interacting with the growth of unemployment too (but as shown by Figure 2.26, the effect is the same across education levels).

FIGURE 2.29: Incremental effect of marital status



Unsurprisingly, married women have more children than non-married women. Of course none of these results are causal, but serve to identify which variable seem to matter more than others as well as the sign of the correlation.

FIGURE 2.30: Incremental effect of education level



The goal however of this regression was not to determine the impact of the variables measured at the individual level. The regression discussed sheds light on associations between the number of children and these variables, but one should not draw any causal conclusions from the results. However, these regressions help determine differences between childbearing decisions between regions. This is the case, as can be seen from Figure 2.15.

2.4 Discussion and conclusion

LIS data has historically been used to study poverty and inequality across countries. However, because the data is harmonized, it can also be used for comparative studies such as this one. This study focused on regional differences in TFR, after controlling for individual fixed-effects. A random effect, in the form of a varying regional intercept, was added. The results showed that for most French regions, the random intercept was positive, or, in the case of the Parisian region, negative. For the FR6 NUTS 1 region (comprised of Aquitaine, Midi-Pyrénées and Limousin), this effect was not significant. For Germany however, the random intercepts were either negative, or not significant. It would thus seem that there is what could be called a baseline TFR that is higher for France than for Germany, and that this baseline TFR is determined by the institutional setting of the country, culture, or both. However, the study by Klüsener et al. (2013) indicates that it is the institutional setting rather than culture that plays a role, at least for Western German regions. Could this also be the case for French regions? It would be of interest to try to include institutional variables as well as variables proxying for culture measured at the country level to try to shed more light into this question. Adding more countries to this analysis would also be of interest, but collecting data, especially on the institutional settings of each country would require large efforts.

In the present study, unemployment growth between the years 2008 and 2009 was found to correlate positively with number of children. This is a result that is in contradiction with some of the papers discussed in the literature review, where it was found that economic downturn correlates with lower fertility. However, one has to keep in mind that the dependent variable here is not fertility per se, but number of own children living in the household. This could then mean that what happened is that younger people, who were hit hard by the Great Recession, decided to come back to their parents home, while waiting for better employment prospects. This is indeed a shortcoming of the present study; the dependent variable is not fertility as such, but rather a proxy for fertility. However, data sources with fertility, socio-demographic variables and that are internationally comparable are not abundant. This is why I settled for the LIS data. Also, since this is cross-sectional data, the marginal effects are not showing changes, but differences. It might also be the case that regions with high unemployment are also those with larger families, thus appearing as if an increase unemployment also increases fertility. Panel data would be needed to settle this question, but as mentioned before,

such data sources are not abundant, and I am not aware of any source that would be internationally comparable.

Bibliography

Alicia Adsera. Vanishing children: From high unemployment to low fertility in developed countries. *American Economic Review*, 95(2):189–193, May 2005.

Pau Baizán. Regional child care availability and fertility decisions in Spain. *Demographic Research*, 21(27):803–842, 2009.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015a. doi: 10.18637/jss.v067.i01.

Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015b.

William P. Butz and Michael P. Ward. The emergence of countercyclical u.s. fertility. *The American Economic Review*, 69(3):318–328, 1979.

Chiara Ludovica Comolli. The fertility response to the Great Recession in Europe and the United States: Structural economic conditions and perceived economic uncertainty. *Demographic Research*, 36(51):1549–1600, 2017.

Stefany Coxe, Stephen G. West, and Leona S. Aiken. The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2):121–136, 2009.

Janet Currie and Hannes Schwandt. Short- and long-term effects of unemployment on fertility. *Proceedings of the National Academy of Sciences*, 111(41):14734–14739, 2014.

John F. Ermisch. Purchased child care, optimal family size and mother's employment theory and econometric analysis. *Journal of Population Economics*, 2(2):79–102, 1989.

Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.

Joshua Goldstein, Wolfgang Lutz, and Maria Rita Testa. The emergence of sub-replacement family size ideals in europe. *Population Research and Policy Review*, 22(5):479–496, 2003. ISSN 1573-7829.

Joshua Goldstein, Michaela Kreyenfeld, Aiva Jasilioniene, and Deniz Dilan Karaman Örsal. Fertility Reactions to the "Great Recession" in Europe: Recent Evidence from Order-Specific Data. *Demographic Research*, 29(4):85–104, 2013.

Joshua R Goldstein, Michaela Kreyenfeld, and Felix Rößger. Gibt es eine trendumkehr in der kinderzahl nach geburtsjahrgängen in deutschland. In *Bundesministerium für Familie, Senioren, Frauen und Jugend (Hrsg) Berliner Demografie Forum 2012: Familie-Kinder-Gesellschaft 4. BMFSFJ, Berlin*, 2012.

Karsten Hank and Michaela Kreyenfeld. A multilevel analysis of child care and women's fertility decisions in western germany. *Journal of Marriage and Family*, 65(3):584–596, 2003. ISSN 1741-3737.

Sebastian Klüsener, Karel Neels, and Michaela Kreyenfeld. Family policies and the western european fertility divide: Insights from a natural experiment in belgium. *Population and Development Review*, 39(4):587–610, 2013.

Daniel Lüdecke. *sjPlot: Data Visualization for Statistics in Social Science*, 2017. URL https://CRAN.R-project.org/package=sjPlot. R package version 2.4.0.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014. URL http://www.R-project.org/.

Concetta Rondinelli, Arnstein Aassve, and Francesco Billari. Women's wages and childbearing decisions: Evidence from Italy. *Demographic Research*, S12(19):549–578, 2010. doi: 10.4054/DemRes.2010.22.19.

Tomáš Sobotka, Vegard Skirbekk, and Dimiter Philipov. Economic recession and fertility in the developed world. *Population and Development Review*, 37(2):267–306, 2011.

Agnese Vitali and Francesco C. Billari. Changing determinants of low fertility and diffusion: a spatial analysis for italy. *Population, Space and Place*, 23(2):e1998–n/a, 2017.

Chapter 3

The birth of a child and its impact on wages and worked hours by education level of the mothers and fathers: evidence from France[1]

3.1 Introduction

Major industrialised countries are confronted to low fertility. France seems to be an exception to this rule however. In France, fertility rates are high relative to other industrialised countries such as Germany, but still below the replacement rate of 2.1. This low fertility rate in industrialised countries may be linked to the emancipation of women. Gayle and Miller (2006) explain that in the USA, a typical woman had four children at the beginning of the 20th century, but that number had decreased to 1.9 at the end of the last century. This decrease in childbearing has also been accompanied by an increase in female labour supply. According to Gayle and Miller (2006), "*the participation of all wives increased by 36% over the last 25 years, the rates of mothers with children under the age of three increased by 83%, and by 91% for women with children one year old or younger*" (p. 2).

Women can now have dynamic careers just like men, but evidence shows that it is still mostly mothers who decrease their labour supply to take care of their children. According to the survey *Families and employers* conducted by the INSEE and the INED and exploited by Pailhé and Solaz (2007), after having their first child, 22% of women perform a professional transition linked to a birth against only 5% for men.[2] These

[1] This chapter is derived from an article co-authored with Vincent Vergnat.

[2] The INSEE is the national institute of statistics and economic studies, the INED is the national institute for demographic studies

transitions can slow down women's careers and the literature on female labour supply and fertility decisions gives a clear picture of this phenomenon. As an example for Australia, Baxter et al. (2008) show that women increase their housework hours greatly after giving birth, which is not the case for men, and men actually increase their labour supply for higher order births.

The reduction of labour supply has an impact on the careers of women which can potentially create a wage gap between mothers and women without children: the family wage gap. During the last two decades, increasing attention has been given to the family wage gap and the literature has been concerned with explaining this family wage gap. For example, Waldfogel (1997) is one of the first contributions to this literature. The author shows, using American data, that labour market experience is not the only explanation for the family wage gap. The author tests two other possible explanations: unobserved heterogeneity and part-time employment. Unobserved heterogeneity does not seem be an important factor to explain the family wage gap, unlike part-time employment. Even after controlling for these factors, an important part of family wage gap remains unexplained. The family wage gap can also be observed in other industrialised countries; Gangl and Ziefle (2009) estimate wage equations for the UK, Germany and the United States of America and show that motherhood is associated with wage penalties ranging from 9% to 18%. German women are especially penalized, because they tend to take long childcare breaks, which is not the case for British and American women; this wage penalty is estimated to be around 16% to 18% in hourly wages for German women, 13% for British women and 9% for American women. Gangl and Ziefle (2009) also show that women with children invested less in education than *comparable* childless women and that women tend to favour child-friendly occupations. Davies and Pierre (2005) focused on the family gap in European countries. To compare the impact of motherhood on earnings in Europe, the authors used the ECHP to estimate wage equations for 11 countries.[3] The authors show that the size in wage penalties in pay is different across countries and depend also on the number of children and the timing of the first birth. They find significant penalties on the wage in many European countries like Germany, the United Kingdom or Denmark. For France, they find a wage penalty of 10% after controlling for selection, but only for mothers with more than 3 children.

This negative impact on a woman's career is more important for educated women, who have invested more in human capital and therefore had a more lucrative career in front of them. For example, Adda et al. (2017) show that in Germany, women with abstract

[3]ECHP: European Community Household Panel Survey

occupations have a much higher skill atrophy rate than women with less abstract occupations (at most 6.9% for women in abstract occupations versus 0.6% for women in routine occupations). These abstract occupations are in general demanded for jobs that require higher education. These highly educated women have also, on average, less children. Adda et al. (2017) also estimate the amenity values of different occupations and show that abstract occupations have a very low amenity value when children are present. They show that if abstract occupations had the same amenity value that routine occupations, the part of women that would choose to work in abstract occupations would increase by 5%. Francesconi (2002) studies the fertility and labour supply decisions of young married women. In his simulations, he shows that "*increased schooling decreases the expected number of children substantially*" (p. 367) and that lower wages increase fertility. These findings seem also to indicate that it is mostly highly educated women that have low fertility and that having a child is more costly for them than for lower educated women.

The family wage gap could also be applicable for men. This issue was studied, among others, by Lundberg and Rose (2000). The authors study the impact of a birth on the wage and worked hours of married men and women in the United States. They show that the negative family wage gap for women depends on the duration of maternity leave and that a birth has a positive impact on the hourly wage of men. Other papers confirm this last result (Lundberg and Rose (2002), Glauber (2008), Hodges and Budig (2010)). Killewald (2013) estimates wage equations using the NLSY[4] 1979 and finds that the wage premium for fathers depends on the family context, namely that biological fathers living with the mother of their children gain around 4% in hourly wages, but unmarried fathers, or stepfathers, do not. However, this 4% wage premium decreases to 1.3% for married, residential fathers who are married to women working full-time. In couples where both the husband and the wife are working full time, specialization cannot occur. Thus, the author argues, these fathers have also household responsibilities which makes it difficult for them to commit more to their careers and thus increase their wages.

Davies and Pierre (2005), discussed above, Meurs et al. (2010), Duguet et al. (2015) and Wilner (2016) are studies that focused on France, which is also the focus of the present paper. Meurs et al. (2010) use the French "Families and Employers" survey to study the impact of children and duration of maternity leave on the gender wage gap. Their results show that having a child creates no direct pay penalties for women and a bonus for men. Moreover, having a child has an indirect negative impact on the hourly

[4]National Longitudinal Survey of Youth

wage of women through the reduction of labour supply (part-time job or time out of labour force to take care of children). Duguet et al. (2015) use difference-in-differences to estimate the family wage gap of women and men working in both the private sector and the public sector. Duguet et al. (2015) find that for women in the public sector, the impact on the wage is a loss from about 3.5% to 6.5%, while it is 9.1% for women in the private sector. Three years later, women still earn less, from about 0.9% to 2.4%, which is lower than their initial wage. Worked hours also decrease. Duguet et al. (2015) also show that for men, having a child is associated with an increase in wages but a decrease in hours worked. Wilner (2016) uses French administrative data and wage equations to test whether the self-selection of women that expect to have children into low wage firms could explain the family wage gap. After controlling for this selection as well as for unobserved heterogeneity and human capital, the author finds that mothers had a penalty in hourly wage of approximately -2.2% per child and fathers do not enjoy any loss or premium.

France has always had very generous social policies, which might be one explanation of the higher fertility rate than its neighbouring countries. Already in 1985, the *Allocation Parentale d'Éducation* (APE) allowed parents of 3 or more children to receive a lump sum if they reduced their labour supply to take care of their children: either by completely stopping working, or by working less hours. However, one condition required that the youngest child was under 3 years old. In 1994, the APE had been extended to parents of two children. The impact of this reform has been studied by Lequien (2012). In 2004, the APE was replaced by the *Complément de libre choix d'activité* (CLCA). The CLCA is also a lump sum, which depends on family resources, given to parents that completely stop, or reduce their labour supply to take care of a child, that must be younger than 3 years old. For their first child, parents received the CLCA during 6 months, and from their second child, they got the CLCA until the third birthday of the last child. It is mostly women who claim this allowance as shown by Boyer and Nicolas (2012). Boyer and Nicolas (2012) and Joseph et al. (2013) show that the proportion of working women reducing their labour supply after the first birth has increased and the impact of the reform depends on the education level of the mother. To incite fathers to stay at home and increase the labour supply of women, the CLCA was replaced in 2015 with the *Prestation partagée d'éducation de l'enfant* (PreParE). The PreParE works in a different fashion than the CLCA: parents both get a lump sum for 6 months until the first birthday of their child. But one parent alone cannot stay 1 year at home and claim the benefits of his/her partner (or reduce his/her labour supply to part-time work for

1 year, except for lone parents). This incites both parents to either stay at home with their newborn child, or to both reduce their labour supply.

This paper contributes to this literature by studying the impact of a child's birth on women's and men's hourly wages and supplied hours in France, using difference-in-differences estimations. We focus on the role of education level and maternity (paternity) leave duration on family wage gap. Our results indicate that having a child reduces the labour market participation of educated and non-educated women but not of men. The birth of the first child also has a negative impact on the hourly wages of highly educated women who take a long maternity leave. As Wilner (2016), we do not find any premium in hourly wages for men.

This paper is structured as follows: Section 3.2 presents the data set used, as well as the econometric methodology. Some descriptive statistics are presented in Section 3.3. Section 3.4 explains how groups were built. Section 3.5 discusses the results and Section 4.6 concludes the paper.

3.2 Data and econometric methodology

3.2.1 The DADS-EDP panel

The data set used in this paper is called the "DADS-EDP" which is a panel provided by the INSEE. This "DADS-EDP" data set is actually composed of two other sources: the DADS panel merged to the EDP. [5] The DADS is an administrative data set with information on wages, the type of employment contract, employment sector, the size of the firm the person is working in, the starting and closing dates of the period of paid work, the number of paid hours, etc... Each year, firms have to make a declaration for each of their employees. Every working person in France is covered by these declarations, except for employees of government bodies, self-employed people and employees of French firms established abroad.

Civil servants working in public institutions of an industrial and commercial nature are included in the DADS (since 1991 and 1992) as well as publicly-employed hospital staff (since 1984), civil servants of territorial communities (since 1988), unemployment benefits recipients (since 2002) and agricultural workers (since 2003). There exists different

[5] DADS stands for *Déclaration Annuelle des Données Sociales*, or Annual Declaration of Social data. EDP stands for *Échantillon Démographique Permanent*, or Permanant Demographic sample.

versions of the DADS, for instance a version that includes every civil servant. We have access to the panel version of the DADS (from 1976 to 2010), which is a 1/25th sample of the DADS until 2001 (we have employees born in October in an even year). Since 2002, the sample size was doubled.

The EDP is a panel with information on marital status, fertility, degrees obtained and the place of residence.[6] From 1967, to 2004, people born from the 1st to the 4th of October are in the EDP. Since 2004, the data set was enriched with individuals born from the 2th to the 5th of January as well as those born between the 1st and 4th of April and July.[7] The data are gathered from civil registries each year, and also from the census, whenever needed to complete the information from the civil registries. For example, for people born between 1989 to 1997, the information on their children comes exclusively from the census. For most people born between 1982 and 1989, this is also mostly the case. Before 2004, only people living in continental France were in the EDP. Since 2004, people living in the French overseas territories are also included.

The merged DADS-EDP panel to which we have access through the CASD[8] is composed of individuals born on EDP days that are also in the DADS panel. Therefore, the data set does not include civil servants of national public services, men or women who have never worked and self-employed people. French nationals born abroad are not included in the data.

3.2.2 Data preparation

Before using the data for analysis, a lot of data preparation and cleaning was done. The raw data is in spell format, not very well suited for analysis. We describe briefly how we transformed the data into a panel. The same operations were applied for both women and men, so this subsection only describtes the steps for women.

First of all, we had to order the births of the children. In the description of the data it was written that for each child the variable `aeni` gave the birth year of the i

the year of birth of the oldest child was `aen2` for a woman who had two children, but for a woman who had 3 children it was then `aen3`. Something as simple as computing the mean age of childbearing by cohort for the first child was thus impossible to do with a single line of code. The first step was thus to order the births so that the first child, and the oldest, was in `aen1`, the second child to be born to a woman in variable `aen2` and so on. The second step was to remove obvious errors; for example, we had some women that were born after their children. We completely removed such lines. We filled up incomplete data if possible: for example if `starting date of contract` and the `duration of payment` were both available, but not the `ending date of contact`, it was easy to deduct the `ending date of contract`. We then created two variables that gave the age in years of the individuals (current year minus year of birth) and the number of children someone had at a given year. Until then we only had the total number of children someone ever had.

One of the first important steps was to create a variable that counted the number of days a person stayed out of the labour force due to having a child. The starting and ending days of work contracts are reported, so it was possible for us to compute this variable. For women for whom no date of exit or entry into a firm is indicated around birth, we have deducted the length of maternity leave by the reduction in hours worked or wages in comparison to hours worked and hourly wage of the year preceding the birth. [9] The variable "duration of maternity leave" is useful to compare the impact of the birth of a child for women that had short maternity leaves (less than 6 months after the birth) to mothers who took longer maternity leaves. We refer here to maternity leave as the period of exit from the labour market following the birth of a child. This maternity leave therefore includes, within the framework of this article, statutory maternity and parental leave.

We removed the parallel spells by defining the main employment for each woman and for each year. If a woman has had more than one job in a given year, we only kept the one where she worked the most days. If there were two jobs (or more) where she had worked an equal number of days in, we kept the job with the highest wage. For each job, we also had the worked hours, the number of days worked as well as wages. We summed up all these variables and finally removed every other spell. After these operations, a line in our data set gives, for a given year, a woman's total number of days worked, the total of the earned wages and only her main job.

[9] On request, authors can provide more details on how this variable is calculated.

We also created a variable giving the attained education level in five categories: no information on education, less than high school education, completed high school (or similar degree), 2-3 years of higher education (university or similar institution) and 4-8 years of higher education.[10]

Finally we created hourly wages by dividing the annual wage by the number of hours worked in the year.

3.2.3 Econometric methodology

In order to identify the impact of the birth of a child on hourly wages and the number of hours worked, we use a standard DiD approach, as in Duguet et al. (2015). The following description is fairly standard and can be skipped for readers familiar with DiD.

Card and Krueger (1994) use the DiD method to study the impact of a raise of the minimum wage in New Jersey that occurred on the 1st of April 1992. DiD works by considering two groups, a control group and a treatement group. In experiments, creating such groups is relatively easy: it suffices to assign the treatment (for example, taking a certain medication) randomly. The people that were randomly selected to get the treatment thus become the treated group.

One famous example of randomly assigned treatments in economics are the *Vietnam lotteries*. Angrist (1990) studies the impact of these lotteries on the wages of men and shows that "*[...] as much as ten years after their discharge from service, white veterans who served at the close of the Vietnam era earned substantially less than nonveterans*" (p. 330). But in most cases, however, treatment is rarely assigned randomly in economics.

In cases were random treatment is not possible, quasi-experimental methods such as DiD have proved to provide consistent estimates of the parameters of interest. The problem econometricians face is that they do not observe what would have been the outcome variable (for example, hourly wages) for the control group if it were treated, because it might be that people self-selected into the treatment group. Let us illustrate how DiD works. Let y

$$E[y$$

$E[y$

wrong or unreported. As our study focuses on births of children in 2002-2003, we will present descriptive statistics only for the most recent cohorts.

TABLE 3.1: Age at which women and men are first observed

	Date of birth					
	1964 - 1973		1974 - 1983		1984 - 1993	
	Women	Men	Women	Men	Women	Men
Mean	24.64	24.16	21.19	20.99	18.55	18.12
Standard deviation	6.67	6.72	3.15	3.29	1.47	1.61
1st quartile	19	19	19	19	18	17
Median	31	31	23	23	19	19
3rd quartile	21	21	20	20	19	18
Observations	475092	551005	348546	393392	133769	149177

TABLE 3.2: Age at which women and men are last observed

	Date of birth					
	1964 - 1973		1974 - 1983		1984 - 1993	
	Women	Men	Women	Men	Women	Men
Mean	41.06	41.00	31.27	31.42	23.27	23.14
Standard deviation	3.94	3.93	3.26	3.22	2.24	2.33
1st quartile	39	38	29	29	22	22
Median	44	44	34	34	25	25
3rd quartile	42	42	31	32	24	24
Observations	475092	551005	348546	393392	133769	149177

Table 3.1 and 3.2 give the mean age of the first and last observation, for women and men. People in the data set are first observed when they start working and are last observed whenever they leave the country, go into retirement, or die. The cohorts with the most people are 1964-1973 and the 1974-1983 cohorts.

FIGURE 3.2: Fertility rate in France

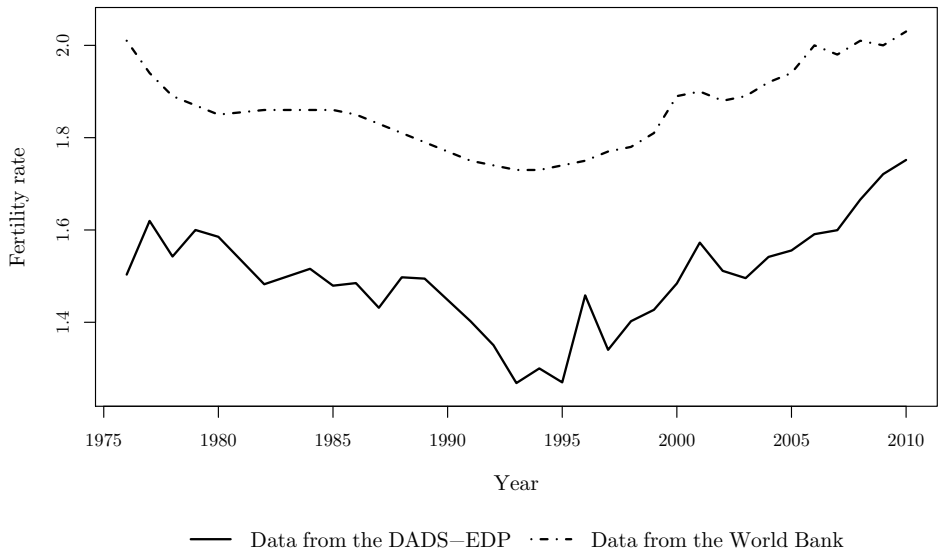


Figure 3.2 shows the fertility rate in France. As can be seen, the true fertility rate is underestimated by our data. The dotted curve is the fertility rate estimated by the World Bank, while the solid curve is the fertility rate estimated by the DADS-EDP. The trend, however, is the same for both curves. Also, as mentioned above, for some years, the information on fertility is missing and completed with data from the census, which counts children currently living with their parents. Children that already left home are not included in the census. Another issue is that we do not have all civil servants in the data nor women who have never been involved in the labour market at all in their life.

FIGURE 3.3: Age at birth for women and across cohorts [11]

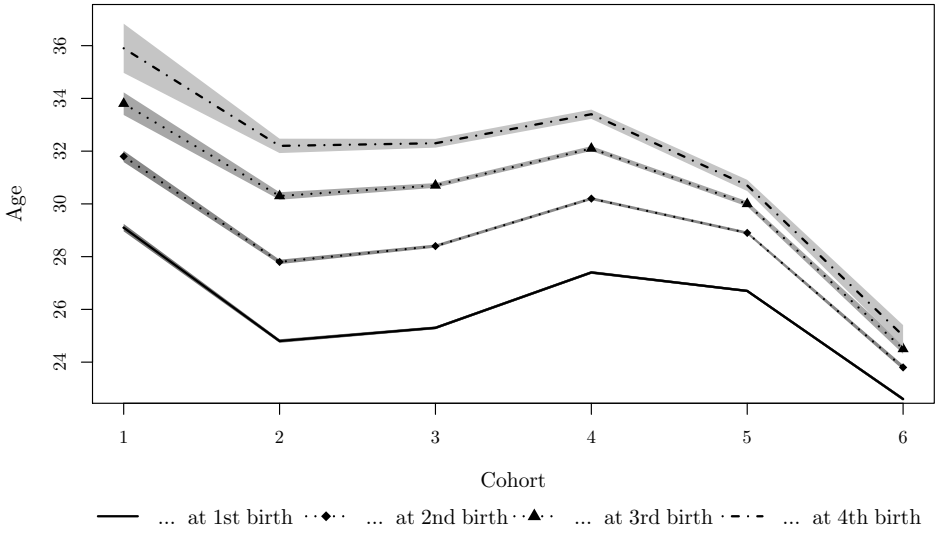
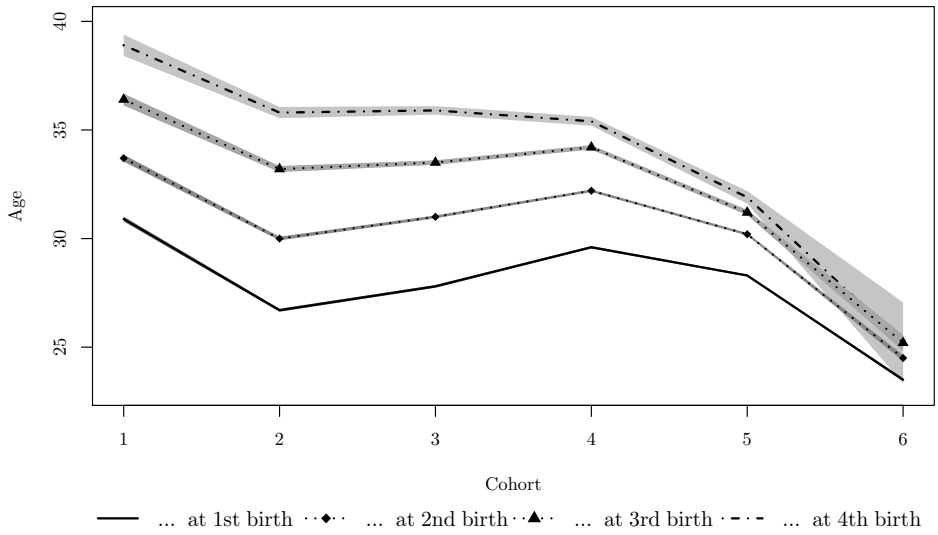


FIGURE 3.4: Age at birth for men and across cohorts



Figures 3.3 and 3.4 show the age at the first, second, third and fourth birth for women and men respectively. Confidence intervals are represented by the grey areas. Ignoring the first and last cohorts, which suffer from censoring, we see that the age at which people have their children has increased. Figures 3.5 and 3.6 show the timing of births

[11] Cohort 1: 1934-1943, cohort 2: 1944-1953, cohort 3: 1954-1963, cohort 4: 1964-1973, cohort 5: 1974-1983, cohort 6: 1984-1993

for different cohorts. The same conclusions can be drawn from these figures and from figure 3.3 and 3.4: younger cohorts are having their children later.

FIGURE 3.5: Timing of births for women by cohorts

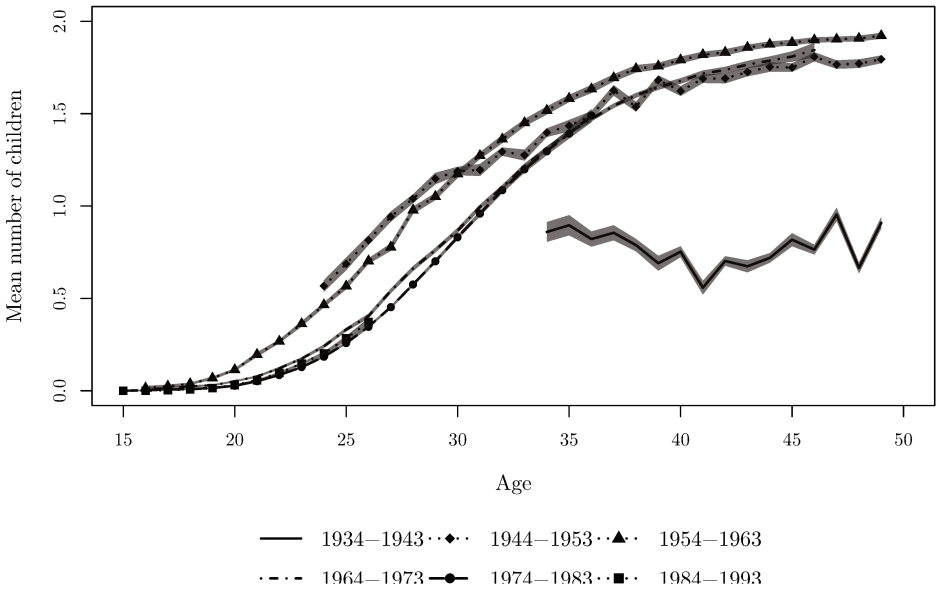


FIGURE 3.6: Timing of births for men by cohorts

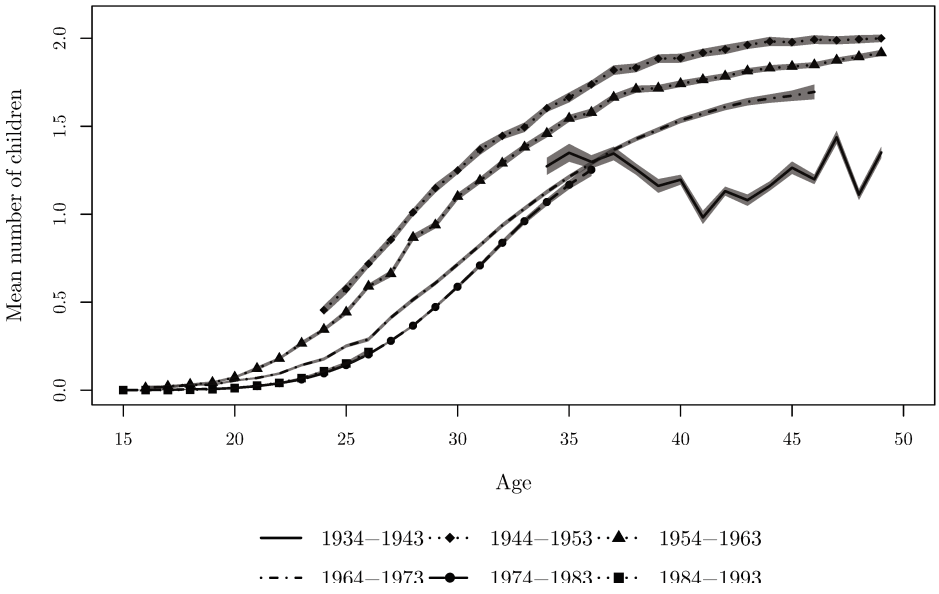


FIGURE 3.7: Timing of births for women by education

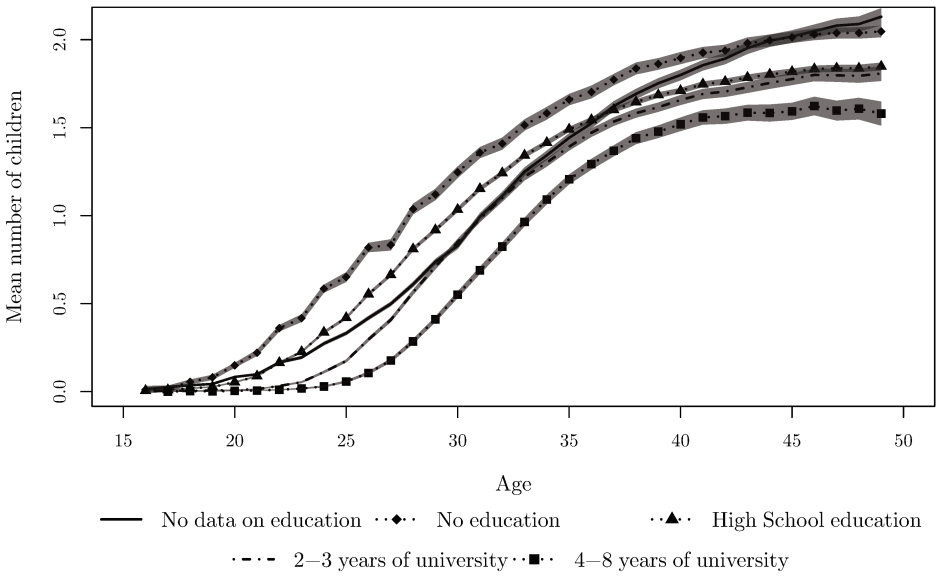
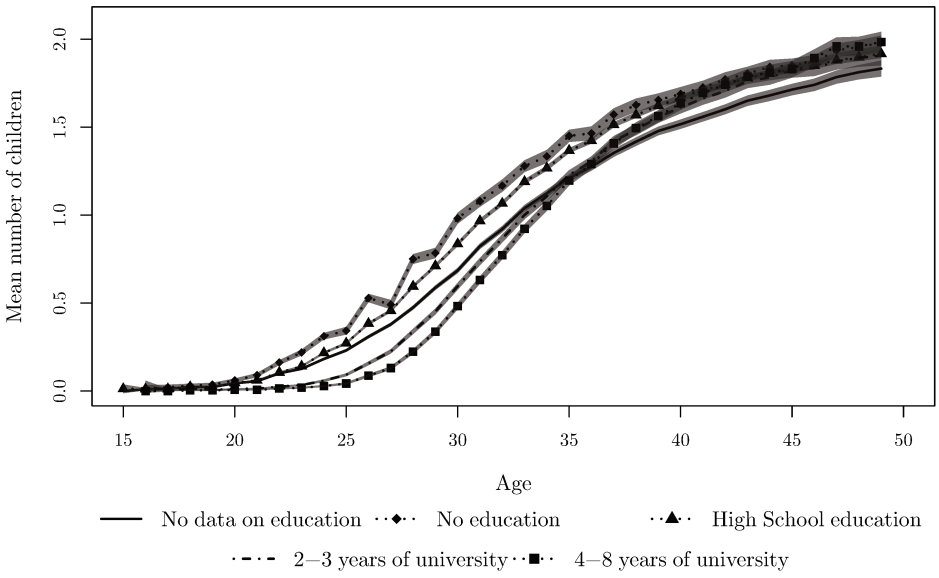


FIGURE 3.8: Timing of births for men by education



Figures 3.7 and 3.8 are both made using the "central cohorts"; due to censoring and low number of observations, we removed the two oldest cohorts and the youngest cohort. These figures suggest that women time childbearing conditionally on their education level, which does not seem to be the case for men.

FIGURE 3.9: Yearly net wage by number of kids for 3 cohorts of women pooled together (1954-1983)

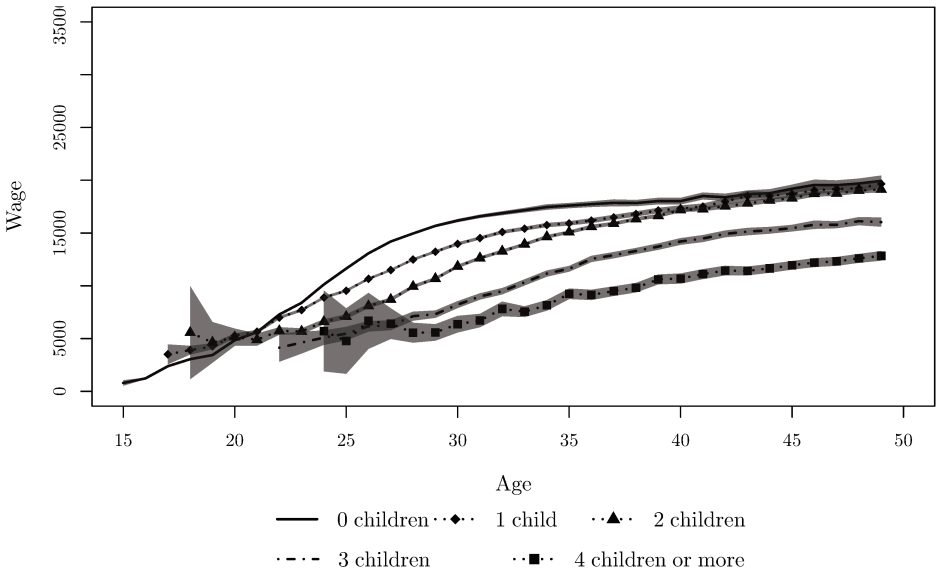
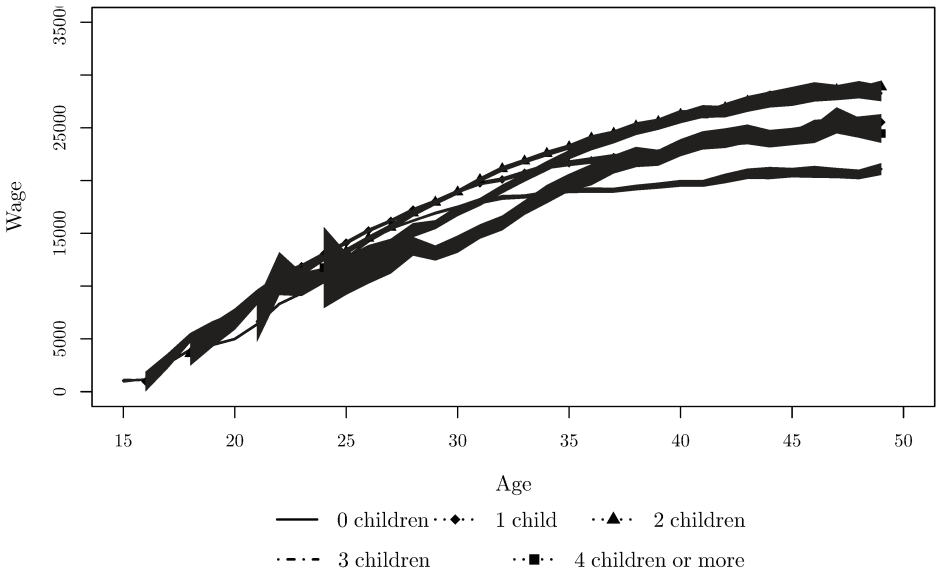


FIGURE 3.10: Yearly net wage by number of kids for 3 cohorts of men pooled together (1954-1983)



Figures 3.9 and 3.10 show the yearly net wage by number of children. The more children a woman has, the less she seems to earn. There seems to be a difference between having no, one or two children for women younger than 40, but then these women catch up to their peers with less children. Starting with three children however, women never seem to catch up. Interestingly, before the women reach the age of 25, there does not seem to

be much selection effect. For men, having children seems to be correlated with higher yearly wages as can be seen on Figure 3.10.[12]

FIGURE 3.11: Share of part-time workers by number of children

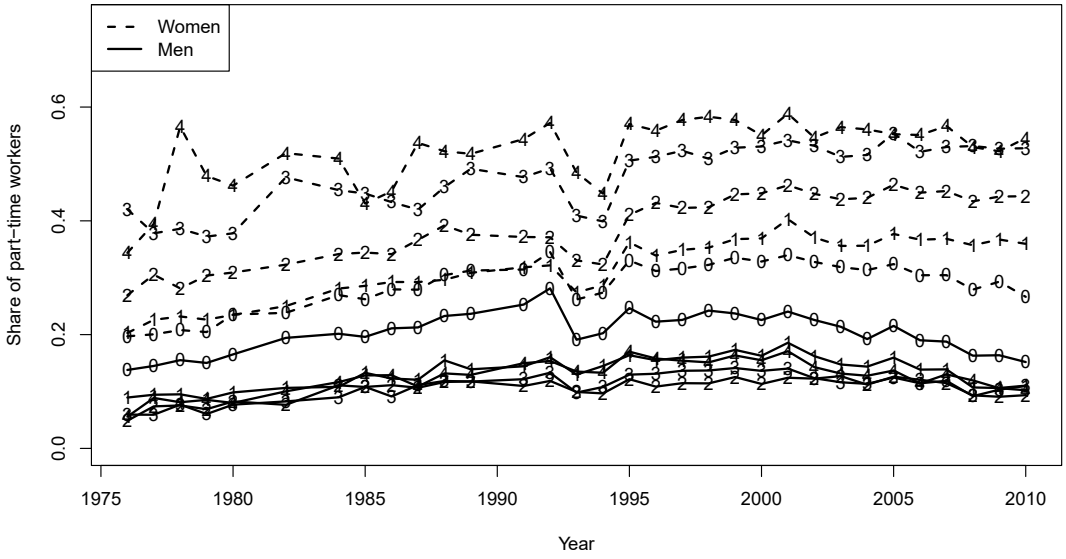


Figure 3.11 shows the share of part-time workers by number of children. The numbers are the number of children women (dashed curve) and men (solid curve) have. Throughout the years this share has remained somewhat stable for both men and women, but the more children women have, the more likely it is that they work part-time. In recent years especially, we see that even after only one child, women seem to decrease their labour supply. For men however, the number of children does not seem to have any impact on part-time work, apart from 0 children. A possible explanation for this might be that men that do not have children tend to be young and thus work part-time to pay for their studies for example.

[12] One might be surprised to see 15 year olds with 4 to 8 years of university education. This is because we only have the final education level in our data. For the purposes of the estimations, we removed people that were too young to have such high degrees.

FIGURE 3.12: Yearly worked hours by number of kids for 3 cohorts of women pooled together (1954-1983)

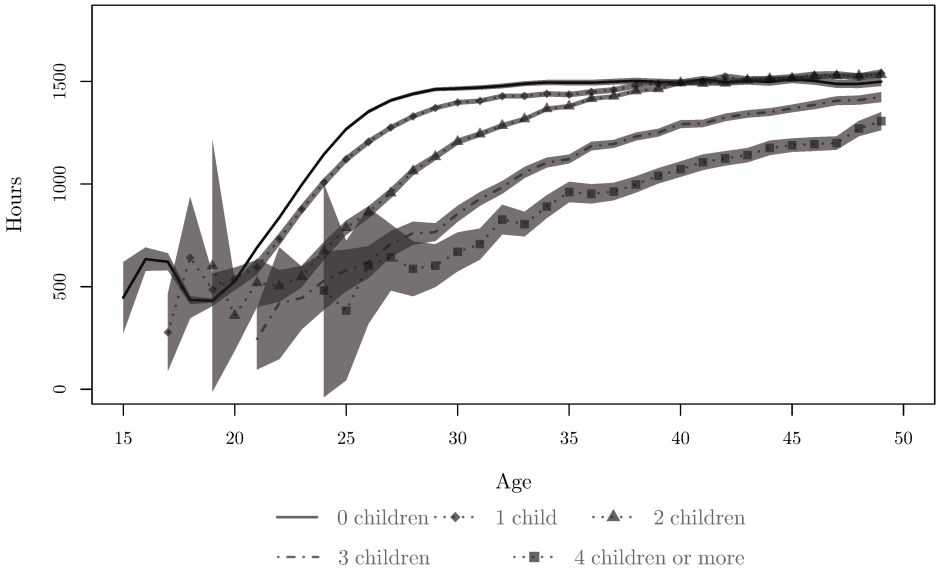


FIGURE 3.13: Yearly worked hours by number of kids for 3 cohorts of men pooled together (1954-1983)

Figures 3.12 and 3.13 show the number of hours worked in a year for women and men by number of children. These last two graphs are very interesting; indeed, having more children seems to be correlated with less worked hours for women, but this is reversed for men.

What do all these graphs and summary statistics tell us? They do seem to go in the same direction as the evidence from the literature. For women, higher education is correlated with later childbearing and having more children is correlated with lower yearly wages. For men, having children does not seem to carry a penalty, quite the contrary. The graphs suggest that having children might be associated with higher yearly wages. As for hours worked, women reduce their labour supply the more children they have, while labour supply seems to increase for men. In the next section, we show how we estimate the impact of the birth of one, two and three children on the following outcome variables: hourly wages and hours worked. We estimate this impact by distinguishing the attained education level of the parents as well as by maternity or paternity leave duration. As stated in the introduction, the literature suggests that higher educated women suffer from a higher penalty in wages. We hypothesise that highly educated women have their children later because they study longer, and thus suffer from a penalty in hourly wages. This paper does not try to give an answer as to why there could be a penalty in hourly wages, but the literature on female labour supply tells us that it may be due to a higher depreciation rate of human capital for highly educated women.

3.4 Control and treatment groups

To analyse the impact of the birth of a child on hours worked and on the hourly wage of a woman or a man, we run a difference-in-differences estimation. This allows us not to have to specify a functional form for the wage equation and also deals with unobserved heterogeneity. It is thus important to discuss how we constructed the control and treatment groups. We only kept individuals that had a birth in either 2002 or 2003. We assume that having a child in 2002 and 2003 is very similar, and thus pooling the individuals is not a problem. We selected both these years in order to increase the number of observations for our analysis. Then, we observed what the hourly wages and worked hours were for these individuals two years before treatment. This counts as our first observation. Then we also observed them again two, four and six years after treatment. The reader might wonder why we chose the years 2002 and 2003 as the treatment years. The first reason is that we wanted to use the most recent data at our disposal. The second reason is the requirement that nothing but the treatment could influence the variables of interest. So if we focus on the 2000's, our options are actually very limited. In the year 2000, the socialist government at the time in France introduced the 35-hour workweek. Before this reform, the normal workweek in France was 39 hours long. We had to be careful not to select individuals before and after this reform, or else

we would have attributed a decrease in hours due to this reform to the birth of a child. Another reform was implemented in 2004, called the *complément de libre choix d'activité* (CLCA). The CLCA is a lump sum, which depends on family resources, given to parents that completely stop, or reduce their labour supply to take care of a child, that must be younger than 3 years old. For their first child, parents received the CLCA during 6 months, and from their second child, they got the CLCA until the third birthday of the last child. Before this reform, only parents of two children or more could receive such an allowance. Here again, we had to be careful, as this reform could have changed the behaviour of people. Since we also wanted to see the labour supply and hourly wages of treated individuals six years after treatment, we could not have chosen the year 2005 (or later) as a treatment year, as our data is available until 2010.

To create our control and treatment groups we separated the individuals by education levels. However, having five categories for the education variable decreased the number of individuals inside the groups too much. We decided to pool individuals together into two categories: having at most a high school degree (or similar) or having more than a high school degree (such as any university degree). We also made sure that people in groups would not have another child before 2011. We separated the treatment and control groups further into women (men) that stayed out of the labour force for less than 6 months, for more than 6 months but less than than 2 and a half years and women (men) that stayed out of the labour force for more than 6 months but less than 6 years. This allows us to distinguish the impact for people that stay out of the labour force for different lengths of time.

Then, to analyse the impact of having a first child, we proceeded in a similar manner to Duguet et al. (2015): the control group is composed of women without children. For the impact of a second child, the control group is made up of women with one child and for the impact of a third birth, the control group is composed of women with two children.

After creating the control and treatment groups, we made sure that the individuals in the control group only differ from those in the treatment group by the fact of not being treated. This means that the individuals composing these groups should be as homogeneous as possible. Thus, we ensured that the age structure between these groups was similar. So we created hundreds of control and treatment groups by selecting individuals that were of a certain age in the year of the first observation. For each pair of group created, we compared the distribution of age using a Kolmogorov-Smirnov test on the empirical distributions of age and a simple t-test to compare the means of ages in both groups and only kept the biggest groups were these tests passed. This means that we

do not have too many observations in each group, but we are fairly confident in the quality of the groups. The impact of the birth on the wages and hours worked might still not be causally determined as the decision to have a child is endogenous. However, by proceeding in the described manner, we removed a lot of potential sources of bias, and by comparing women with one (two) child(ren) to women with two (three) children, we are fairly confident in our results.

After all these steps we ran our DiD analysis for each subgroup. Our variables of interest were *number of hours worked* in a year and *hourly wages*. We bootstrap confidence intervals using Efron's bootstrap percentile t method as described in Chernick (2007). Efron's bootstrap percentile t method is simple to implement and is second-order accurate.

3.5 Results and discussion

Tables 3.3 and 3.4 show the impact of the birth of a first child on hourly wages as well as on yearly worked hours. The impact is shown for 2004, 2006 and 2008 (or 2005, 2008 and 2009 for children born in 2003), so two, four and six years after treatment. We further distinguished between women that went on maternity leave for less than 6 months, more than 6 months but less than 2 years and a half and finally for more than 6 months but less than 6 years. The results are in 2007€. For women with at most a high school degree, the results indicate no penalty in hourly wages, and a premium four years after treatment for women who took a short maternity leave. For women with more than a high school degree, we see a negative impact, but only for women that stayed more than 6 months out of the labour force. For women who stayed out of labour force between 6 months and two and half years, the negative impact of child bearing increases over time. Women that stayed less than 6 months do not suffer any penalty. This might be evidence of human capital depreciation that only occurs if women stay out of the labour force for longer periods of time.

Hours decrease significantly for both educated and non-educated women in the short and in the long run. The impact on labour supply of birth does not seem transitory.

TABLE 3.3: Estimation of the impact of the birth of a child on the hourly wages of women.

		Less than High School	More than High School
Year of 2nd observation	Maternity Leave Duration	Estimate	Estimate
2004	Short	0.33*	−0.33
		[−0.04, 0.69]	[−0.96, 0.25]
		[0.00, 0.64]	[−0.86, 0.11]
2006	Short	0.44**	−0.34
		[0.02, 0.83]	[−1.01, 0.30]
		[0.06, 0.77]	[−0.90, 0.20]
	Intermediate	−0.54	−1.63**
		[−1.69, 1.41]	[−3.24, −0.10]
		[−1.52, 1.20]	[−3.05, −0.39]
2008	Short	−0.20	−0.49
		[−3.41, 1.41]	[−2.02, 0.82]
		[−2.59, 1.16]	[−1.78, 0.64]
	Intermediate	−0.65	−2.22**
		[−2.85, 1.33]	[−3.77, −0.78]
		[−2.47, 1.02]	[−3.52, −1.01]
	Long	−0.08	0.29
		[−2.42, 1.89]	[−3.53, 2.61]
		[−2.02, 1.53]	[−2.95, 2.18]

5% (**) and 10% (*) bootstrapped confidence intervals (in brackets) obtained with 1000 replications.
Size of the groups for lower educated women: *Short Leave* (Treated = 187/Control = 2526), *Intermediate Leave* (Treated = 107/Control = 2768), *Long Leave* (Treated = 145/Control = 2978).
Size of the groups for highly educated women: *Short Leave* (Treated = 149/Control = 1885), *Intermediate Leave* (Treated = 41/Control = 2186), *Long Leave* (Treated = 62/Control = 2360).
Short Leave, *Intermediate Leave* and *Long Leave* "less than 6 months out of the labour force after giving birth", "more than 6 months but less than 2 and a half years" and "more than 6 months but less than 6 years"

Tables 3.5 and 3.6 show the impact of having a second child, relative to women with one child already. Hourly wages seem to be mostly unaffected. There is a negative impact but only for lower educated women four years after treatment who took an intermediate leave. Worked hours have, in most cases, decreased yet again significantly.

Finally, Tables 3.7 and 3.8 show the impact of having a third child, but only for women who took a short maternity leave break (there were very few women that had three children and took a long maternity leave in our data). Hourly wages are unaffected. Supplied hours decrease particularly two years after birth. The reduction in labour supply is also greater among less educated women, whatever the time horizon.

To summarize, whatever the rank of birth, a birth has a negative impact on women's labour supply and especially for lower educated women. As for hourly wages, it is

TABLE 3.4: Estimation of the impact of the birth of a child on hours worked of women.

		Less than High School	More than High School
Year of 2nd observation	Maternity Leave Duration	Estimate	Estimate
2004	Short Leave	-199.89^{**}	-150.16^{**}
		$[-286.81, -115.31]$	$[-246.98, -51.49]$
		$[-269.08, -128.55]$	$[-231.86, -66.90]$
2006	Short Leave	-186.58^{**}	-125.59^{**}
		$[-265.65, -109.85]$	$[-219.13, -27.04]$
		$[-254.48, -121.39]$	$[-207.68, -40.71]$
	Intermediate Leave	-182.97^{**}	-89.48
		$[-316.10, -50.44]$	$[-251.54, 103.28]$
		$[-303.11, -77.67]$	$[-225.31, 71.59]$
2008	Short Leave	-192.47^{**}	-107.83^{**}
		$[-279.63, -100.67]$	$[-207.46, -8.61]$
		$[-267.92, -113.19]$	$[-192.10, -22.37]$
	Intermediate Leave	-124.78^{*}	-150.48^{*}
		$[-257.60, 4.45]$	$[-332.55, 13.34]$
		$[-240.82, -11.69]$	$[-305.41, -14.44]$
	Long Leave	-95.01	-197.21^{**}
		$[-217.05, 25.71]$	$[-379.99, -25.23]$
		$[-195.47, 13.02]$	$[-350.26, -55.86]$

5% (**) and 10% (*) bootstrapped confidence intervals (in brackets) obtained with 1000 replications. Size of the groups for lower educated women: *Short Leave* (Treated = 187/Control = 2526), *Intermediate Leave* (Treated = 107/Control = 2768), *Long Leave* (Treated = 145/Control = 2978). Size of the groups for highly educated women: *Short Leave* (Treated = 149/Control = 1885), *Intermediate Leave* (Treated = 41/Control = 2186), *Long Leave* (Treated = 62/Control = 2360). *Short Leave*, *Intermediate Leave* and *Long Leave* "less than 6 months out of the labour force after giving birth", "more than 6 months but less than 2 and a half years" and "more than 6 months but less than 6 years"

essentially the first birth that has an impact, especially for educated women who have taken an intermediate maternity leave. Births of higher rank do not have a significant additional impact.

For the fathers, not many results were statistically significant. So instead of showing uninteresting tables, we discuss the few results that were significant.

A first birth significantly decreases the hours of work of highly educated fathers who had paternity leave longer than 6 months but shorter than two and a half years, 4 years after the birth, by -411.43^{-709}

TABLE 3.5: Estimation of the impact of the birth of a second child on the hourly wages of women.

		Less than High School	More than High School
Year of 2nd observation	Maternity Leave Duration	Estimate	Estimate
2004	Short Leave	0.55	0.86*
		[−0.13, 1.21]	[−0.11, 1.72]
		[−0.03, 1.13]	[0.06, 1.61]
2006	Short Leave	0.77	0.37
		[−0.62, 1.88]	[−0.64, 1.25]
		[−0.47, 1.67]	[−0.45, 1.12]
	Intermediate Leave	−1.11**	−0.23
		[−2.14, −0.06]	[−1.92, 1.17]
		[−1.95, −0.21]	[−1.62, 0.97]
2008	Short Leave	−0.67	0.71
		[−3.09, 1.40]	[−0.67, 1.93]
		[−2.66, 1.00]	[−0.43, 1.69]
	Intermediate Leave	−0.56	2.32
		[−3.57, 1.83]	[−6.34, 7.62]
		[−3.05, 1.44]	[−5.10, 6.92]
	Long Leave	−0.67	−0.16
		[−4.10, 1.80]	[−4.89, 3.97]
		[−3.44, 1.33]	[−4.24, 3.44]

5% (**) and 10% (*) bootstrapped confidence intervals (in brackets) obtained with 1000 replications.
Size of the groups for lower educated women: *Short Leave* (Treated = 116/Control = 452), *Intermediate Leave* (Treated = 42/Control = 510), *Long Leave* (Treated = 91/Control = 548).
Size of the groups for highly educated women: *Short Leave* (Treated = 83/Control = 113), *Intermediate Leave* (Treated = 16/Control = 124), *Long Leave* (Treated = 28/Control = 132).
Short Leave, *Intermediate Leave* and *Long Leave* "less than 6 months out of the labour force after giving birth", "more than 6 months but less than 2 and a half years" and "more than 6 months but less than 6 years"

after the birth, for long paternity leaves the increase is 167.24 [30]

TABLE 3.6: Estimation of the impact of the birth of a second child on hours worked of women.

		Less than High School	More than High School
Year of 2nd observation	Maternity Leave Duration	Estimate	Estimate
2004	Short Leave	-505.44^{**}	-486.12^{**}
		$[-670.91, -348.13]$	$[-674.61, -306.94]$
		$[-642.21, -375.64]$	$[-641.41, -324.96]$
2006	Short Leave	-214.79^{**}	-371.42^{**}
		$[-355.79, -70.47]$	$[-538.96, -187.27]$
		$[-338.87, -92.36]$	$[-515.32, -219.20]$
2006	Intermediate Leave	-101.79	-408.73^{**}
		$[-314.75, 120.56]$	$[-923.83, -3.61]$
		$[-284.84, 86.90]$	$[-857.26, -71.74]$
2008	Short Leave	-210.44^{**}	-304.84^{**}
		$[-339.95, -58.27]$	$[-480.38, -130.42]$
		$[-319.60, -85.01]$	$[-450.04, -166.12]$
	Intermediate Leave	-137.79	-352.86
		$[-375.53, 104.86]$	$[-864.69, 85.01]$
		$[-337.66, 59.79]$	$[-780.07, 7.51]$
	Long Leave	-78.40	-340.45^{*}
		$[-240.02, 71.28]$	$[-669.78, 5.23]$
		$[-206.98, 53.00]$	$[-617.21, -52.24]$

5% (**) and 10% (*) bootstrapped confidence intervals (in brackets) obtained with 1000 replications.
Size of the groups for lower educated women: *Short Leave* (Treated = 116/Control = 452), *Intermediate Leave* (Treated = 42/Control = 510), *Long Leave* (Treated = 91/Control = 548).
Size of the groups for highly educated women: *Short Leave* (Treated = 83/Control = 113), *Intermediate Leave* (Treated = 16/Control = 124), *Long Leave* (Treated = 28/Control = 132).
Short Leave, *Intermediate Leave* and *Long Leave* "less than 6 months out of the labour force after giving birth", "more than 6 months but less than 2 and a half years" and "more than 6 months but less than 6 years"

after the birth of a child and/or leave the labour market for a relatively long period. Our results point to evidence that women then suffer a depreciation of their human capital. This depreciation is especially large for women who have accumulated human capital the fastest, namely, educated women. Also, over time the difference in remuneration between mothers and women without children increases. This would reflect a slower wage increase for the most educated women. Moreover, it seems that the negative impact of birth on hourly wages is accentuated over time, reflecting the slower career progression for educated mothers. This *human capital explanation* could explain a great part of our results. Mothers could also choose a job with less constraints but with a smaller hourly wage and/or less professional opportunities. These family-friendly firms are in general low wage firms, as pointed out by Wilner (2016). Mothers could also be less productive because they spend a lot of time and energy to take care of children. For most men, labour supply as well as the hourly wage are not impacted by the birth of

TABLE 3.7: Estimation of the impact of the birth of a third child on the hourly wages of women.

		Less than High School	More than High School
Year of 2nd observation	Maternity Leave Duration	Estimate	Estimate
2004	Short	1.02^{**}	0.55
		$[0.09, 1.96]$	$[-1.67, 2.51]$
		$[0.24, 1.78]$	$[-1.27, 2.14]$
2006	Short	1.24^{*}	0.82
		$[-0.16, 2.32]$	$[-1.98, 3.58]$
		$[0.04, 2.10]$	$[-1.47, 3.10]$
2008	Short	0.47	-0.27
		$[-1.77, 2.21]$	$[-4.39, 3.26]$
		$[-1.60, 1.87]$	$[-3.69, 2.63]$

5% (**) and 10% (*) bootstrapped confidence intervals (in brackets) obtained with 1000 replications.
Size of the groups for lower educated women: *Short Leave* (Treated = 27/Control = 2576)
Size of the groups for highly educated women: *Short Leave* (Treated = 27/Control = 1024)
Short Leave, "less than 6 months out of the labour force after giving birth"

TABLE 3.8: Estimation of the impact of the birth of a third child on hours worked of women.

		Less than High School	More than High School
Year of 2nd observation	Maternity Leave Duration	Estimate	Estimate
2004	Short Leave	-524.70^{**}	-373.78^{**}
		$[-747.12, -263.07]$	$[-665.03, -113.99]$
		$[-712.59, -304.91]$	$[-618.56, -153.72]$
2006	Short Leave	-247.41^{**}	-171.20^{*}
		$[-362.58, -128.44]$	$[-397.09, 26.62]$
		$[-338.66, -147.48]$	$[-365.01, -6.68]$
2008	Short Leave	-377.91^{**}	-42.14
		$[-563.13, -150.60]$	$[-252.16, 139.03]$
		$[-537.05, -196.52]$	$[-208.86, 116.00]$

5% (**) and 10% (*) bootstrapped confidence intervals (in brackets) obtained with 1000 replications.
Size of the groups for lower educated women: *Short Leave* (Treated = 27/Control = 2576)
Size of the groups for highly educated women: *Short Leave* (Treated = 27/Control = 1024)
Short Leave, "less than 6 months out of the labour force after giving birth"

their children. These gender differences seem to indicate an important specialization of tasks within French households. One of the goals of the new policy put in place by the French government in 2015 called *Prestation partagée d'éducation de l'enfant* (PreParE), which reduces the allowances for parental leave of the mother in favour of the father, is to better spread the effect of a birth between men and women.

3.6 Conclusion

Industrialised countries want to solve three seemingly contradicting problems; (i) increase the labour market participation of women, as well as (ii) increase the fertility rates of families but also (iii) reduce the family pay gap. In this article, we sought to test the hypothesis that a birth negatively impacts the career of a woman and positively the career of a man. We distinguished the impact by different education levels and by maternity (paternity) leave duration. We focused on France, a country with a long history of generous social policies and high fertility rates. Our results show that, whatever the education level, women decrease their supplied hours. Highly educated mothers who take a long maternity leave have a lower hourly wage compared to similar childless women. This could be evidence for human capital depreciation, but this is outside the scope of this paper. In future work, we plan on investigating human capital depreciation of highly educated women in France. Our empirical results confirm that the loss in wages after the birth of a child is proportional to the education level of the mothers. This result raises the question of whether the compensation should be higher for this group of women (as is the case in Germany). These women suffer a greater loss in wages, and thus have less children. This was not an important issue some decades ago, as highly educated women were not numerous, but nowadays, more and more women have university degrees. Our results do not show that men systematically receive a bonus after the birth of a child. These results are in line with those of Wilner (2016). In some cases though, lower educated men do increase their supplied hours. This is probably to compensate the loss in wages of their wives.

3.7 Appendix: data cleaning[13]

Before selecting people for our study, we cleaned the DADS-EDP data (2010). The following list details exactly what we did:

- Creation of a variable : total number of children.
- Creation of a variable : cumulated number of children for each year.
- Pooling the education variable into 5 categories for the descriptive statistics:
    1. Missing information
    2. Low education (Lower than High school diploma)
    3. Middle education (High school diploma)
    4. Low university degree (2 or 3 years after High school)
    5. High university degree (4 years or more after High school)
- Pooling the education variable into two categories for the DiD:
    1. Having at most a high school degree
    2. Having more a high school degree
- Deletion of individuals born after their children.
- Parallel spells: each line of the data set corresponds to a person, a year and a job. Thus, if a person has multiple jobs in the same year, it will have multiple lines for this year. To keep only one line per individual and per year we computed the following variables for each individuals:
    1. Number of hours worked in the year. (Whatever the number of jobs)
    2. Number of days of work
    3. Sum of the gross wage
    4. Sum of the net wage (after social contributions but before income tax)
    5. Sum of the net wage with fringe benefits
    6. Sum of the fringe benefits

[13] Data munging was made possible thanks to the dplyr package (Wickham and Francois, 2015). Tables were made thanks to (Hlavac, 2015).

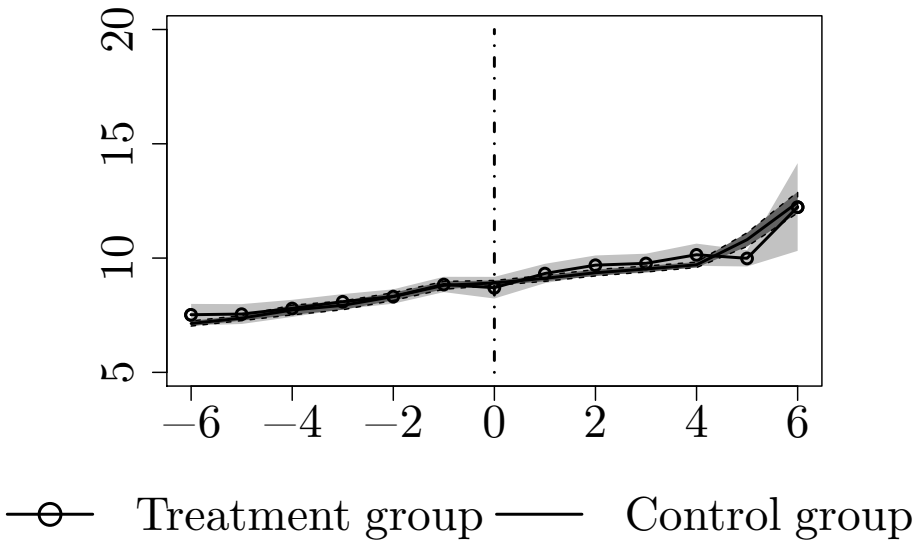
- After computing these variables, we wanted to keep one line per individual and per year. In case there are multiple lines per person and per year we have followed the following rules :
  1. Keep the line with the most days of work
  2. If number of working days are equal to several positions in the same year, we keep the job with the highest wage
  3. If the number of days of work and the wage are equal for various positions in the same year, we randomly keep one line per individual and per year
- We created hourly wages by dividing the annual wage by hours of work
- We deleted observations with extremely high hourly wages or extremely high worked hours

3.7.1 Appendix 2: Common trend graphs

The following graphs show the pre and post-treatment trends for our outcome variables of interest. Gray and dark gray areas are 95% confidence intervals around the mean for the treatment and control groups respectively. Lower educated women (men) are on the left, and higher educated women (men) are on the right.

3.7.1.1 Impact of the first child

FIGURE 3.14: Hourly wages for women, less than 6 months out of labor force after giving birth



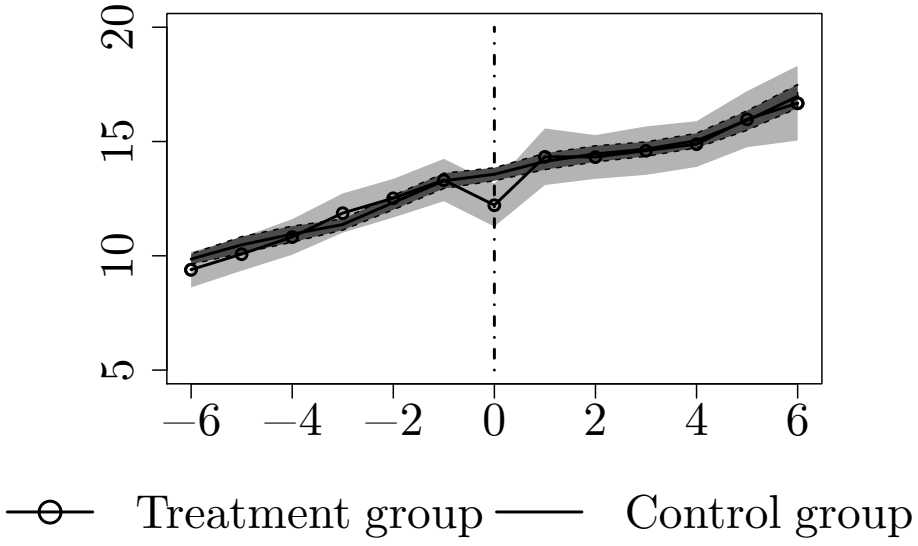
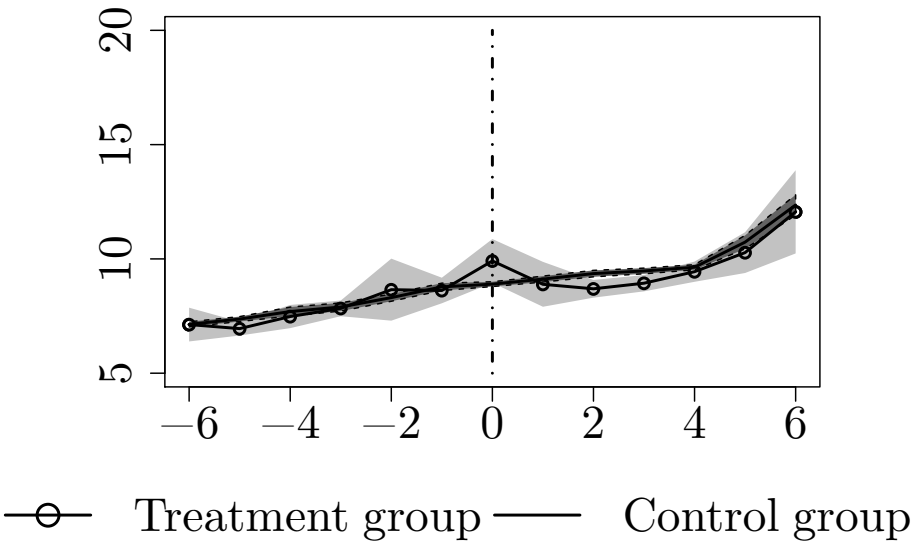


FIGURE 3.15: Hourly wages for women, more than 6 months but less than 2 and a half years out of labor force after giving birth



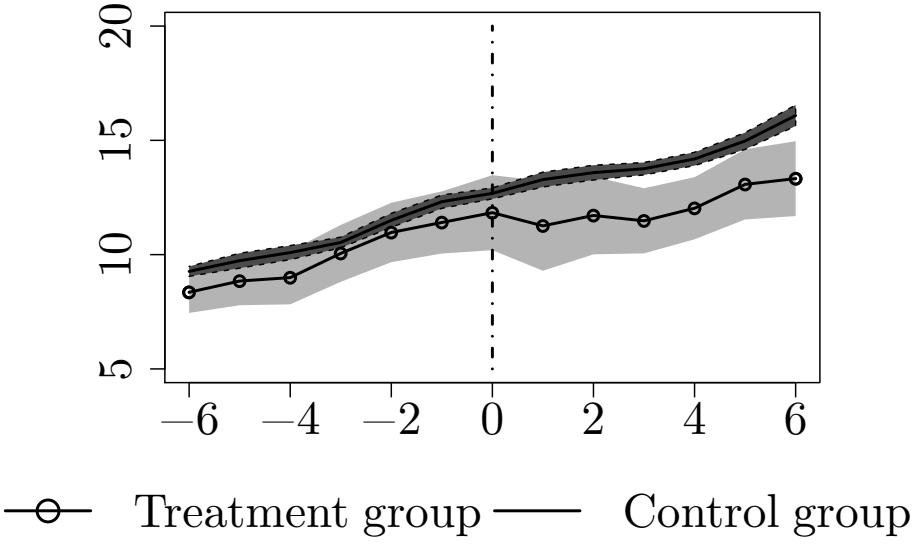
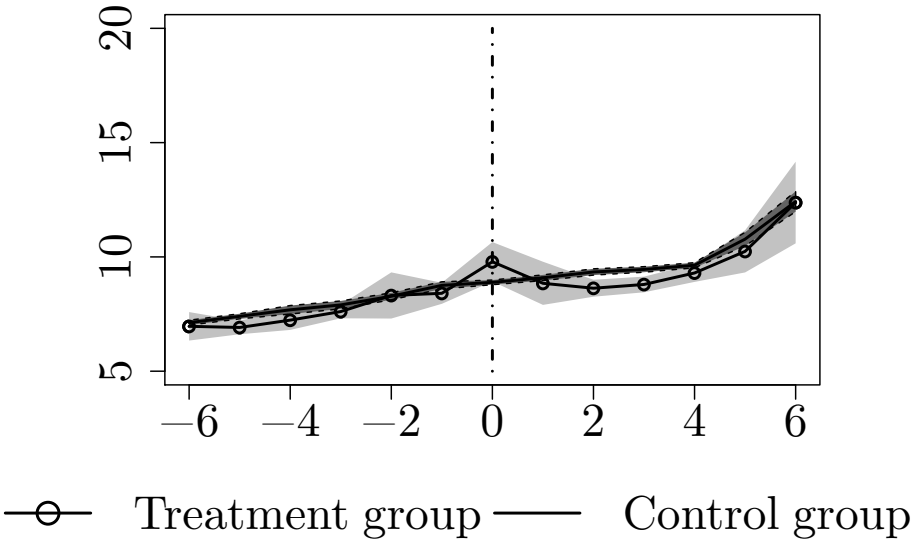


FIGURE 3.16: Hourly wages for women, more than 6 months but less than 6 years out of the labor force



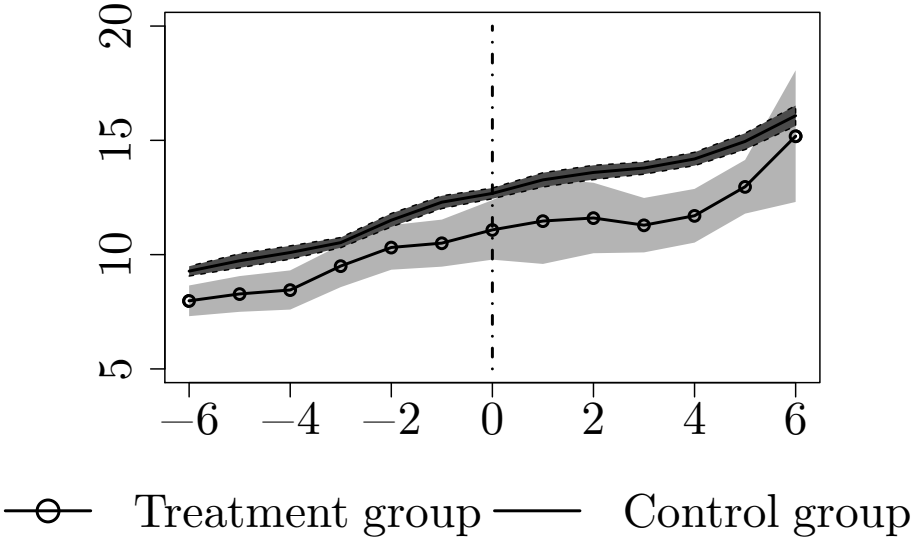
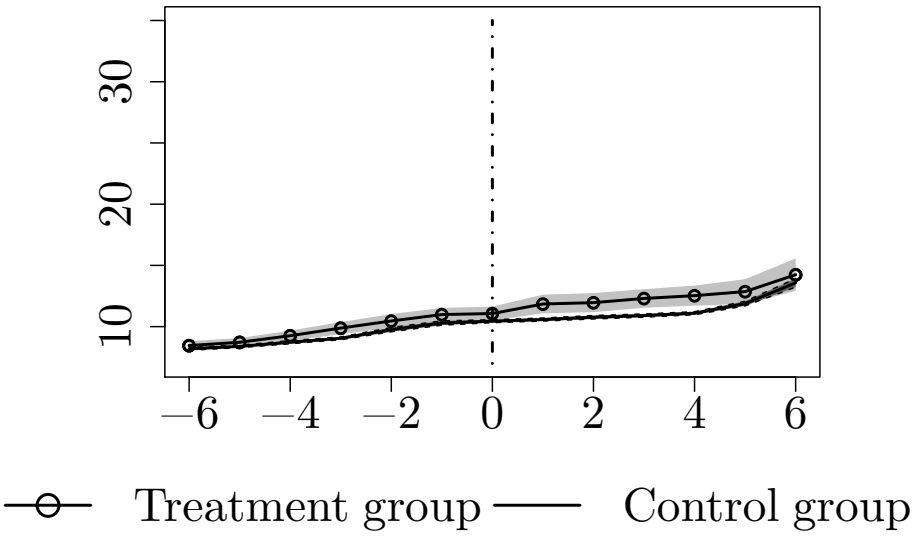


FIGURE 3.17: Hourly wages for men, less than 6 months out of labor force after birth



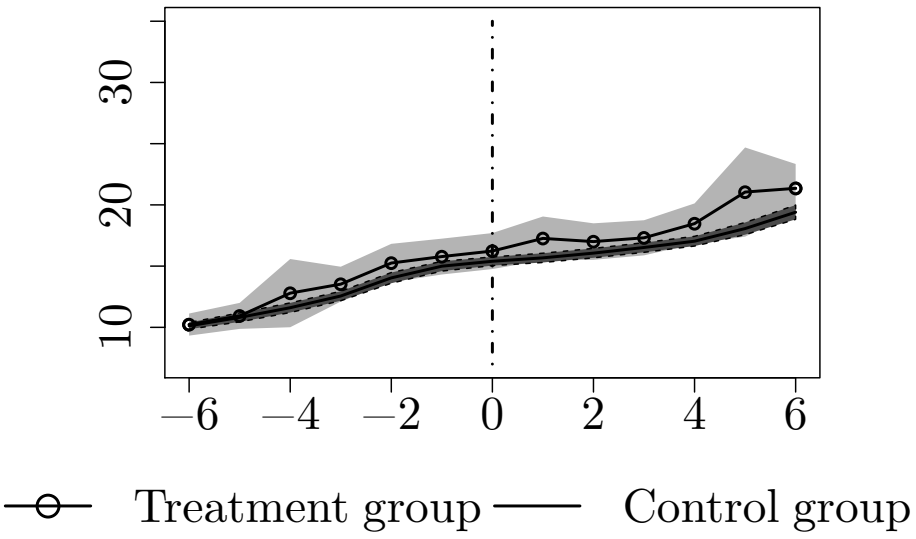
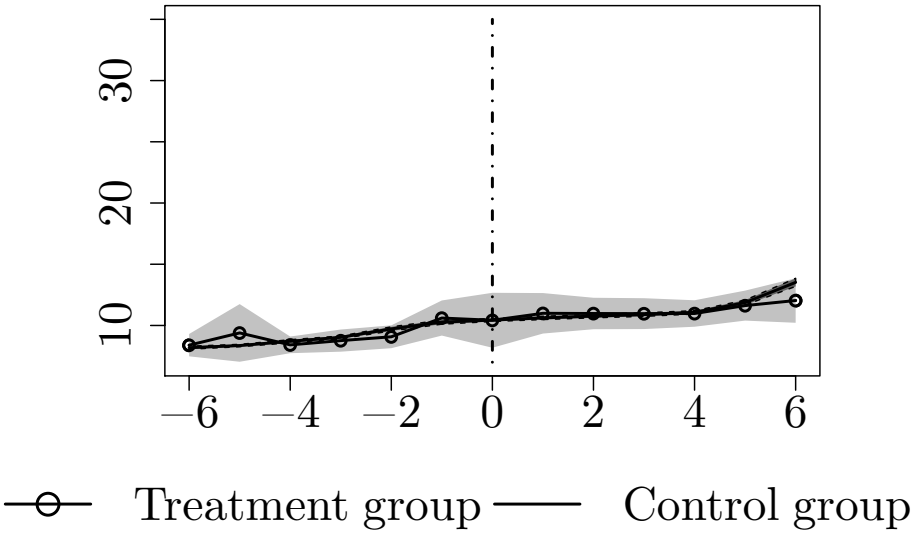


FIGURE 3.18: Hourly wages for men, more than 6 months but less than 2 and a half years out of labor force after birth



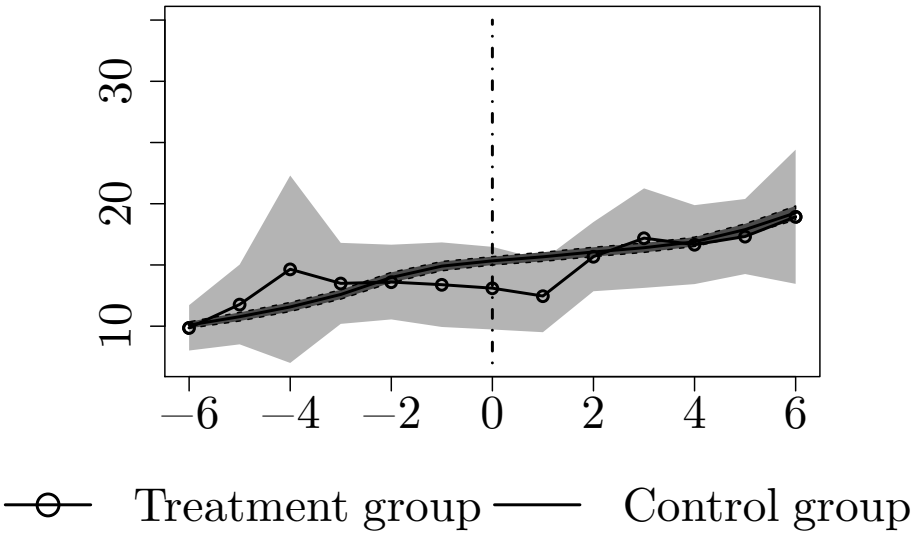
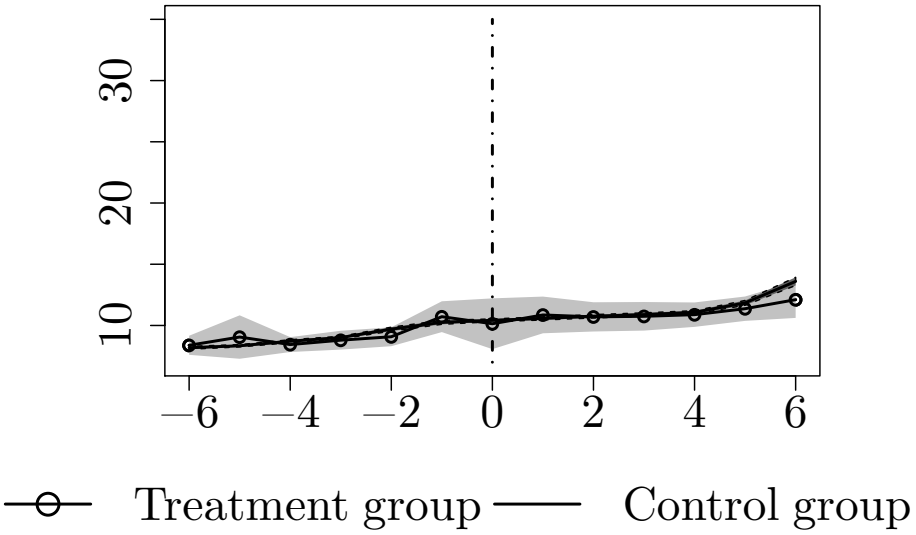


FIGURE 3.19: Hourly wages for men, more than 6 months but less than 6 years out of the labor force



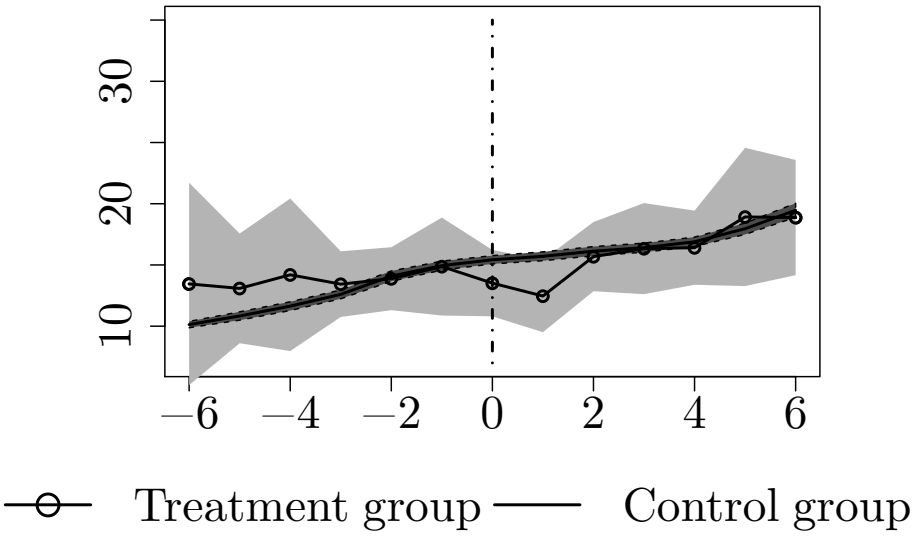
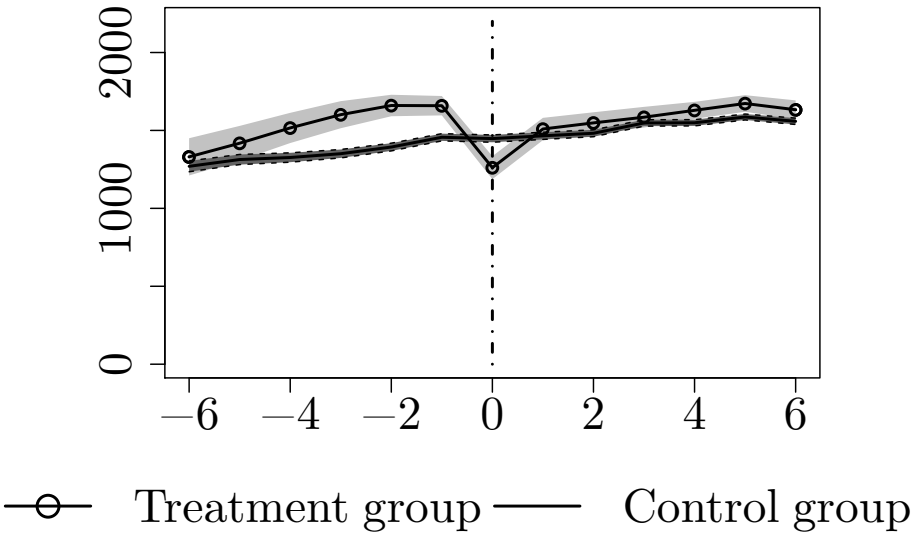


FIGURE 3.20: Worked hours for women, less than 6 months out of labor force after giving birth



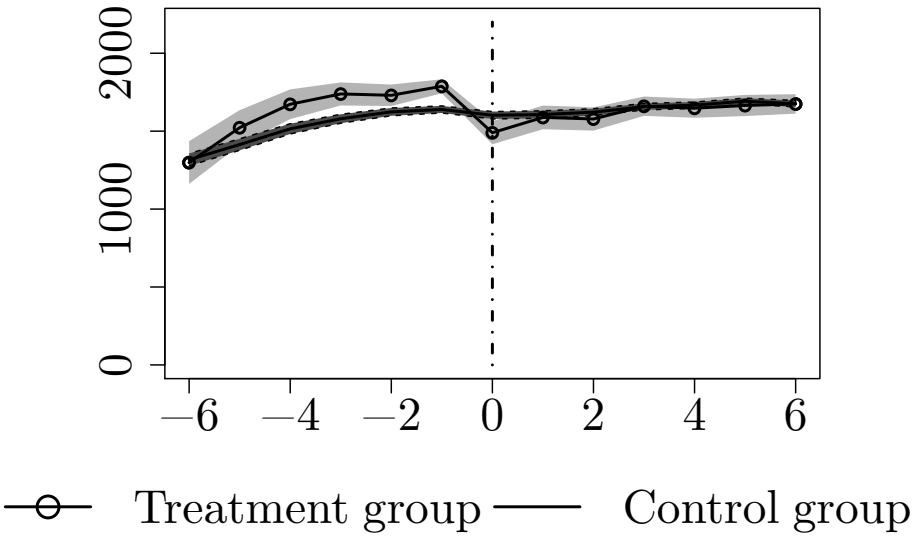
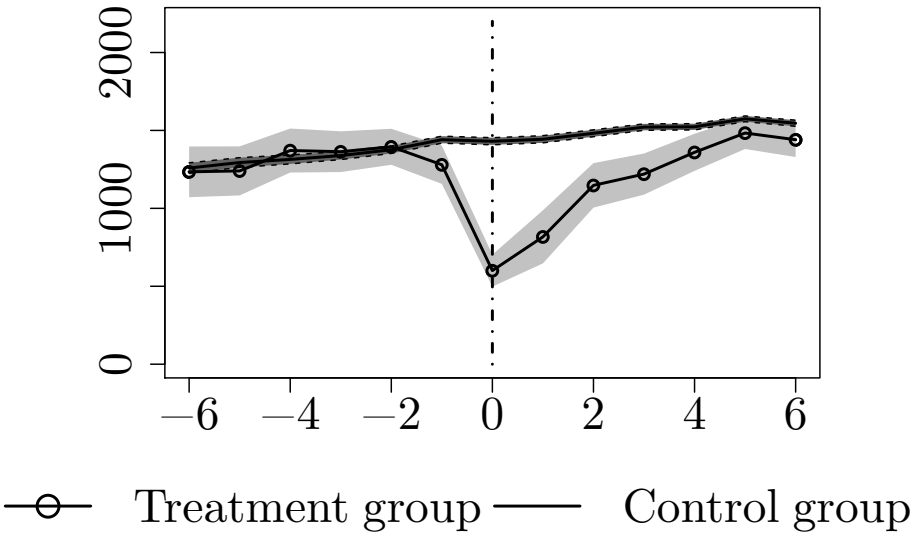


FIGURE 3.21: Worked hours for women, more than 6 months but less than 2 and a half years out of labor force after giving birth



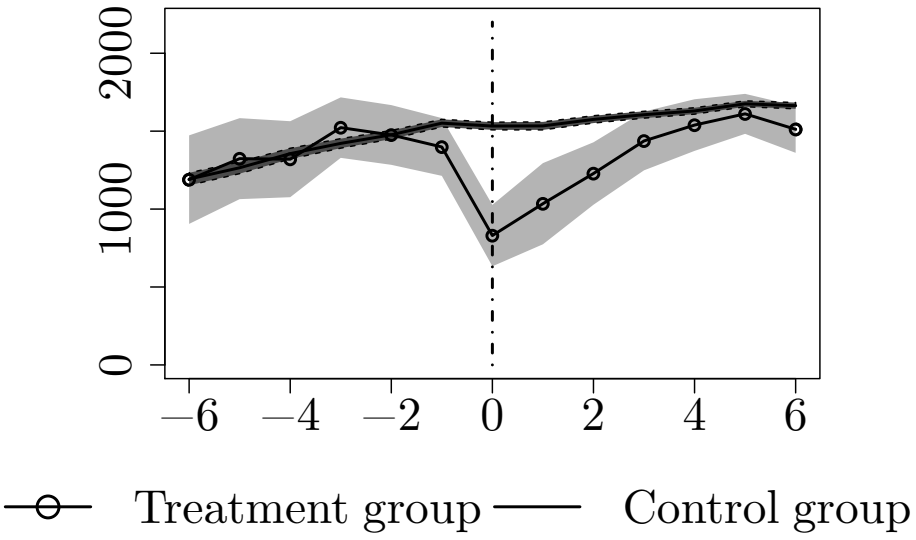
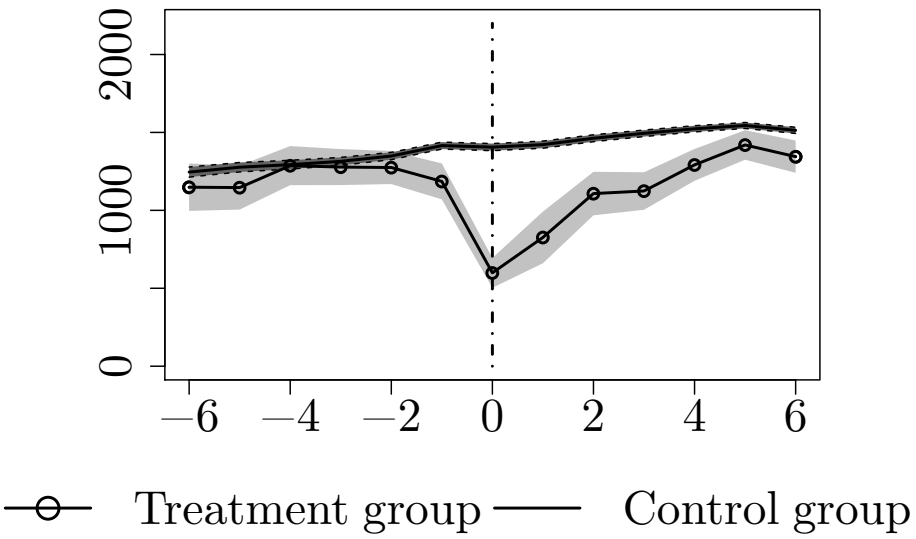


FIGURE 3.22: Worked hours for women, more than 6 months but less than 6 years out of the labor force



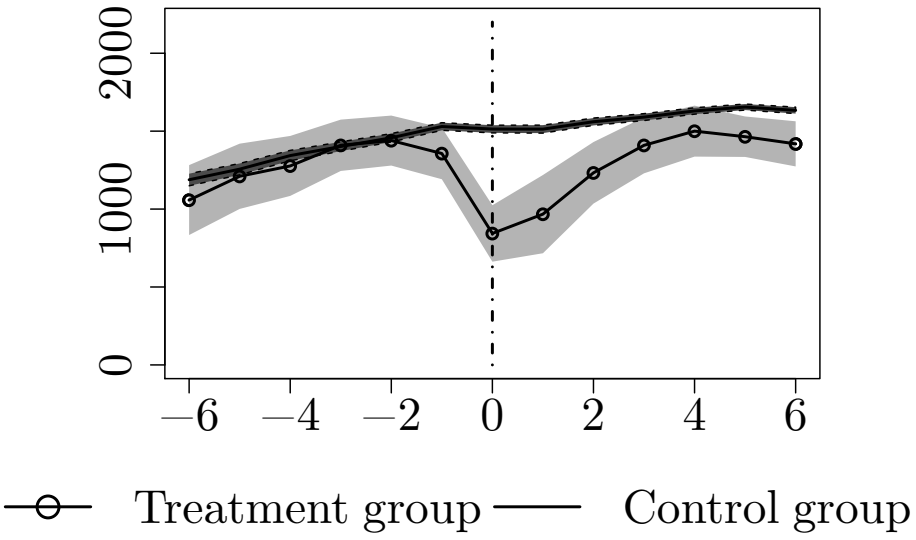
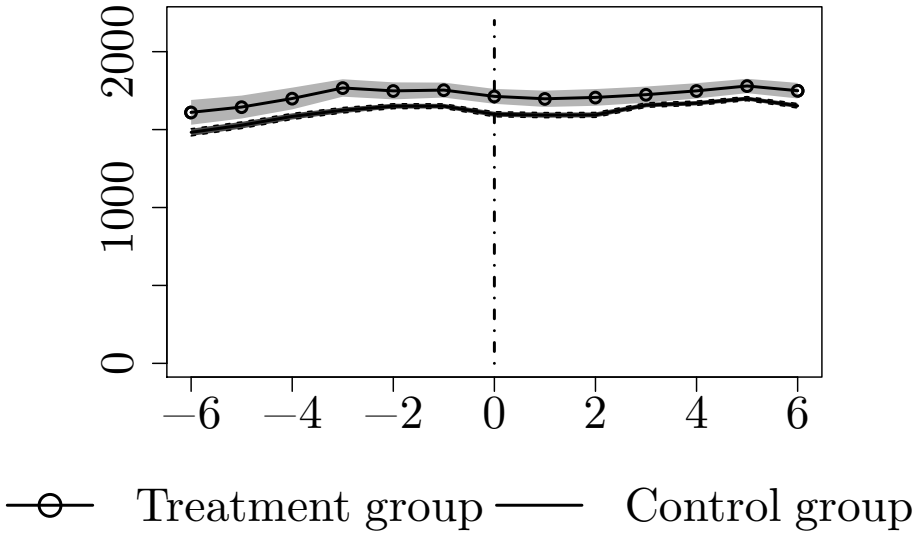


FIGURE 3.23: Worked hours for men, less than 6 months out of labor force after birth



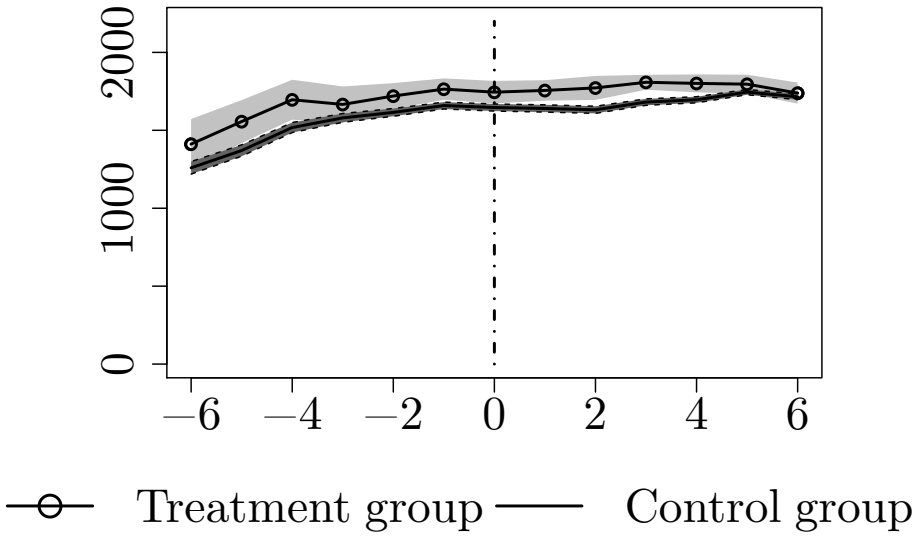
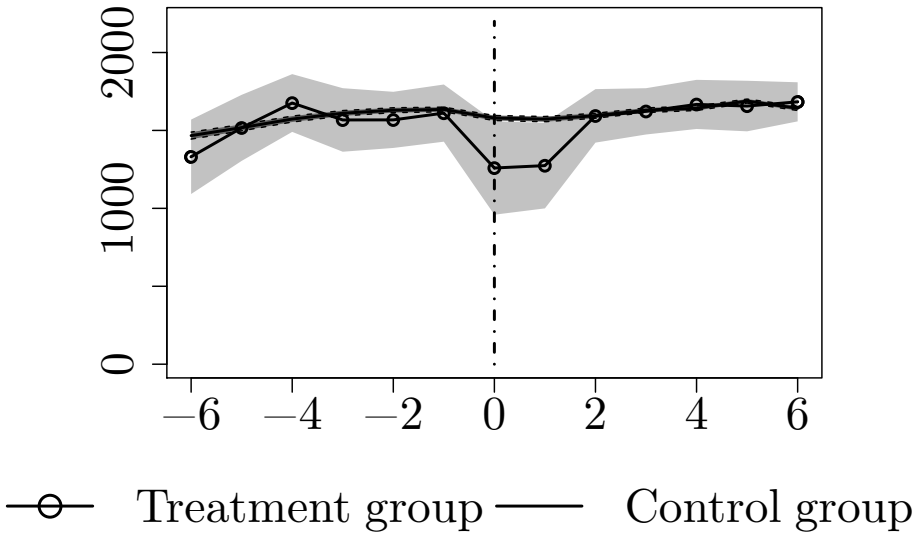


FIGURE 3.24: Worked hours for men, more than 6 months but less than 2 and a half years out of labor force after birth



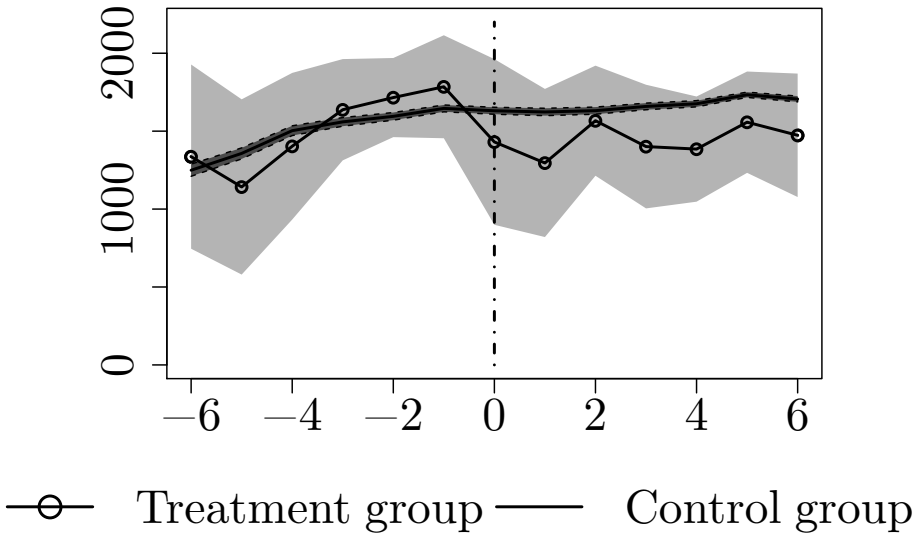
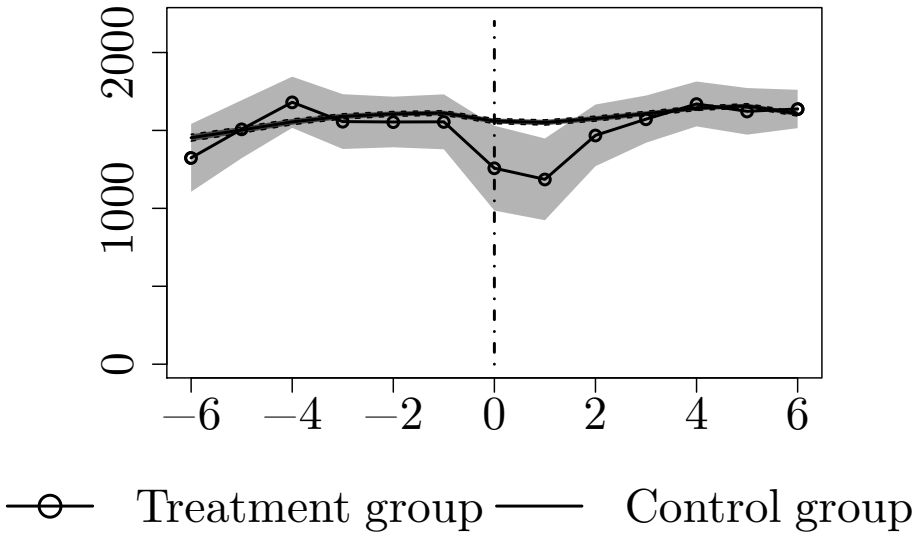
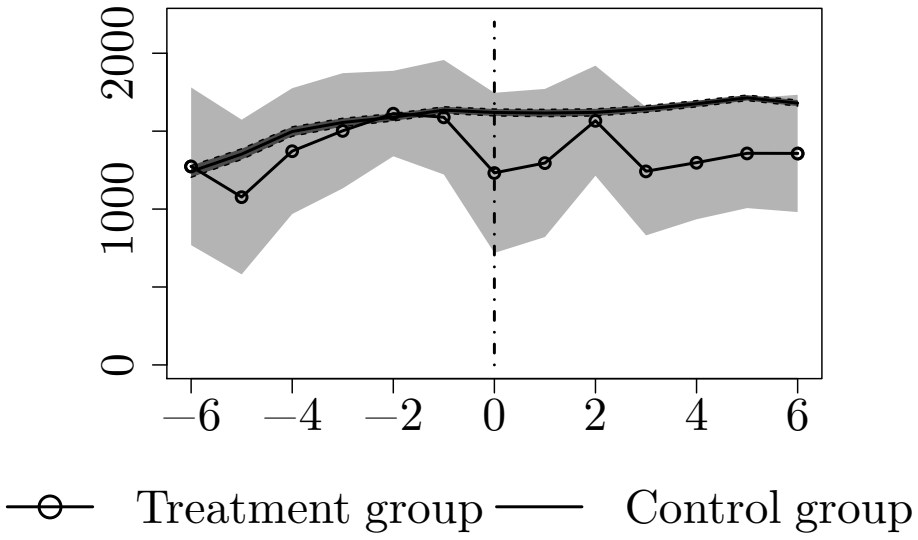


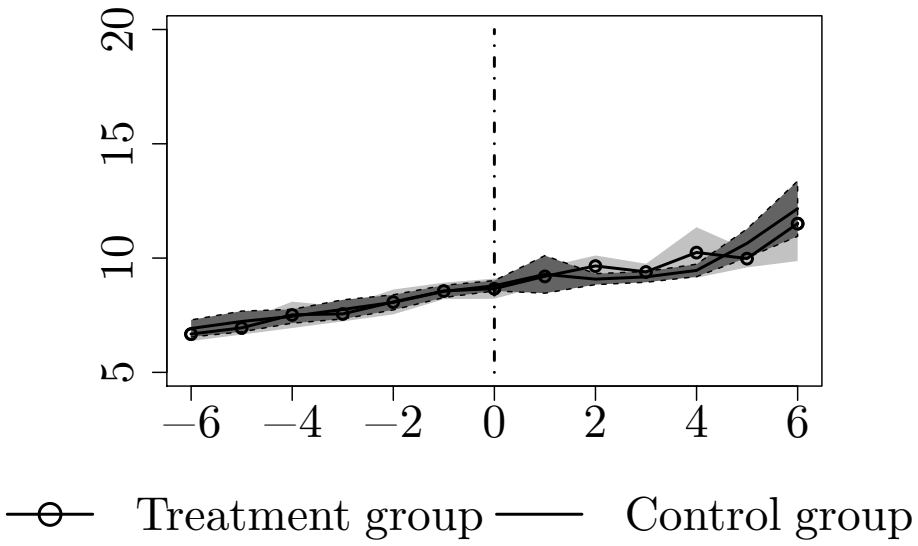
FIGURE 3.25: Worked hours for men, more than 6 months but less than 6 years after birth





3.7.1.2 Impact of the second child

FIGURE 3.26: Hourly wages for women, less than 6 months out of labor force after giving birth



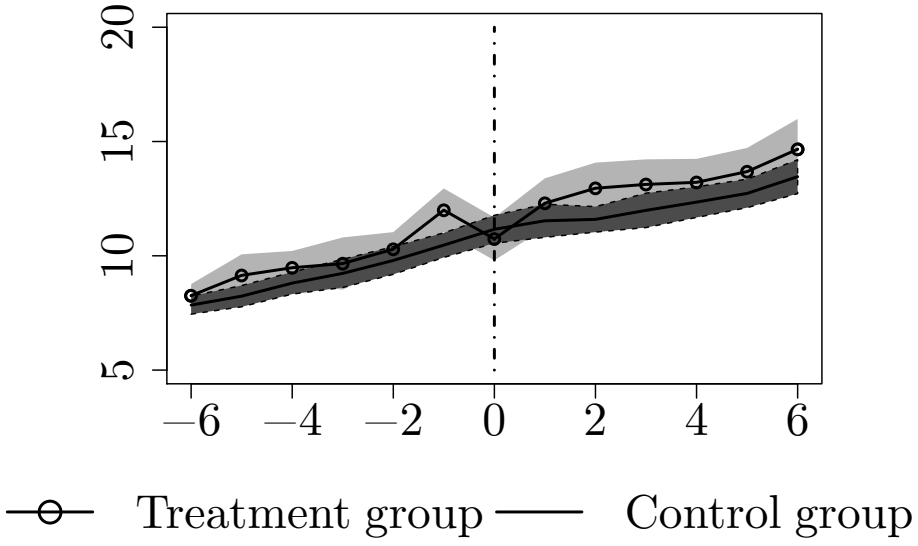
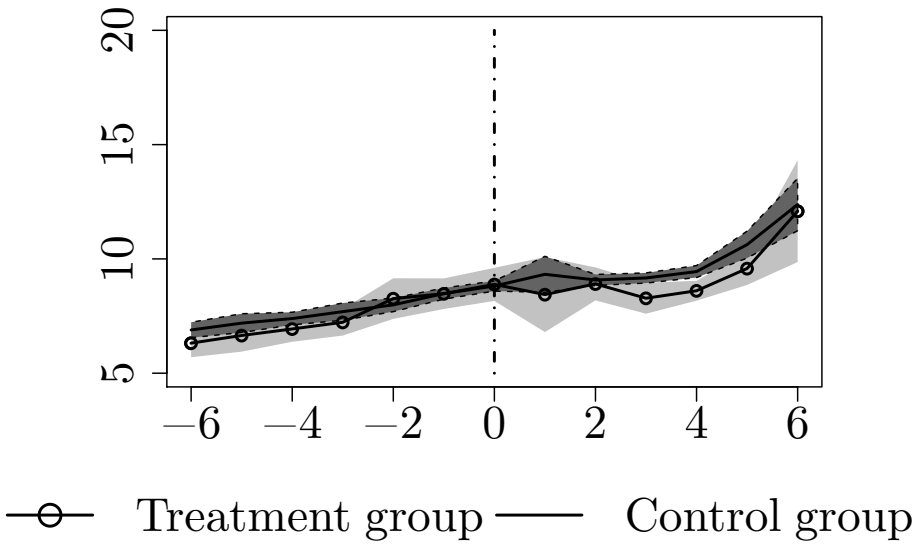


FIGURE 3.27: Hourly wages for women, more than 6 months but less than 2 and a half years out of labor force after giving birth



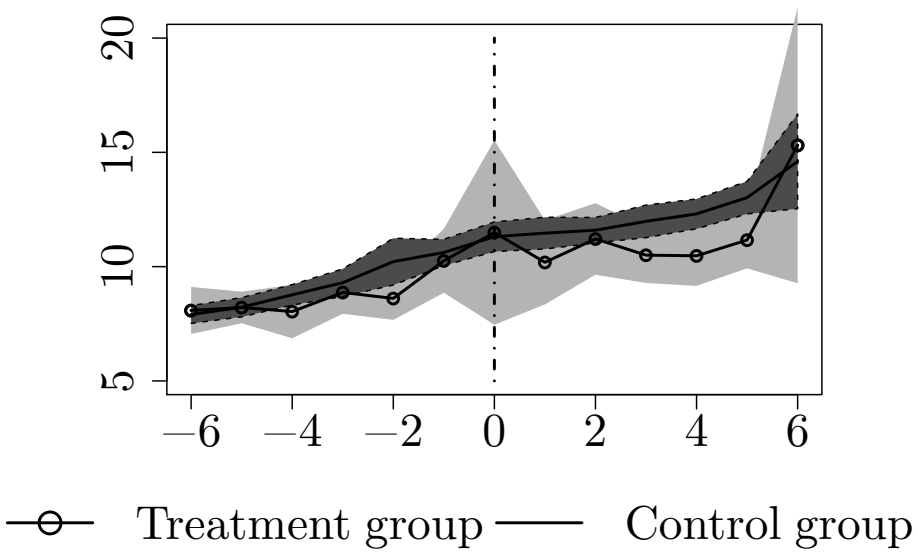
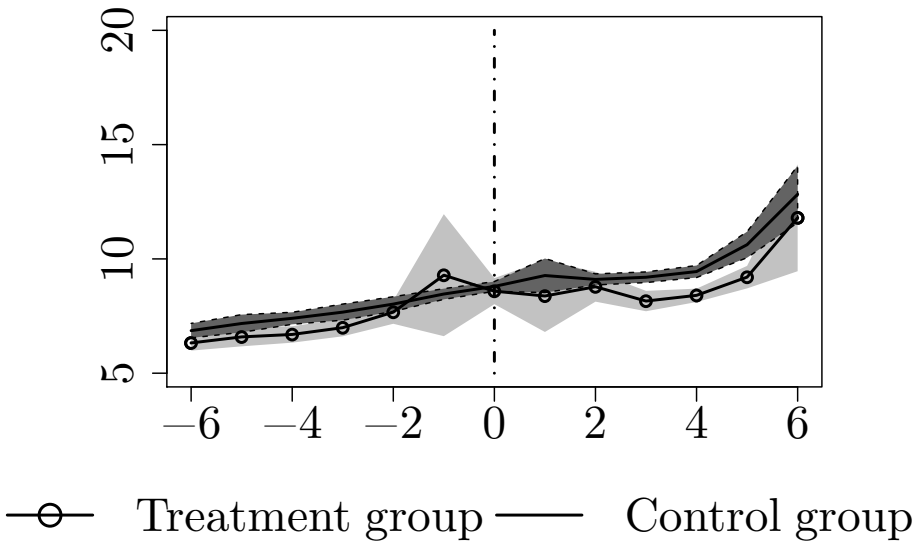


FIGURE 3.28: Hourly wages for women, more than 6 months but less than 6 years out of the labor force



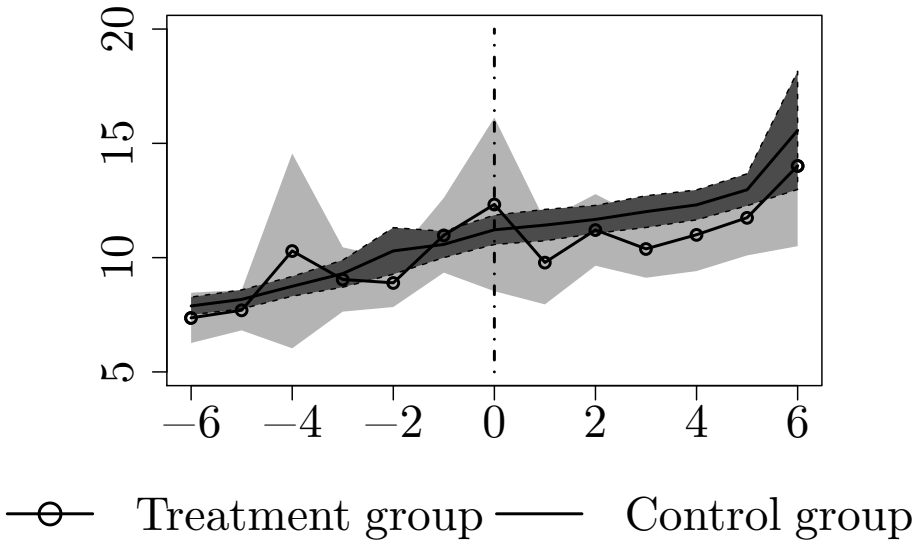
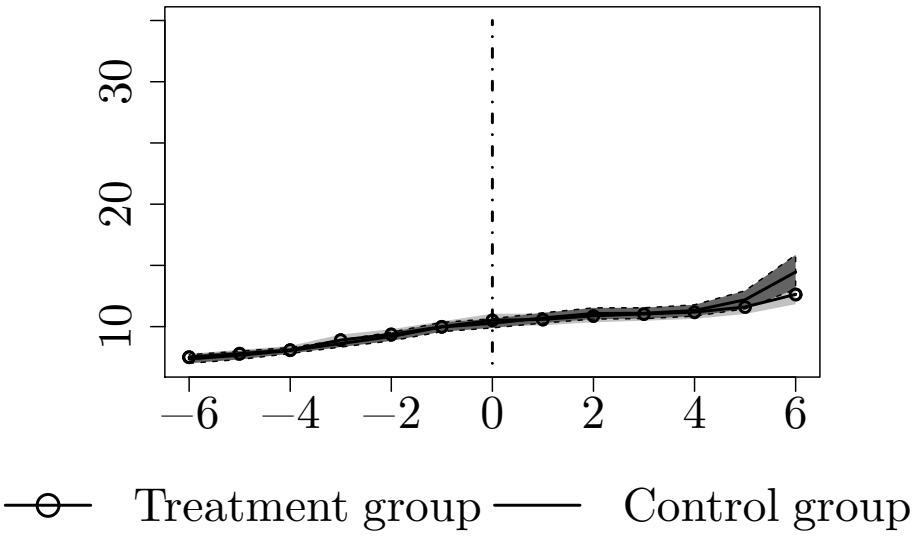


FIGURE 3.29: Hourly wages for men, less than 6 months out of labor force after birth



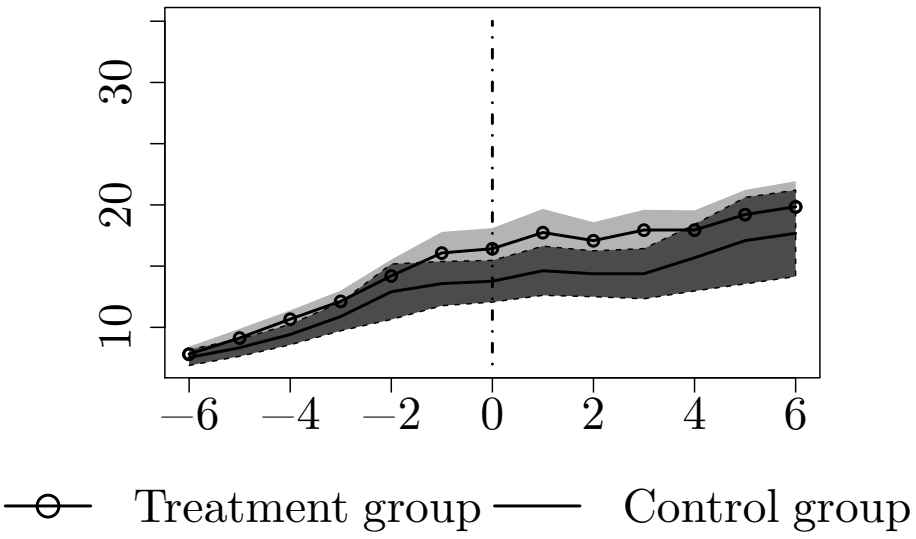
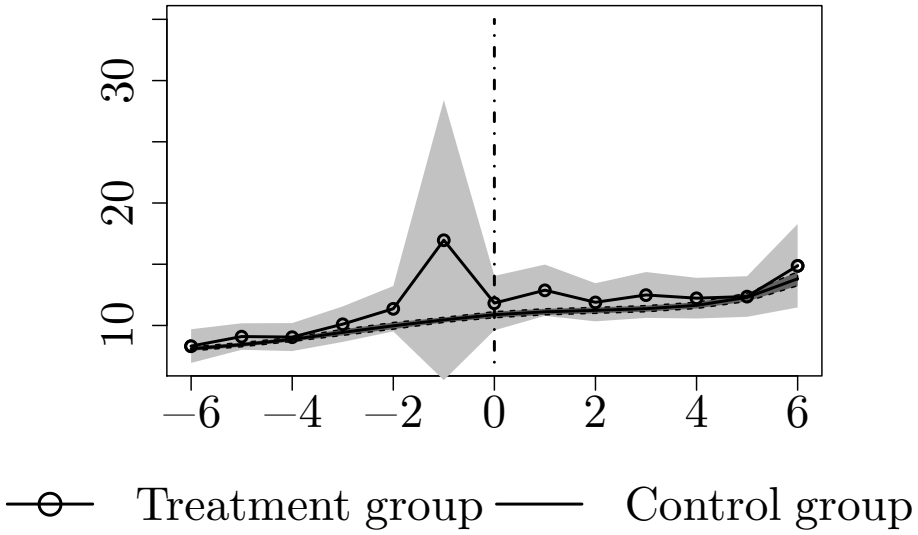


FIGURE 3.30: Hourly wages for men, more than 6 months but less than 2 and a half years out of labor force after birth



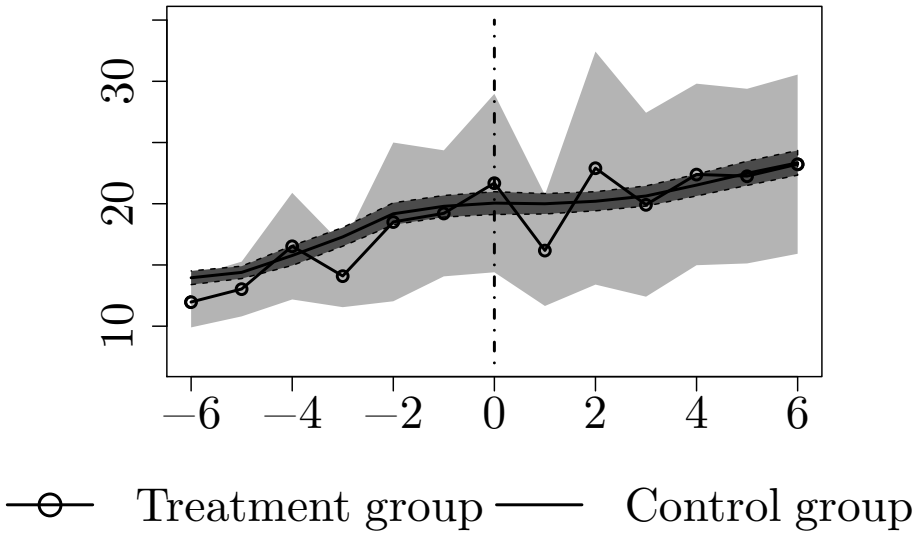
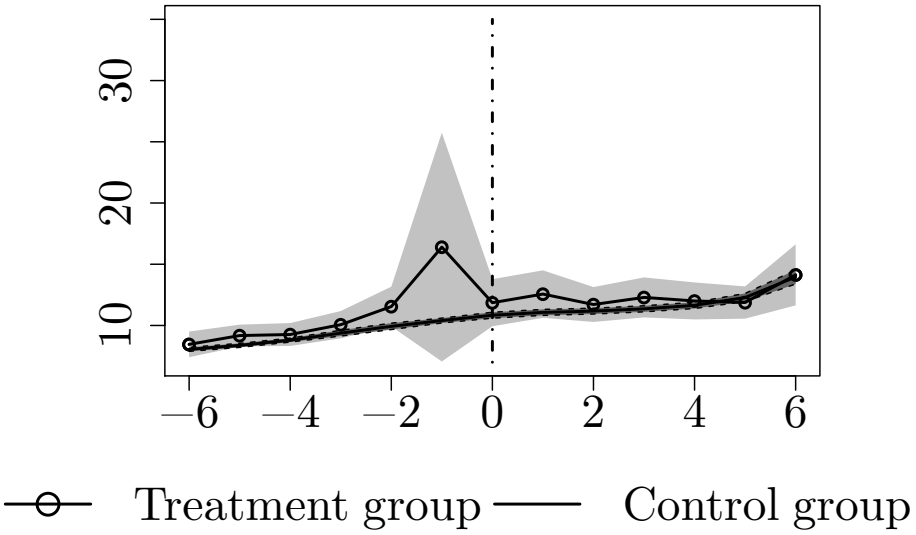


FIGURE 3.31: Hourly wages for men, more than 6 months but less than 6 years out of the labor force after birth



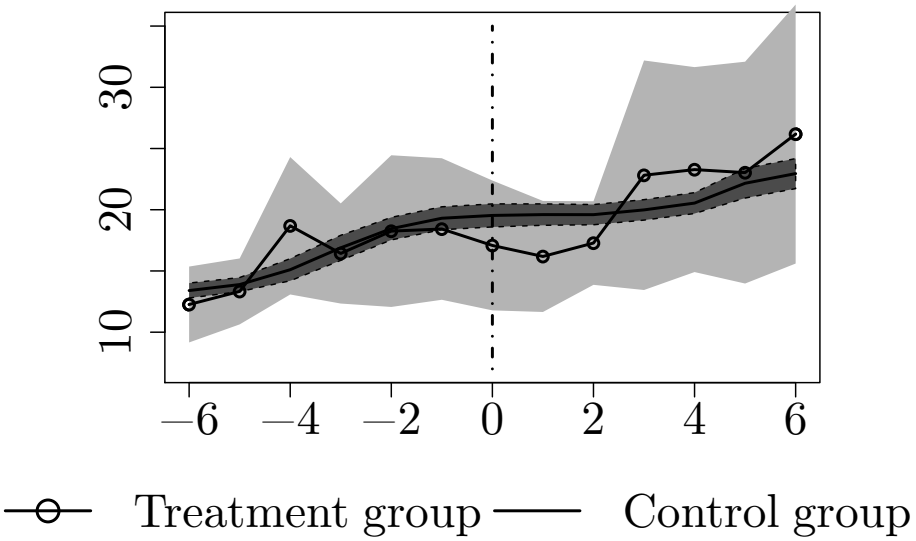
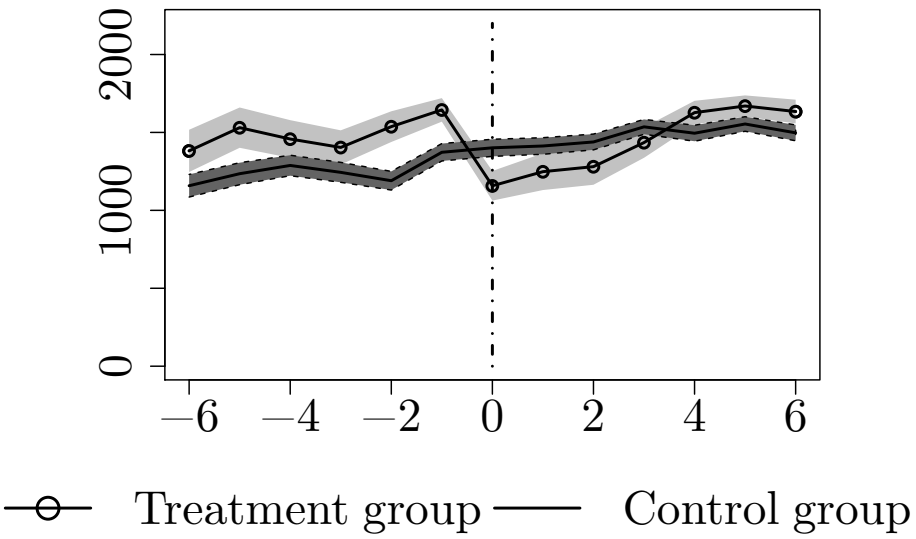


FIGURE 3.32: Worked hours for women, less than 6 months out of labor force after giving birth



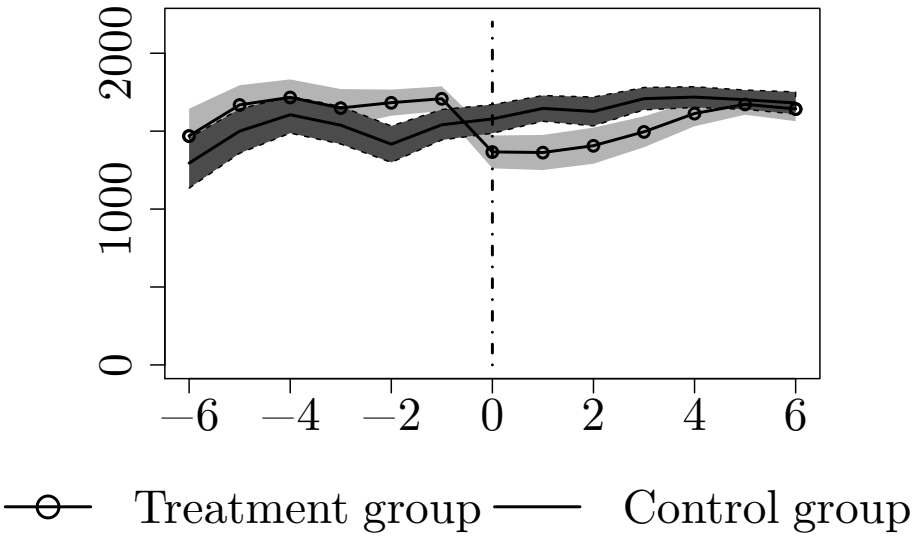
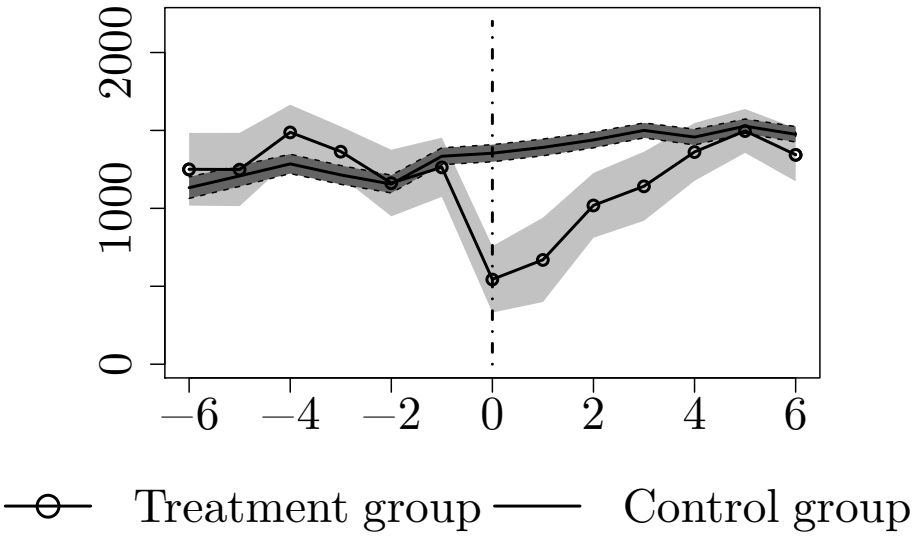


FIGURE 3.33: Worked hours for women, more than 6 months but less than 2 and a half years out of labor force after giving birth



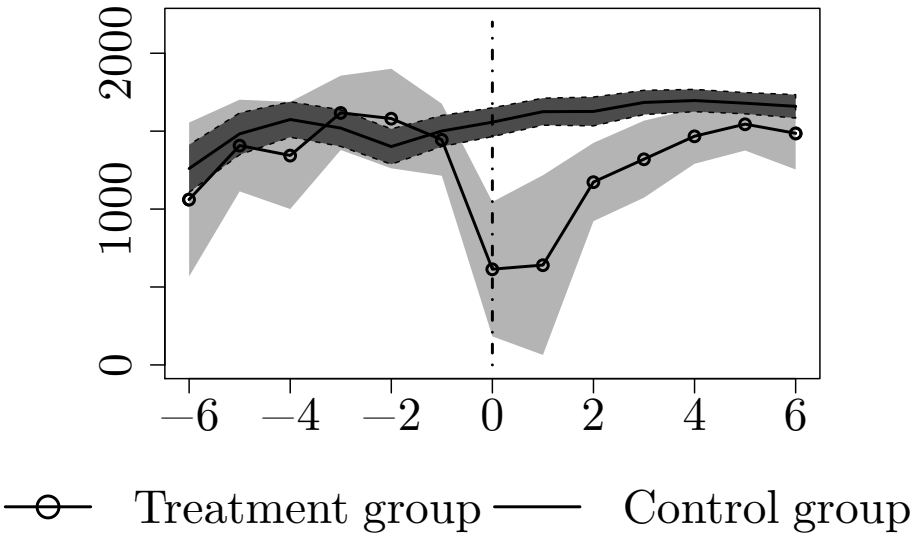
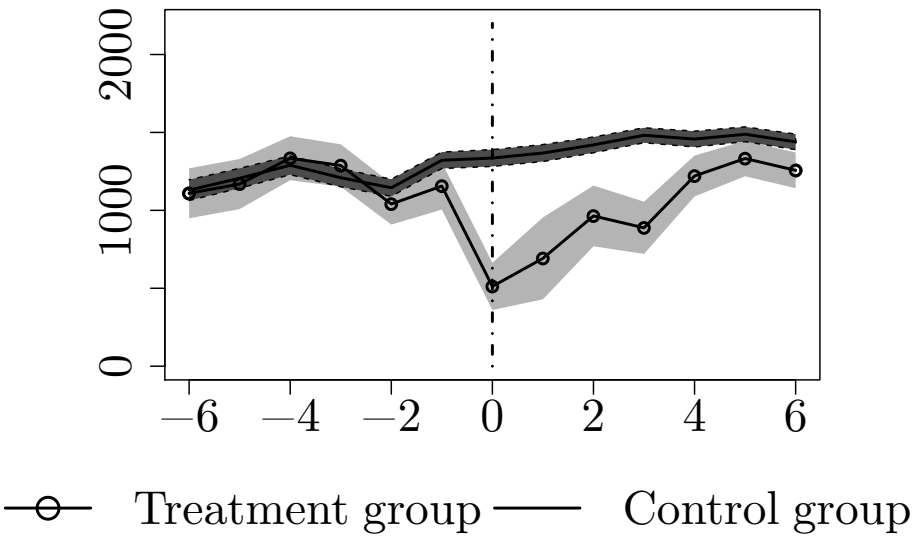


FIGURE 3.34: Worked hours for women, more than 6 months but less than 6 years out of the labor force



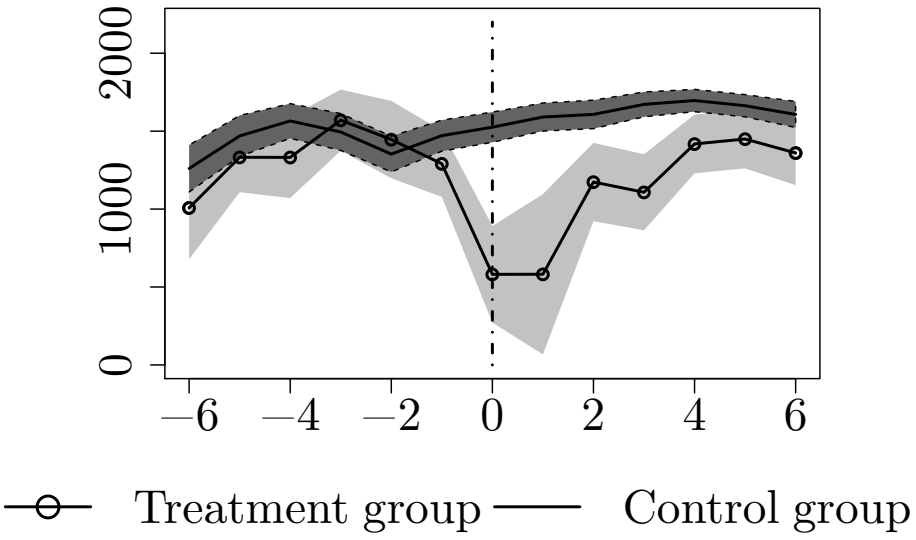
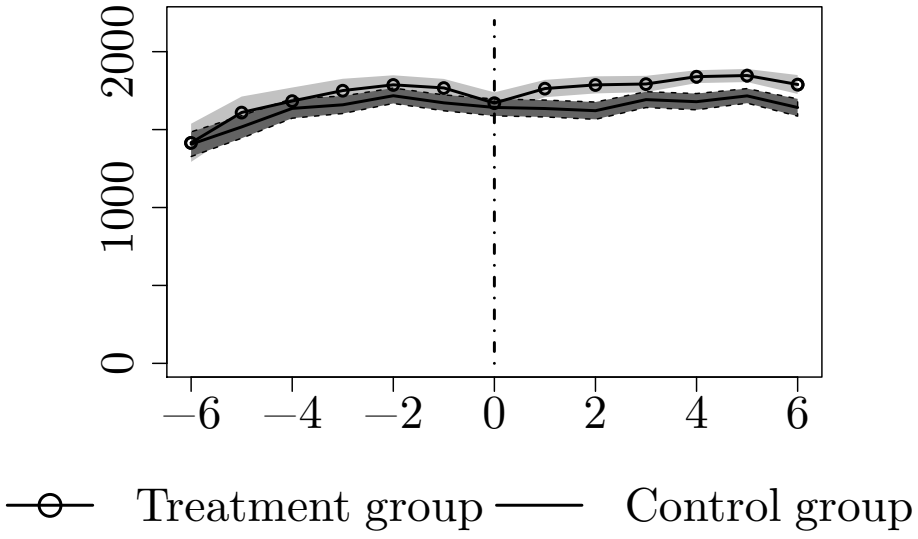


FIGURE 3.35: Worked hours for men, less than 6 months out of labor force after birth



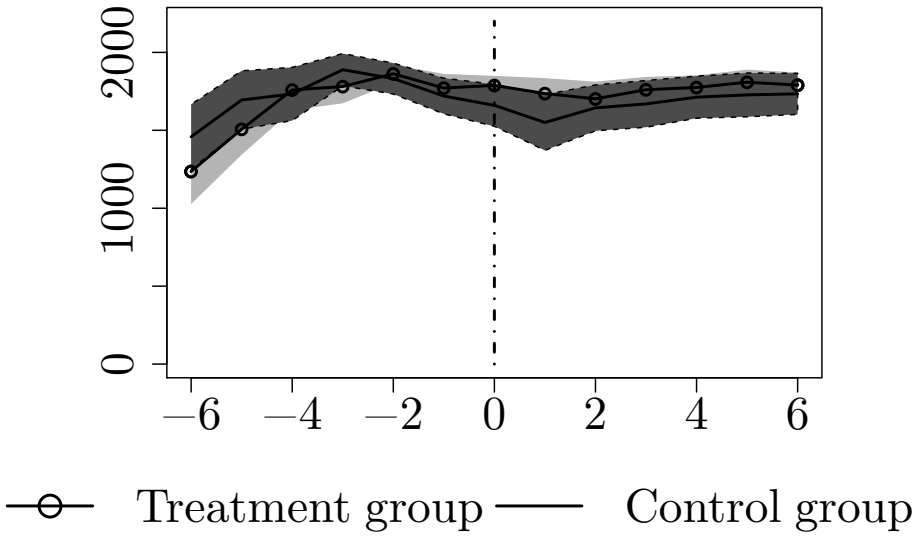
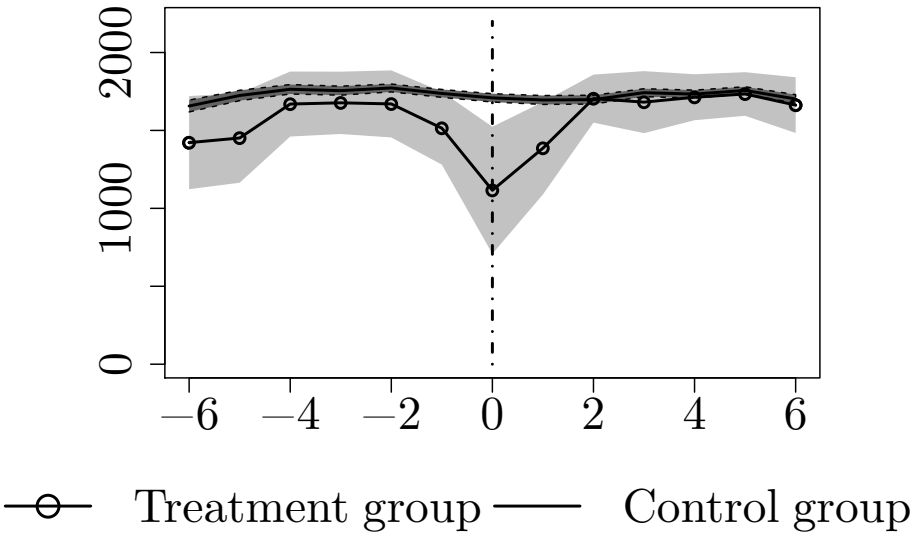


FIGURE 3.36: Worked hours for men, more than 6 months but less than 2 and a half years out of labor force after birth



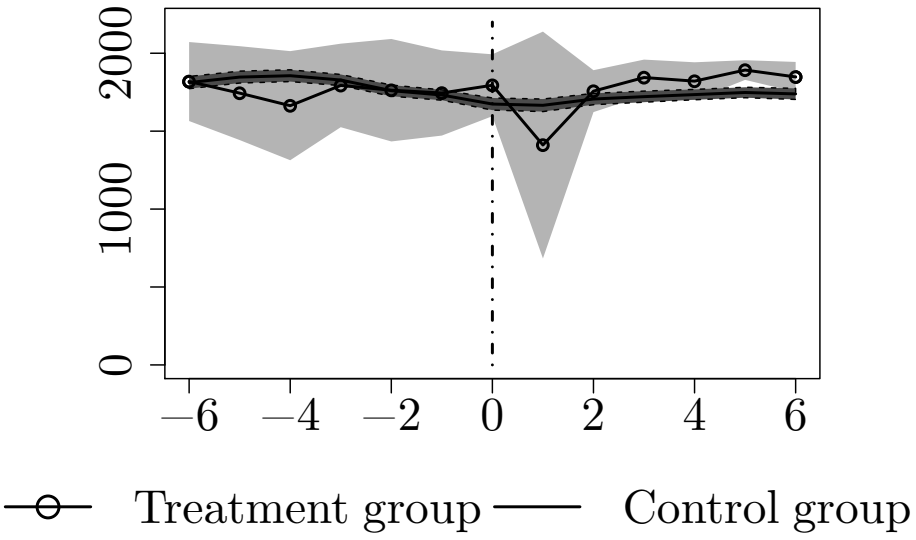
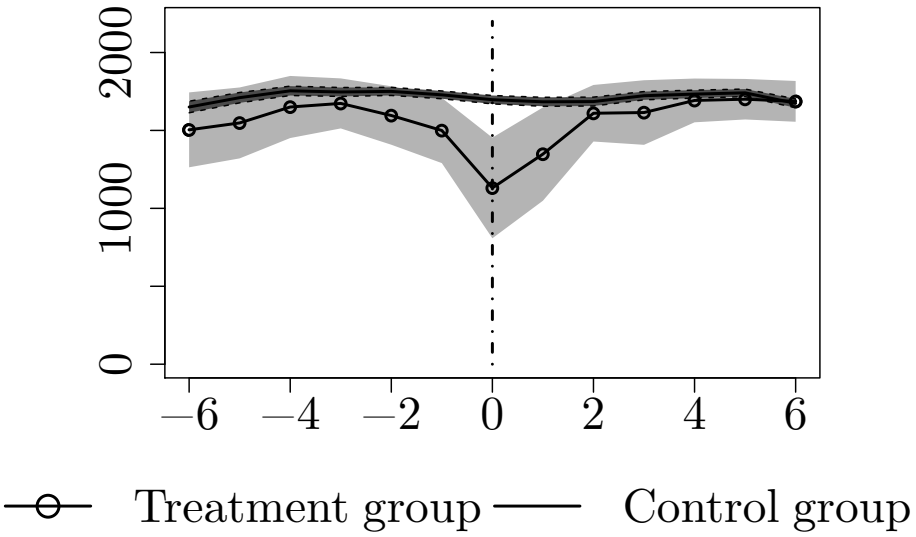
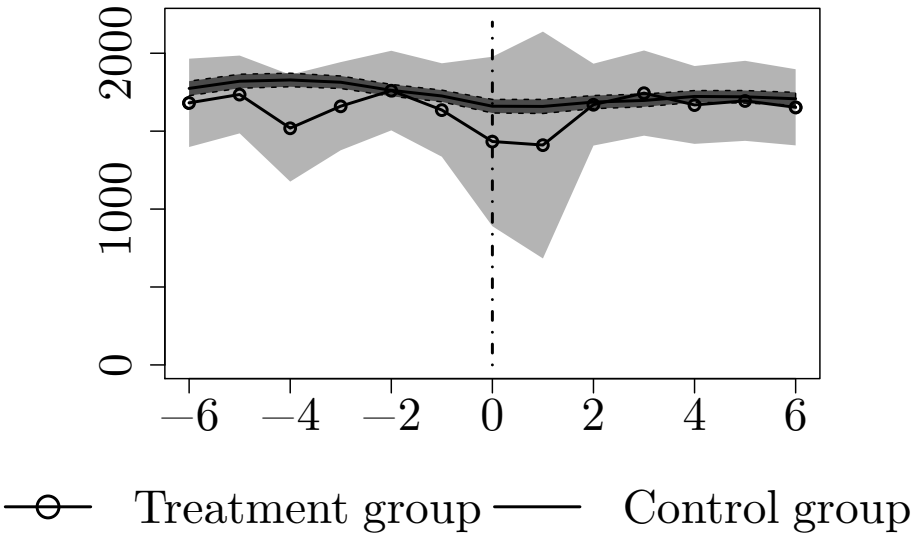


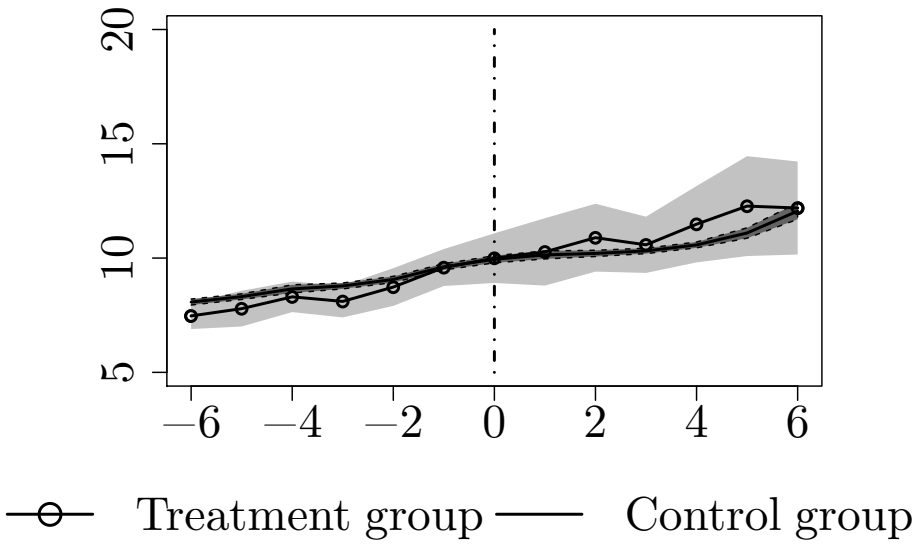
FIGURE 3.37: Worked hours for men, more than 6 months but less than 6 years out of the labor force after birth





3.7.1.3 Impact of the third child

FIGURE 3.38: Hourly wages for women, less than 6 months out of labor force after giving birth



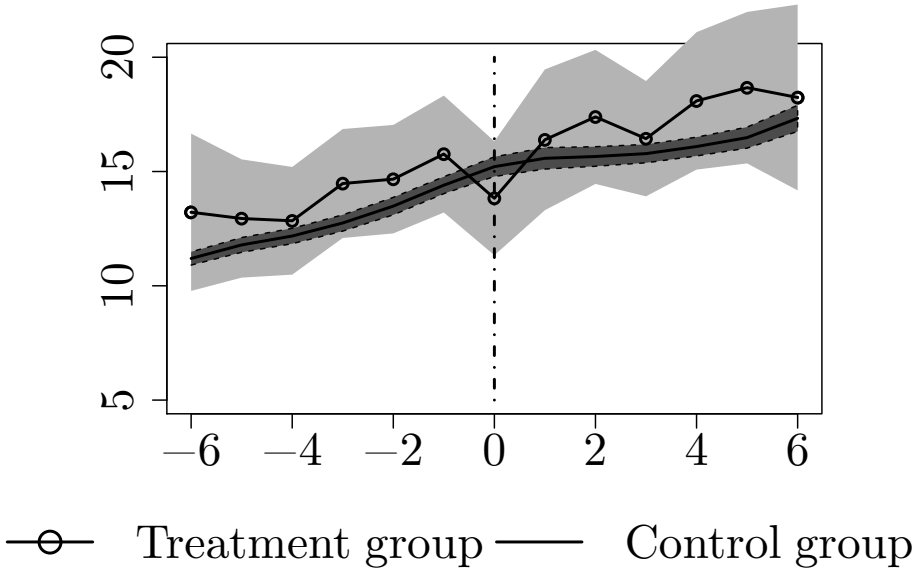
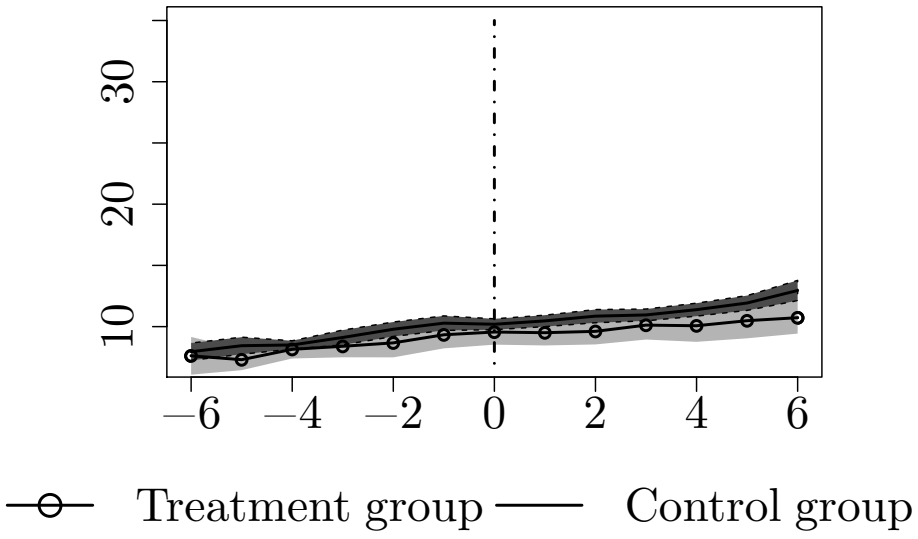


FIGURE 3.39: Hourly wages for men, less than 6 months out of labor force after birth



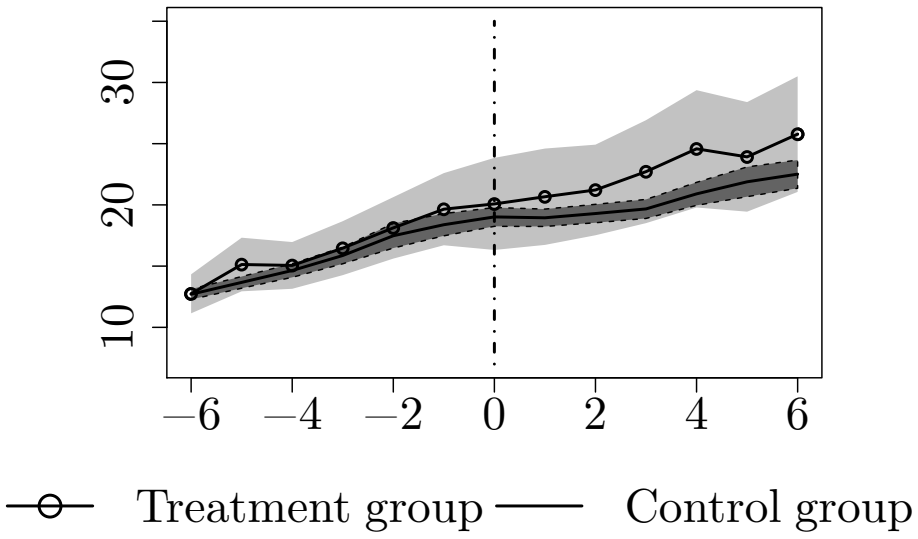


FIGURE 3.40: Hourly wages for men, more than 6 months but less than 2 and a half years out of labor force after birth (only for lower educated men)

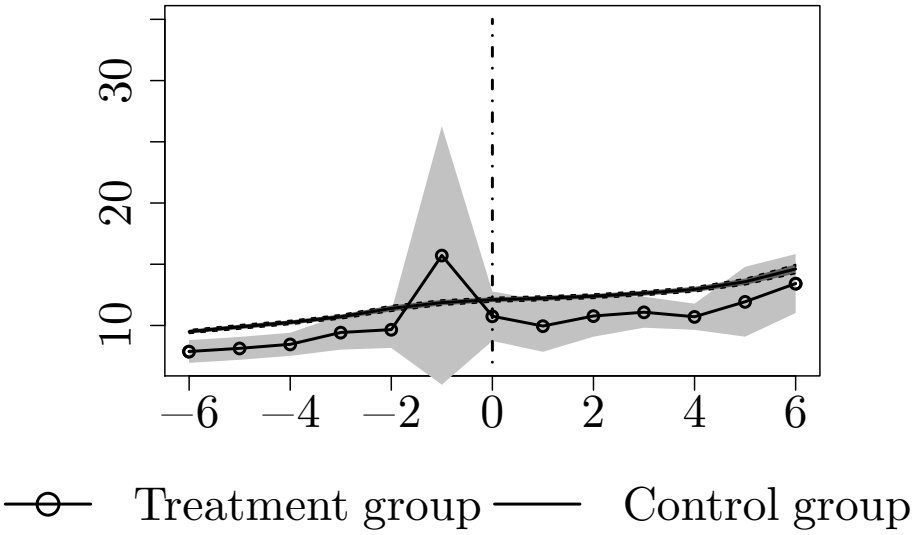
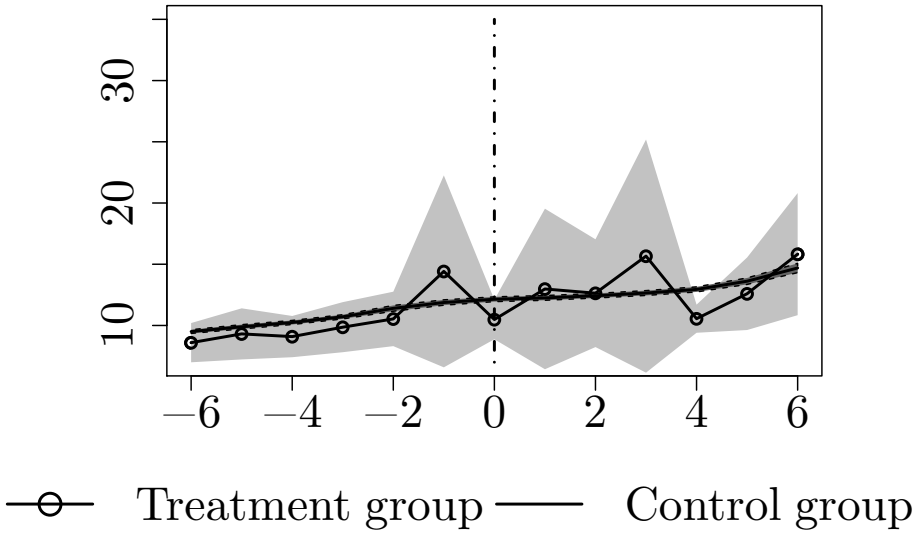


FIGURE 3.41: Hourly wages for men, more than 6 months but less than 6 years out of the labor force after birth



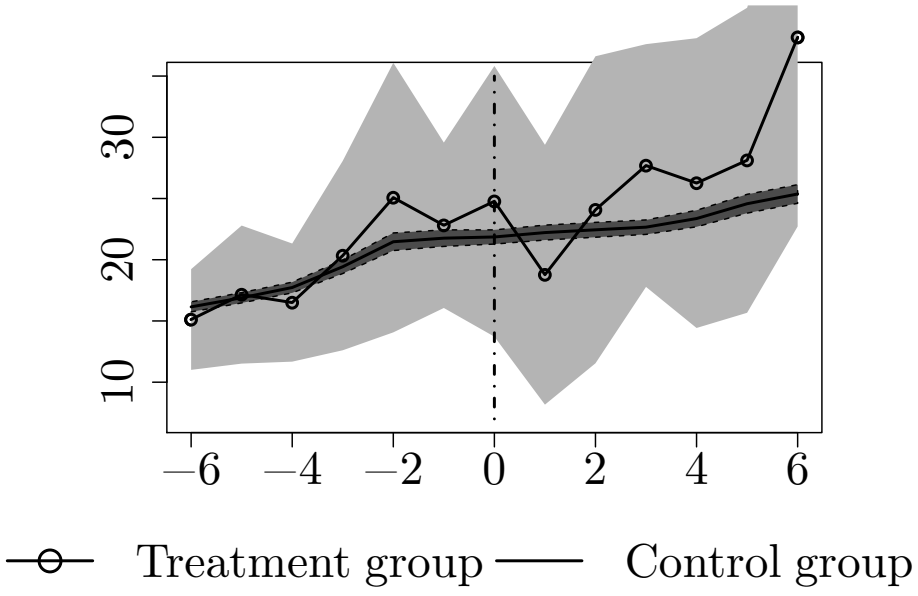
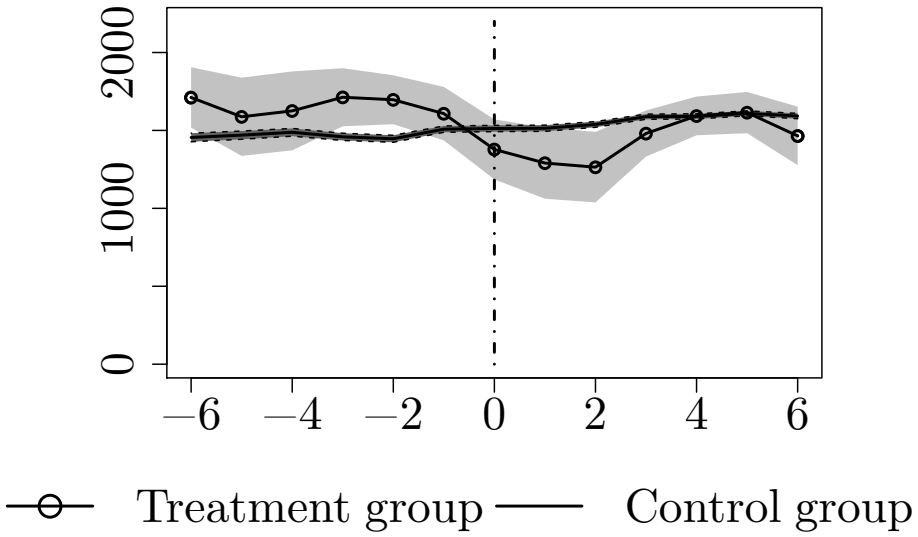


FIGURE 3.42: Worked hours for women, less than 6 months out of labor force after giving birth



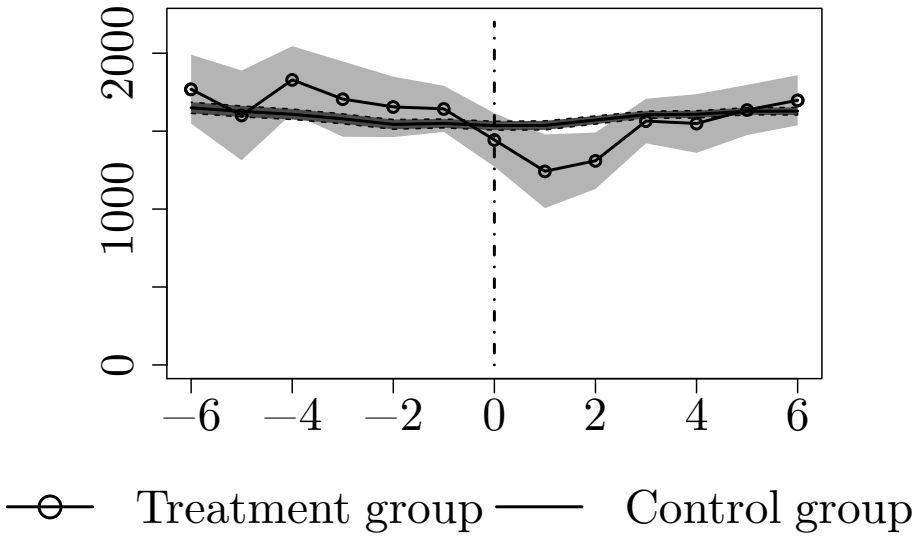
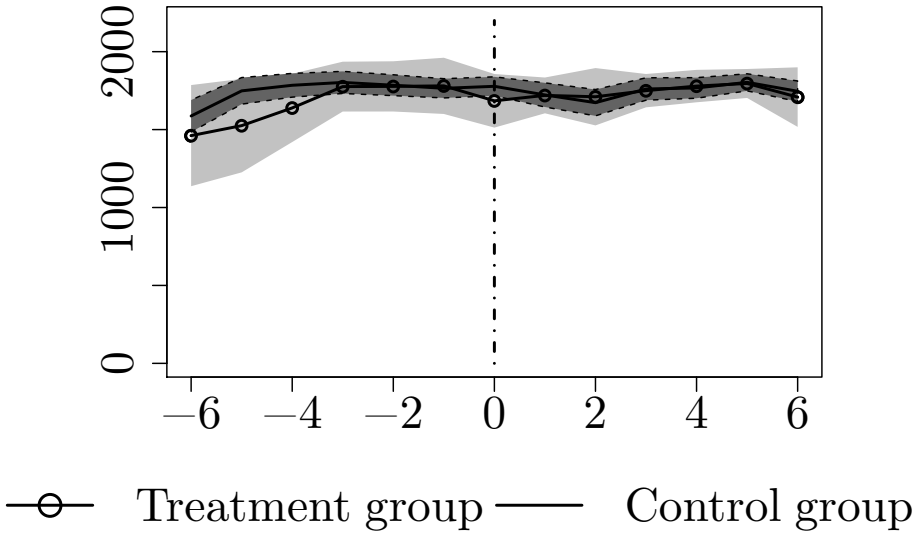


FIGURE 3.43: Worked hours for men, less than 6 months out of labor force after birth



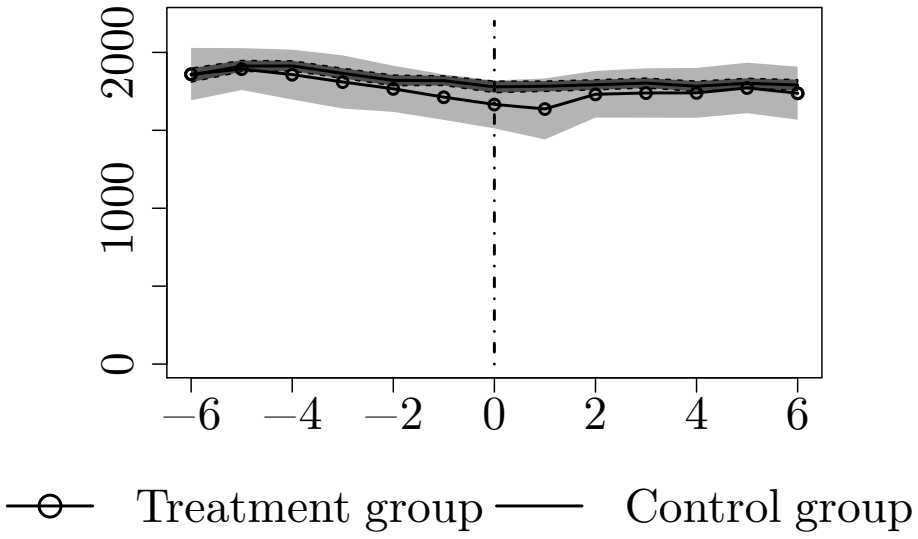


FIGURE 3.44: Worked hours for men, more than 6 months but less than 2 and a half years out of labor force after birth (only for lower educated men)

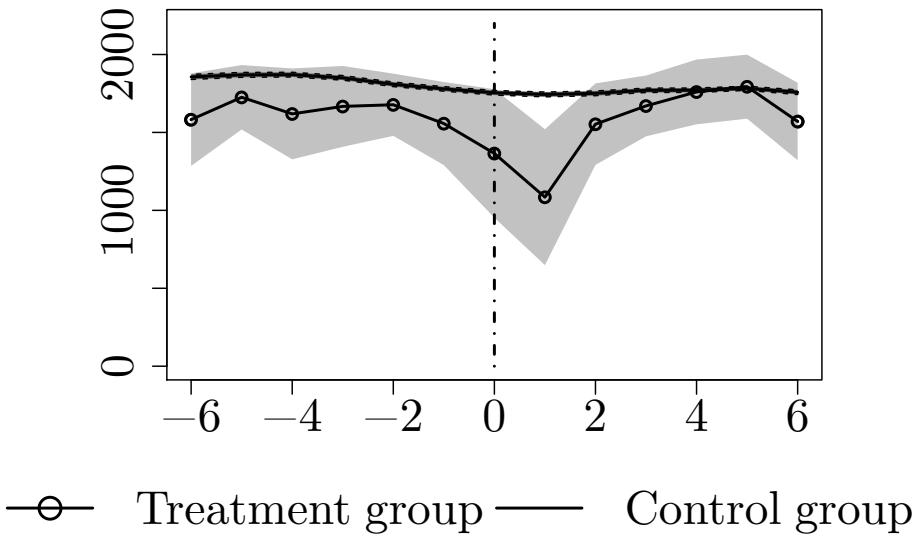
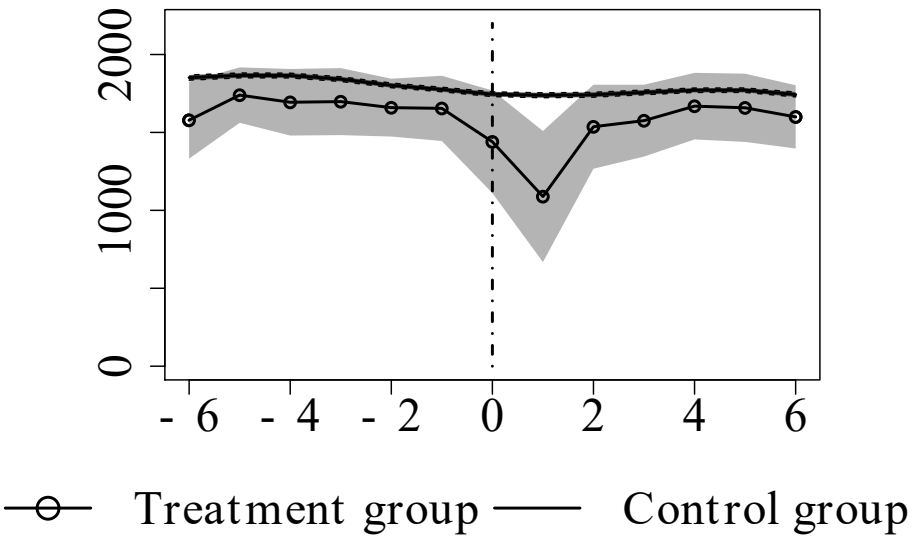
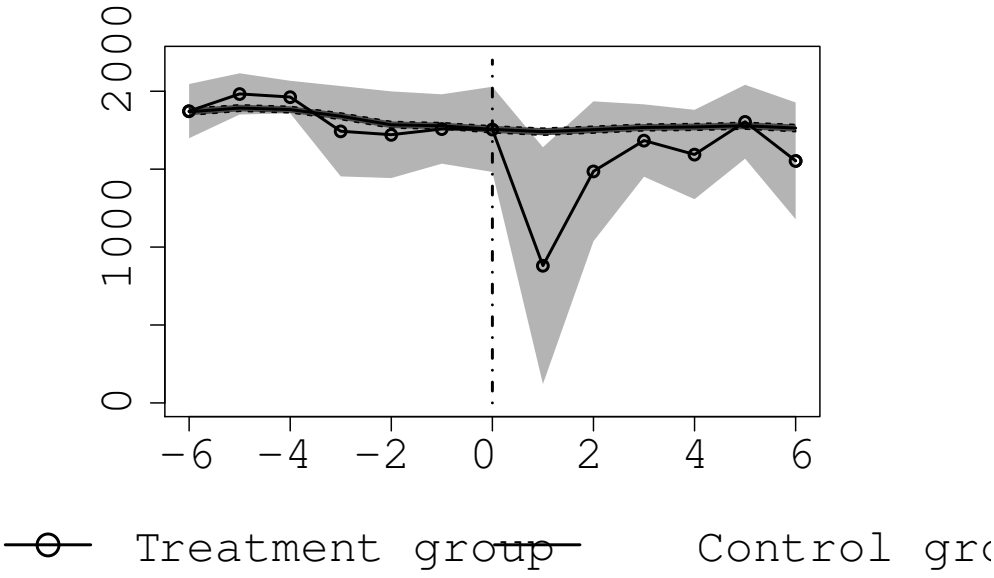


FIGURE 3.45: Worked hours for men, more than 6 months but less than 6 years out of the labor force after birth





Bibliography

Jérôme Adda, Christian Dustmann, and Katrien Stevens. The career costs of children. *Journal of Political Economy*, 125(2):293–337, 2017.

Joshua D. Angrist. Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records. *The American Economic Review*, 80(3): pp. 313–336, 1990.

Janeen Baxter, Belinda Hewitt, and Michele Haynes. Life Course Transitions and Housework: Marriage, Parenthood, and Time on Housework. *Journal of Marriage and Family*, 70(2):259–272, 2008.

Danielle Boyer and Muriel Nicolas. Les comportements d'activité des mères à la suite du CLCA à taux partiel. Technical Report 1, 2012.

David Card and Alan B. Krueger. Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. *The American Economic Review*, 84(4):772–793, 1994.

Michael R Chernick. *Bootstrap methods: A guide for practitioners and researchers*. John Wiley & Sons, second edition, 2007.

Rhys Davies and Gaelle Pierre. The family gap in pay in Europe: a cross-country study. *Labour Economics*, 12:469–486, 2005.

Emmanuel Duguet, Chloé Duvivier, Joseph Lanfranchi, and Mathieu Narcy. Identification et quantification des sources de l'écart de rémuné-ration entre hommes et femmes au sein de la Fonction publique. Technical Report 89, 2015.

Marco Francesconi. A joint dynamic model of fertility and work of married women. *Journal of Labor Economics*, 20(2):336–380, 2002.

Markus Gangl and Andrea Ziefle. Motherhood, labor force behavior, and women's careers: An empirical assessment of the wage penalty for motherhood in britain, germany, and the united states. *Demography*, 46(2):341–369, 2009.

George-Levi Gayle and Robert A Miller. Life-Cycle Fertility and Human Capital Accumulation. *Tepper School of Business. Paper*, 128, 2006.

Rebecca Glauber. Race and Gender in Families and at Work : the fatherhood wage prenium. *Gender and Society*, 22(1):8–30, 2008.

Marek Hlavac. *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Harvard University, Cambridge, USA, 2015. URL http://CRAN.R-project.org/package=stargazer. R package version 5.2.

Melissa Hodges and Michelle Budig. Who gets the daddy bonus, Organizational Hegemonic Masculinity and the Impact of Fatherhood on Earnings. *Gender and Society*, 24(6):717–745, 2010.

Olivier Joseph, Ariane Pailhé, Isabelle Recotillet, and Anne Solaz. The economic impact of taking short parental leave : Evaluation of a French reform. *Labour Economics*, 25: 63–75, 2013.

Alexandra Killewald. A reconsideration of the fatherhood prenium: marriage, coresidence, biology, and father's wages. *American Sociological Review*, 78(1):96–116, 2013.

Laurent Lequien. The impact of parental leave duration on later wages. *Annals of Economics and Statistics/ANNALES D'ÉCONOMIE ET DE STATISTIQUE*, pages 267–285, 2012.

Shelly Lundberg and Elaina Rose. Parenthood and the earnings of married men and women. *Labour Economics*, 7:689–710, 2000.

Shelly Lundberg and Elaina Rose. The effects of Sons and Daughters on Men's Labor Supply and Wages. *The Review of Economics and Statistics*, 84(2):689–710, 2002.

Dominique Meurs, Ariane Pailhé, and Sophie Ponthieux. Child-related Career Interruptions and the Gender Wage Gap in France. *Annals of Economics and Statistics*, (100):15–46, December 2010.

Ariane Pailhé and Anne Solaz. Inflexions des trajectoires professionnelles des hommes et des femmes après la naissance d'enfants. *Recherches et prévisions*, 90(1):5–16, 2007.

Jane Waldfogel. The effect of children on Women's Wages. *American Sociological Review*, 62(2):209–217, 1997.

Hadley Wickham and Romain Francois. *dplyr: A Grammar of Data Manipulation*, 2015. URL http://CRAN.R-project.org/package=dplyr. R package version 0.4.1.

Lionel Wilner. Worker-firm matching and the parenthood pay gap : Evidence from linked employer-employee data. *Journal of Population Economics*, 29(4):991–1023, 2016.

Chapter 4

The time and the transitions back to work in France after maternity[1]

4.1 Introduction

Numbers for 2015, from Eurostat, show that the employment rate of women between 20 and 49 years old without children is 76,4% and decreases to 70,9% for women with only one child (under 6 years old). The proportion of part time workers among women between 20 and 49 years old without children is 20.4% and 24.9% for women with one child (under 6 years old). However, employment rate can have different meanings. Lower educated women often do not find full time jobs, and thus have to work part time. Guedj (2013) shows that 23% of childless women, under 40 years old, without any education are forced to work part time. This figure decreases to 5% when women have, at least, two years of higher education. Still in Guedj (2013) wee see that the employment rate, for each level of education, also depends on age; for both women and men, the relationship between employment rate and age has an inverted U shape. Thus the employment status of women seems sensitive to education, number of children and age. In Rodrigues and Vergnat (2016), we have shown that women, on average, decrease their supplied yearly worked hours 2, 4 and 6 years after giving birth and that the magnitude of this reduction depends on the duration of maternity/parental leave. However it would also be interesting to know more about which characteristics of women influence their decision to participate, if at all, to the labour market again. Which types of women are more likely to come back to full time work? Is having a higher education correlated with a higher probability of coming back to full time work, and with shorter maternity leaves?

[1] This chapter is derived from an article co-authored with Vincent Vergnat.

Does tenure in a firm play a role? Are past wages an important factor when deciding between going back to the previous employer, or switch bosses? In other words, when and how are mothers coming back to the labour market after giving birth? This paper attempts to answer these questions in the case of France, by estimating a competing risks model as in Arntz et al. (2017).

The links between the duration of maternity leave and the employment of women have already been addressed in the literature. For example, in Leibowitz et al. (1992), the authors study how the availability of child care impacts the employment of mothers two years after giving birth to their first child. Leibowitz et al. (1992) estimate a probit model and take the duration of maternity leave into account by running the regression twice; once when the children are 3 months old and again when the children are 2 years old. Less educated women are less likely to return to work, but women with a high school degree do not differ much from university educated women. Family income is negatively correlated with the probability to return to work; i.e. the higher the family income, the less likely the woman will start working again. Higher predicted wages when returning to work also increase the probability of entering the labour market again. Young mothers were more likely to go back to work in the three months after giving birth to their child if these young mothers were living with their own mothers.

More recently, Geyer et al. (2015), shows that introducing universal child care increases female labour supply by an average of 7 percentage points. Asai (2015) investigates the effect of an increase in cash benefits which is not accompanied by an increase in maternity leave duration for Japan. The author finds no evidence that an increase in cash benefits increase the likelihood of mothers returning to work. The author stresses however that this result might be caused by the lack of child care facilities in Japan and by the specificity of the Japanese labour market. In Japan, a very high commitment to the workplace, with very long and inflexible working hours, leads mothers to prefer (or be socially pressured) to stay at home to take care of their children. Lalive and Zweimüller (2009) study the impact of two maternity leave reforms in Austria on future fertility decisions and on the career of mothers. The first reform, in 1990, was an extension from the job protection period, which, before the reform ended with the first birthday of the first child, and after the reform with the first birthday of the second child. The second reform, in 1996, was a six months reduction of the job protection period. The first reform increased fertility and return to work was delayed even after the job protection period had ended. The second reform shortened the spacing between the first and the second birth, but without major impacts on total fertility. In a subsequent article, Lalive et al.

(2014) study the impact of a third reform which was implemented in the year 2000. This reform increased the maximum duration of cash benefits the parents were entitled to for having a first child. This reform increased the time mothers spent at home before returning to work. For France, Pailhé and Solaz (2007) study the impact of a birth on the probability, for mothers, to invest less in the labor market. They show that the higher the birth order, the greater the likelihood of reducing working hours. Age at birth and degree level also play an important role in the transitions made by mothers. A similar study from Domingo and Marc (2012), analyzes the professional trajectories of mothers in France and in particular the effects of a break or a reduction of work. They show the importance of the demographic characteristics and the conditions of the last job on the occupational situation of the mothers after a birth.

All the above articles study similar questions to ours, but use different methods. The authors of these articles consider probit models or (multinomial) logit to estimate the probability of the mothers going back to work. In this paper, we estimate a competing risks model, because we want to study not only how long it takes for mothers to go back to work, but also how. Competing risks models have been used extensively in economics to study the duration of unemployment. Edin (1989), Narendranathan and Stewart (1993), McCall (1996), Mussida (2007), Portugal and Addison (2008) are some examples. Güell and Petrongolo (2007) discuss the timing of conversion from temporary work contracts to permanent work contracts or other states (such as unemployment) for Spain.

The papers from which we draw the most inspiration are Fitzenberger et al. (2016) and Arntz et al. (2017). Fitzenberger et al. (2016) use a dataset for a large German company to study the return to job of women after giving birth to a first child. Their data, covering the years from 2000 to 2008, show that return-to-job after a first birth is a source of high uncertainty for firms. Indeed, an important part of first mothers do not return to work after the parental leave. This is less the case for women who are more involved in their careers. This paper offers a very interesting and detailed analysis, however the conclusions are drawn from female workers of a single firm. Results from Fitzenberger et al. (2016) must thus be taken with some reservations. Arntz et al. (2017) use German data from 1985-2005, and focus on the link between labour market conditions, legislation, the length of maternity duration and the return to work. The authors show that there have been important changes in the behaviour of women during the last three decades. Women are more likely to give birth to a second child or to work again for the same employer, but in part time, than getting back to work full

time or dropping completely out of the labour market. It seems that the maternity leave legislation plays a role in this pattern. The authors show that the longer the job protection period, the longer the maternity leaves taken by the mothers. This fact leads to important costs for the employer and for the economy as a whole. These findings are not necessarily generalizable to France. Fertility and labour supply decisions of French women are quite different from those of their German counterparts. For example, total fertility rate is 1.47 in Germany while it is 2.01 for France in 2014. The percentage of part time working women aged 25 to 54 is 47.6% for Germany, against 28.7% for France in 2015 (Eurostat).

Using a rich administrative dataset, we study the young mothers' decisions; how long mothers decide to stay out of the labour force, and how do mothers come back. Different mothers, with different education levels, careers, and working in different environments will make different choices as to how long their break will be, and how they will come back to the labour market. We use, as Arntz et al. (2017), a competing risks model to study this question for France. We will, compare our results to those of Arntz et al. (2017). The paper is constructed as follows: in Section 4.2 we present the institutional setting of France to give more context to the interested reader, Section 4.3 describes the data, Section 4.4 the methodology employed, Section 4.5 presents and comments the results. We conclude in Section 4.6.

4.2 Institutional setting

In France, since 1980, maternity leave is decomposed into two components: the prenatal leave (6 weeks for the first child) and the postnatal leave (10 weeks for the first child). During this leave, mothers recieve a compensation if they have contributed to a social security scheme some time before the leave and take a minimum of 8 weeks of maternity leave. The compensation is equal to the average income of the last three months before maternity leave with a daily ceiling of 83.58€ (in 2016). During maternity leave, a woman cannot be fired from her job. After the maternity leave, the employer has to give the mother her previous job or a similar job with at least the same wage. The maternity leave is considered as a period of effective service, thus the mother enjoys the same wage increase as employees of the firm with the same occupation. After the maternity leave, a mother (or father) can reduce her working hours to take care of her child until the child's third birthday.[2]

[2] This is the so-called *congé parental d'écucation* in French.

Since 1990, the following parental leave provisions are available to parents of one child:

- Before 2004, only parents of at least two children were eligible for the parental leave provision.
- Since 2004, a new benefit was created, called *Complément de Libre Choix d'Activité* (CLCA). It allows one parent of a child under three, who has paid contributions for pension during at least 8 quarters in the last 2 years, to decrease their professional activity during six months. The amount of the benefits, per month, on January 2017, is 390.92€ if the parent stops working, 252.71€ for half-time work or less, and 145.78€ if the parent works for 50% to 80% of full time work. Before April 2014, depending on the household income, these amounts could have been increased by about 185€.
- Since 2015, yet another benefit was introduced. The *Prestation partagée d'écucation de l'enfant* (PreParE) which, like the German *Elterngeld*, introduces an incentive for fathers to take a parental leave. The amounts and conditions, for parents of a first child, are the same as for CLCA but now, a second parent can also take the benefits for six months.

Once the parents want to get back to work, many opportunities for childcare exist in France. In France there is a fairly large supply of collective structures for young children and parents can also employ childminders. According to the DREES [3], there are on average in each department of metropolitan France, 68 places in preschool childcare facilities per 100 children in 2013. The French family allowance funds support some of the expenses related to child care expenses (for parents who are working or searching for a job). Two major cases occur:

1. If the child goes to institutions hosting young children, such as nurseries:
   - The price of childcare is calculated based on the household's resources (this calculation is possible thanks to subsidies granted by the French family allowance funds to these institutions).
   - Parents get a tax credit for their expenses for childcare.
2. If the child is cared for by a childminder or by a nanny at the parent's home:

[3] Research Division of the French Ministry of Social Affairs and Health

- In the case of the employment of a childminder, parents are exempt of all social security contributions, and in the case of employment of a nanny, of 50% of all social security contributions.
- Benefits to help parents finance the childminder or the nanny (*Allocation Garde d'Enfant à Domicile* (AGED) and *Aide à l'emploi d'une assitante maternelle agréée* (AFEAMA) before 2004 and *Complément de libre choix du Mode de Garde*(CMG) afterwards).
- Tax credit for childcare expenses (after taking into account the potential benefits for childcare).

These different policies allow parents, especially mothers, to reconcile family and work life. Mothers are able to choose either to come back to their previous employer, or change employers, work full time or part time, or decide to have a long career break and take care of their child. These possibilities are of course also available to fathers, but for reasons that are not discussed in this paper, it is mostly mothers that are confronted to this choice. Our study focuses on the time a mother is not effectively working after giving birth, which encompasses the legal post-maternity leave and the potential parental leave.

4.3 Data and summary statistics

4.3.1 The DADS-EDP panel

The data used in this paper resulted from the merger of the DADS panel and the EDP dataset and is called "DADS-EDP".[4] The "DADS-EDP" panel covers the period from 1976 to 2010. These data are provided by the INSEE and give information on a sample of the French population.[5] The sample is composed of persons born between the 1st and the 4th of October (only those born during an even year for the years 1967 to 2002 are included in the DADS-EDP panel). People born abroad, who never worked, who are self-employed or who work as civil servants of national public services are excluded from the data. Between 1967 and 2010, the sample evolved. Since 1991 and 1992, civil servants working in public institutions of an industrial and commercial nature are included in the panel as well as publicly-employed hospital staff (since 1984) and civil

[4] *Déclaration Annuelle des Données Sociales* or Annual Declaration of Social Data, *Échantillon Démographique Permanent* or Permanent Demographic Sample

[5] INSEE: National Institute of Statistics and Economic Studies. The French national statistical institute.

servants of territorial communities (since 1988). Unemployment benefits recipients were also included in the panel in 2002 and agricultural workers in 2003. Residents of overseas territories were included in 2004. The merged "DADS-EDP" data provides information on socio-demographic variables such as date of birth, date of wedding, place of residence, level of education, number and date of birth of children... Data on wages, hours worked, type of employment contract, the starting and closing dates of the period of paid work are available for each individual and for each year of the panel. We also know the size, the employment sector and the location of the firm the person is working in. This merged dataset, provided as is, unfortunately does not contain every variable from either the DADS nor the EDP. For example, the EDP contains much more information on children for instance, such as whether the birth occurred at home or in a hospital, which is not available in the merged DADS-EDP.

Our variable of interest, duration of post-natal leave, is not available as such in the data, so we had to compute it. The raw data is in a spell format (one line for each job), so it is possible to compute the variable quite easily, for most cases. Here are the steps to create this variable:

1. We observe the start date and the end date of paid work in the data as well as the date of the birth of the child. We can therefore deduce the time that elapsed between the date of the birth of the child and the date of return to work.

2. In the data, we observe that some women never exit their jobs to enter maternity leave. This can be women still considered in the company's payroll software but already on maternity leave, for example. For these women, we "spot" the exit date by the reduction in hours worked between the year before the birth and the period around the birth. To control the results, or in case of missing information, we divide the wage earned in the period around the birth by the hourly wage the year before to deduce the reduction of the working time during this period. Thus, we get the reduction of working time around the birth that we convert into days and we subtracted 42 days (the legal duration of the pre-natal leave) as we are interested in the duration of post-natal leave.

3. Finally, there are some women for whom there are no exits from the labor market at the date of birth of the child (as case 2) but for which an exit takes place before the end of the statutory post-natal leave (before 70 days). For these women we computed the maternity leave duration using the method described in point 2 (for the period surrounding the birth) to which we added the diffence of days between

the date of exit before the end of the statutory maternity leave and the date of return to work.

4.3.2 Control variables

In this paper we are interested in the following transitions. A mother may return to:

- the same employer, full time.
- the same employer, part time.
- another employer, full time.
- another employer, part time.[6]

There are mothers that never go back to work, but we cannot know the reason from the data. These observations are thus considered as censored. Some of these mothers may very well never return to work, ever, which in our context would constitute a defective risk. However, of these censored mothers, 88.3% had their child in 2010, the last year of observation, which suggests that most mothers end up going back to work again, and that censoring is thus orthogonal to mothers' characteristics, and thus, no bias in our estimations should be present. Also, one should keep in mind that our sample is constituted of mothers that were quite well integrated into the labour market, and that they have the right to return to their previous job if they so wish. We are thus fairly confident that there is no defective risk present.

To explain women's decisions we considered a rich set of control variables; women's age, tenure and experience until the birth of their child, as well as their pre-birth wage. We divided the annual wage the year before birth into three classes (less than 20 000 € , between 20 000 and 30 000 € and more than 30 000 €). We added a dummy equal to 1 if the mother has already worked part-time in the past. Occupation is also included, again as a categorical variables divided into four levels: executives, intermediate occupations, clerks and blue-collar workers. The educational variable has a lot of missing values, so instead we chose to use the occupation categories as a proxy for education. We also added another categorical variable for the activity sector the mother worked in before giving birth, divided into four levels: industry, construction, trade and services.

[6] We define part-time work using the variable *CE* of DADS-EDP : 1: Full time and 2: Part time.

Another variable that might influence women's decisions is the size of the firms where they worked before giving birth. We considered firms with less than 11 employees, with less than 50 but more than 10 and firms with more than 49. We defined these three classes to take French legislation into account. Indeed, firms with more than 10 employees have to have staff representatives, and firms with more than 50 employees have to set up an employee representative committee. We further added a binary variable to take the policy change of parental leave of 2004 into account. This dummy equals 1 if the birth took place in 2004 or later, and 0 otherwise. We also added a dummy for eligibility to the CLCA (8 quarters of social contributions in the last two years). A marriage dummy is introduced in order to control for the family situation (we have access only to the information of the marriage, unmarried women living with a partner are not indicated). Finally, we created a variable measuring the distance between the city of work and the city of residence for each women, the year before birth. To do this, we used the geographical coordinates provided by INSEE for each municipality. This will allow us to test the hypothesis that women who have a long commute will be more inclined to reduce their labour supply or to change to an employer that might be closer to their place of residence.[7]

Macroeconomic variables included are the national GDP growth rate[8] and the unemployment rate at the French department level[9] which control for the general state of the economy and the labour market. Unlike Arntz et al. (2017), we do not have information on the availability of childcare facilities at the local level. This might not be an issue in the specific case of France, since, as explained in the introduction, childcare facilities are very common in France, and a large number of benefits exist to help parents of any social class cover the costs.

4.3.3 Descriptive statistics

Our sample is restricted to women in metropolitan France who gave a first birth between 1995 and 2010 and who worked full time for at least 12 months at the same firm during the year before giving birth. The second condition makes it possible to have a relatively homogeneous pre-birth sample of women in terms of attachment to the labour market

[7] The distance we computed is the great circle distance, and it might actually be a bad proxy or commuting time. Indeed, it is entirely possible to live close to one's working place and take very long to get there because of lack of public transportation or too much traffic. However, with this data, this is the best proxy we could come up with.

[8] Source: World Bank.

[9] Source: INSEE

and only covers mothers who can access parental leave. We removed public servants because accounting for hours and wages does not allow for an adequate deduction of exit time from the labour market (see section 3.1). We suppressed observations where data the year before birth seemed to be wrong (hourly wages less than 6€ and annual hours less than 1600, as we focus only on women working full time the year before birth), which amounts to a loss of 1203 observations. We removed women for which data needed for the study were missing such as the activity sector, occupational categories or distance from home to work (128 observations). We further removed women who gave birth to a child during the 42 first days of a year (770 observations) because they may have started their maternity leave in the civil year preceding the birth. They may therefore have a reduction in earnings and hours worked in the civil year preceding the birth, which may impact the approximation of post-natal leave. Finally, we suppressed women who had a post-natal leave duration of 0 day (1060 women). We think that this situation can occur in companies that have implemented the full maintenance of the salary for mothers on maternity leave, preventing us from correctly approximating the post-natal leave because we do not observe a reduction in working time or salary the year of birth.[10] The final sample is thus composed of 8467 women.

TABLE 4.1: Age at which women have their first child (n=7534)

	Min	1st quartile	Median	Mean	3rd quartile	Max
Age	20.0	27.0	29.0	30.0	32.0	47.0

Table 4.1 presents the age of mothers at the birth of their first child. The average age of childbearing is 29.7. In France, Volant (2017) shows that the average age varied over the period from 1989 to 2010 between 25.9 and 28 years and nearly 42% of the children born (in rank 1, 2 or 3) in 2010 have a mother aged between 27 to 32. In our sample this number is 50%. This difference can be explained by the presence in our sample of women who are relatively well integrated into the labour market. We have women who have been in full-time for at least 1 year in a same firm. These characteristics may not be relevant to younger mothers, which move our age distribution to the right.

[10] We did not find any particular characteristics associated with the wrong approximation of post-natal leave (post-natal leave of 0 day). This seems to be independent of the activity sector of the firm or occupational categories.

FIGURE 4.1: Distribution of the duration of maternity leave

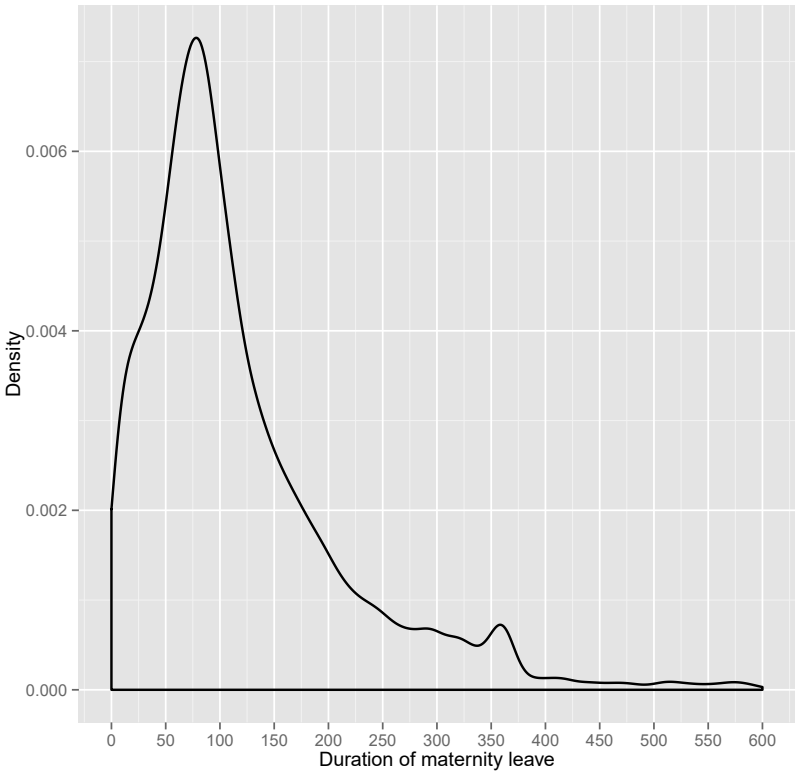


Figure 4.1 shows the distribution of the duration of maternity leave in the sample, i.e. the number of days between the birth of the child on the return to work. Our definition of maternity leave included therefore also possible parental leave. The distribution's mode appears to be around 75 days, which is slightly longer than the statutory duration of maternity leave in France. There is also a small bump around the first birthday of the child. After the child's first birth however, the number of women still in maternity leave decreases quickly.

TABLE 4.2: Transitions characteristics

Transition	Duration of post-mat. leave (days)	Age	Tenure (days)	Experience (days)	Live in the capital region (%)	Child's birth $\geq$ 2004 (%)
Censored	529.4 (985	30.8	1291.4	2531.4	23.5	94.2

this region. The last column of table 4.2 shows that 89.9% of the censored mothers had their child in or after 2004 against 50 to 60% for the other transitions.[12]

TABLE 4.3: Mother's characteristics before each transition

Women who transitioned to...	Censored	Full time at the same employer	Full time at another employer	Part time at the same employer	Part time at another employer
(%)	10.8	57.6	13.4	13.3	4.9
Education					
No education (%)	5.4	53.7	15.5	16.9	8.4
High school education (%)	9.6	57.6	12.8	13.9	6.1
2-3 years of higher education (%)	11.2	58.7	14.2	12.3	3.6
4-8 years of higher education (%)	14.1	56.5	12.9	12.9	3.7
Occupation					
Executive (%)	13.6	56.7	13.3	12.6	3.8
Intermediate occupation (%)	9.2	59.6	14.2	13.0	4.0
Clerk (%)	11.6	56.2	13.0	13.5	5.6
Blue-collar worker (%)	8.4	58.7	12.5	14.4	6.0
Establishment size					
<20 employees (%)	12.4	53.7	15.3	11.8	6.8
≥ 20 and <500 (%)	9.3	60.1	12.8	13.9	3.8
≥ 500 (%)	11.8	59.8	9.5	16.1	2.9

Table 4.3 shows some characteristics of the different transitions on which our study will focus (only for completed spells). The top row shows the proportion of mothers who transitioned to the different states. For instance, 57.8% of the women in the sample came back to full time work to their previous employer, while around 16% changed employers. Almost 17% returned to part time work. In all, around 70% of young mothers returned to their previous employer. What is striking is that post-natal leave duration, tenure and experience have very large standard deviations, which suggest very wide distributions for these variables, as shown in figure 4.2. However, it appears that the duration of post-natal leave is lower when women decide to return to the same employer. It is also these women who have more important professional experiences and tenure. These women seem, thus, to be better rooted in their firm. Age of mothers is, for all risks, around 29-30. Mothers who come back to full time work at the same employer are those that have the highest average wage the year before the birth. 60.1% of the mothers who come back to full time work to the same employer had their child in or after 2004 against 44.7% for

[12] 2004 is a year were the CLCA was introduced, as explain in section 4.2.

those who return to part-time work at another employer. Since this reform, women seem to come back more often at the same employer.[13] For each transitions, less than 10% of mothers are blue collar worker. The proportion of clerks is more important among mothers who come back to part-time work at another employer. Very few mothers work in the construction sector, but an important part of mothers work in services sector. The proportion of mothers working in services is less important among mothers who came back to part-time work at another employer. On the contrary, the proportion of mothers in the trade sector is higher. The share of mothers working in large firms is higher among women who came back at the same employer. Larger companies offer certainly more opportunities for mothers to take up their jobs on a part-time basis or to have more flexible hours of work. This might be much harder for smaller firms, which have less human resources at their disposal. The existence of employee representative committees or at least representatives of the staff may also foster more family-friendly situations in larger companies.

4.4 Econometric analysis

The exposition here follows Kleinbaum and Klein (2005), as well as Fine and Gray (1999) and Arntz et al. (2017). We refer the interested reader to these texts for more details.

A woman enters maternity leave for one reason only, but could then exit maternity leave in different ways. Young mothers are exposed to these risks during post-natal leave, so we are interested in the duration of the post-natal leave after the birth of their first child and on which variables influence the different risks. Because of data limitations, we do not know if a woman that never came back to work has moved out of the country, had a leave that ended after 2010 (the last year of observation) or died. It is also not possible to know if these mothers decided to have a second child during their first post-natal leave. These observations are thus considered as censored. From table 4.3, only 10.8% of observations are censored.

Typically, survival data is analyzed by first plotting the survival function. This is usually done with the non-parametric Kaplan Meier (KM) estimator. Then, researchers continue their analysis by estimating a Cox proportional hazards model, which links the hazard function to covariates. In the context of competing risks, one would define the Cox PH model as follows:

[13] 2004 is a year were the CLCA was introduced, as explain in section 4.2.

λ

also used the `cmprsk` package (Gray, 2014), as Arntz et al. (2017) did, to check our results. To allow for individual heterogeneity we included an individual random effect in equation (4.3):

$$\lambda$$

FIGURE 4.3: Kaplan-meier estimate of the lifetime distribution function for each risk

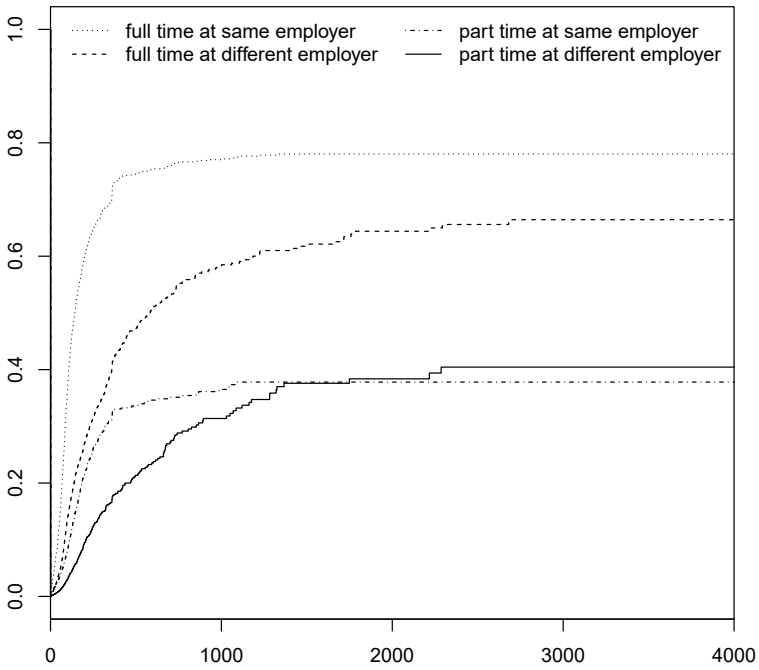
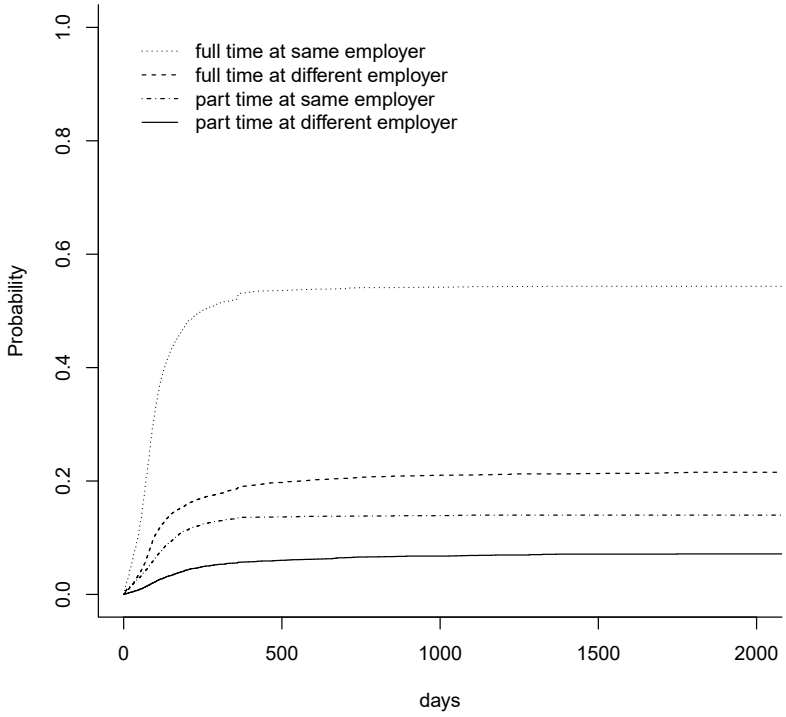


Figure 4.4 represents cumulative incidence functions. In other words, it represents the probability to fail in one risk on or before time t while taking into account the competing risks. The competings risks affect the probability of occurence of an event. If we compare Figure 4.4 with Figures 4.3, we observe that the Kaplan-meier estimates overestimated the probability of occurence of each risk, which explains why it is important to take the presence of the other risks into account and why Fine and Gray (1999) use cumulative incidence functions in their model. We observe in Figure 4.4 that the cumulative probability of occurence of the risk "full time at same employer" is much more important than the three others risks. Each risk occurs relatively soon after birth except the risk "part time with a different employer" for which the cumulative incidence function grows much more slowly.

FIGURE 4.4: Cumulative incidence functions



4.5 Results and discussion

Table 4.4 shows the estimation results of the competing risks regression (equation (4.3)) and table 4.5 shows the estimation of the competing risk model by clustering the observations by activity sector. We do not present the estimation results by clustering by French department and neither with unobserved heterogeneity because the results are very similar to those presented in table 4.4.[15] Tenure and experience have been converted to years in order to have a similar scale between the variables. The reported coefficients are exponentiated, which allows us to present the results as risk ratios. This means that if a coefficient is greater than 1, the variable increases the risk of occurrence of the event studied and will have the opposite effect if the coefficient is less than 1. For simplification purposes, we will refer to the subdistribution hazard by hazard. Let us first analyze the personal characteristics of the mother.

One year of added age decreases the hazard of both full time risks by around 2 to 4%, while increasing the hazard by 6% to return to the same employer part time. Younger

[15] With unobserved heterogeneity, only the random coefficient for the risk "full time at the same employer" is significant and is equal to 0.397. The tables are available upon request.

TABLE 4.4: Fine and Gray competing risks regression results

	Full time same employer	Full time another employer	Part time same employer	Part time another employer
Personal characteristics				
Age	$0,985^{***}$	$0,963^{***}$	$1,060^{***}$	$1,002$
Married	$1,070^{**}$	$1,000$	$0,969$	$0,904$
Ile de France	$0,890^{***}$	$1,302^{***}$	$0,934$	$1,117$
Professional characteristics				
Tenure	$1,025^{***}$	$0,911^{***}$	$1,021$	$0,924^{**}$
Experience	$0,996$	$0,997$	$0,986$	$1,002$
Part time in the past	$0,878^{***}$	$1,035$	$1,021$	$0,943$
Distance work/home at t-1	$0,999^{***}$	$1,001^{**}$	$1,000$	$1,001^{*}$
Annual wage at t-1				
Annual wage $< 20K$	ref	ref	ref	ref
Annual wage $\in [20K, 30K[$	$1,084^{**}$	$1,057$	$0,944$	$0,762^{*}$
Annual wage $\geq 30K$	$1,367^{***}$	$1,073$	$0,687^{**}$	$0,371^{***}$
Occupation				
Executives	ref	ref	ref	ref
Intermediate occupation	$1,089$	$1,105$	$0,898$	$0,532^{***}$
Clerk	$0,989$	$0,993$	$0,982$	$0,694$
Blue collar worker	$0,825^{**}$	$1,195$	$1,027$	$0,635$
Firm size at t-1				
$\leq$10	ref	ref	ref	ref
$\geq$11 et <50	$1,124^{**}$	$0,919$	$1,012$	$0,762^{*}$
$\geq$50	$1,183^{***}$	$0,772^{***}$	$1,236^{**}$	$0,645^{***}$
Activity sector at t-1				
Industry	ref	ref	ref	ref
Construction	$0,909$	$1,057$	$1,222$	$1,179$
Trade	$0,824^{***}$	$1,197$	$1,145$	$1,192$
Service	$0,872^{***}$	$1,134$	$1,156$	$0,997$
Parental leave legislation				
Post-reform CLCA	$0,567$	$1,581$	$0,362$	$1,776$
Eligibility at CLCA	$1,437$	$0,447^{*}$	$2,437$	$0,378^{*}$
Economic environnement				
GDP growth	$0,939^{***}$	$1,059^{***}$	$0,977$	$1,191^{***}$
Local unemployment rate	$0,969^{***}$	$0,993$	$1,010$	$0,974$

Signifiance level : *** p<1%, ** p<5%, * p<10% (H0 : exp(coefficient)=1)

TABLE 4.5: Fine and Gray competing risks regression results with cluster by activity sector

	Full time same employer	Full time another employer	Part time same employer	Part time another employer
Personal characteristics				
Age	$0,985^{***}$	$0,963^{***}$	$1,060$	$1,002$
Married	$1,070^{***}$	$1,000$	$0,969$	$0,904$
Ile de France	$0,890^{***}$	$1,302$	$0,934$	$1,117$
Professional characteristics				
Tenure	$1,025^{***}$	$0,911^{***}$	$1,021^{***}$	$0,924^{*}$
Experience	$0,996^{*}$	$0,997$	$0,986^{***}$	$1,002$
Part-time in the past	$0,878$	$1,035$	$1,021$	$0,943^{*}$
Distance work/home at t-1	$0,999^{***}$	$1,001^{**}$	$1,000$	$1,001^{*}$
Annual wage at t-1				
Annual wage $< 20K$	ref	ref	ref	ref
Annual wage $\in [20K, 30K[$	$1,084^{***}$	$1,057$	$0,944$	$0,764^{***}$
Annual wage $\geq 30K$	$1,367$	$1,073$	$0,687^{***}$	$0,371^{***}$
Occupation				
Executives	ref	ref	ref	ref
Intermediate occupation	$1,088^{**}$	$1,105$	$0,898$	$0,532^{***}$
Clerk	$0,989$	$0,993$	$0,982$	$0,694^{**}$
Blue collar worker	$0,826^{***}$	$1,195$	$1,027$	$0,635^{**}$
Firm size at t-1				
$\leq$10	ref	ref	ref	ref
$\geq$11 et <50	$1,123^{***}$	$0,919^{**}$	$1,012$	$0,762^{***}$
$\geq$50	$1,183$	$0,772^{***}$	$1,236^{**}$	$0,645^{***}$
Activity sector t-1				
Industry	ref	ref	ref	ref
Construction	$0,910^{***}$	$1,057$	$1,222^{***}$	$1,179^{***}$
Trade	$0,825$	$1,197^{***}$	$1,144^{***}$	$1,192^{***}$
Service	$0,872$	$1,134^{***}$	$1,155^{***}$	$0,997$
Parental leave legislation				
Post-reform CLCA	$0,568^{***}$	$1,581^{***}$	$0,362$	$1,776^{**}$
Eligibility to CLCA	$1,436^{***}$	$0,447^{***}$	$2,436$	$0,378^{***}$
Economic environnement				
GDP growth	$0,937$	$1,059^{***}$	$0,977^{*}$	$1,191^{***}$
Local unemployment rate	$0,969^{***}$	$0,993$	$1,010$	$0,974^{***}$

Signifiance level : *** p<1%, ** p<5%, * p<10% (H0 : exp(coefficient)=1)

mothers tend to return more often to full time work, maybe because they want to limit the losses in their wages due to human capital depreciation, as well as signaling attachment to their previous company. Being married at the time of birth increases the risk of returning full-time to the same employer, perhaps reflecting a better division of labour between men and women in married couples that allows women to reconcile their work in full time and family life. This may also reflect the fact that the specialization of married couples takes place before the birth of a child. Thus, married women who have a child, have already decided at the time of marriage that they would invest in the labour market. On the contrary, married women who have decided to specialize more in domestic life may already do so before the birth of the child. These reasons may explain that being married increases the return to full-time work with the same employer. Living in the Île de France, the capital region, decreases the hazard of returning to the same employer by 11%, while increasing the hazard of changing employer by 30%. This clearly shows that mothers living in the capital region have much more job opportunities than in the rest of the country. However, once we cluster by activity sector, the hazard of changing employers is not significant anymore.

As for professional characteristics, tenure increases the hazard to return to the same employer by around 3%, but decreases the hazard of changing employers by almost 10%. Mothers that have been working in the same firm for a long time also tend to return to their previous employers, most probably as a way for these mothers to signal their attachment to their previous employer (or to avoid having to prove their worth again in a new firm), which is a result that is also found in Arntz et al. (2017) or Fitzenberger et al. (2016). They also have accumulated more firm specific human capital, and switching firms would cause that human capital to be, at least in part, wasted. Maybe surprisingly, overall job market experience does not play a role. This may be because we included age and tenure as further controls, which correlate highly with job market experience and may absorb the effect of tenure. Unsurprisingly, for mothers that worked part-time in the past, the hazard of returning full-time to the same employer decreases, perhaps reflecting a lower attachment of these women to the labour market. However, this variable has no effect on other transitions and its effect disappears when clustering by sector. The distance from home to work, while significant, has a very small effect. The probabilities of transition are very close to 0, and this might come from the fact that distance as such does not matter, but rather, the time one takes to go to work. It would be very interesting to analyze the impact of this variable on job transitions after giving birth, but another data source would be needed for that. One possibility would be to

use a routing system such as OSRM[16], but this is not possible via the secured system we have to use to access the data.

The pre-birth annual wage is the variable that has the largest effect. Compared to women in the reference category, the hazard of returning full time to the same employer for women earning between 20K and 30K€ annually is 8% higher and around 25% lower to go work part-time for another employer. For women earning more than 30K€ annually, the hazard to return full-time to the same employer is 37% higher, and the hazard to go work part-time for the same employer or for another employer 31% and 63% lower respectively. Women who earned higher wages also tend to return to their previous employer in full time work. There is potentially a gift exchange game going on here; employers pay high wages to their female employees and thus these young mothers reciprocate by returning to work for them. The opportunity cost is also high for women with high wages (often, educated women) to reduce their labour supply. Of course one would need to investigate this closely, by asking young mothers what is the primary reason they decided to go back to their previous employers, which is outside the scope of this current study. These effects seem robust, as clustering only changes the significativity of returning to the previous employer for women earning a high wage. The effect of wages on the speed of return to work for mothers had already been put forward by Leibowitz et al. (1992) or Arntz et al. (2017).

We used occupations and wages as proxies for education; as explained in section 4.3 the education variable was not of good quality and contained a lot of missing values. Being in an intermediate occupation decreases the hazard of switching to part-time work to another employer by 46% as compared to the reference category of executives, and also decreases the hazard of returning working full time for the same employer for blue collar workers by 17%. When clustering, the estimated coefficients for the risk of changing employer and for part-time work for all occupations becomes significant.

Establishment size is another important predictor. As in Arntz et al. (2017), larger firms tend to have their employees return, either in full time or part time. The hazard of returning full time to the previous employer is 18% (which turns out to be non-significant when clustering) and 24% part-time. Switching employers for full-time work decreases the hazard by 22% and for part-time work by 35%. A similar effect, albeit not so important in size can be seen for middle-sized firms. It could be easier for larger companies to offer part-time jobs. Indeed, in smaller firms, even if the law forces the employer, during the parental leave, to accept a request for part-time work, it is possible

[16] http://project-osrm.org/

that pressure or organizational reasons prevent the mother from actually choosing her desired working hours. The presence of work councils, defending the rights of the staff, can also facilitate part-time returns. Domingo and Marc (2012) also showed that in France, the size of the firms matters for the return to employment of mothers.

The activity sector seems to play an important role. When looking at the estimation results clustered by activity sector, we see that the hazard ratio increases significantly for most risks, but not for returning full time at the same employer. Working in the trade or services decreases the hazard of returning to full time work to the same employer, by 18% and 13% respectively for the model without clustering. After controlling for clusters by sector of activity, for these 2 sectors, we note that the risk of returning to the same employer on a part-time basis or to work full time at another employer increases. In these sectors, the presence of industry-specific policies or agreements can influence the labour supply of mothers. More in-depth studies of different branch policies should be carried out to understand the effect of the industry on returning to work for mothers.

Eligibility for the CLCA becomes significant when clustering by activity sector; for mothers that are eligible to the CLCA, the hazard to return full-time to the same employer increases by 44%. The reverse result is found for full-time at another employer. Eligibility for the CLCA seemed to have a real positive effect on *employer-retention* for young mothers. This law can have given employers an incentive to employers to change attitudes about parental leave and offer more flexibility to mothers.

Macroeconomic variables, such as GDP growth and the unemployment rate (at the departmental level) were also included. GDP growth decreases the hazard of returning working full time at the same employer by 6% (but this effect disappears once we cluster by sectors), while increasing the hazard of changing employers (full-time) by 6%. It also increases the hazard of full-time work at another employer by 20%. The local unemployment rate only decreases the hazard of returning working full-time at the same employer by 3%, but also decreases the hazard of changing employers for part-time work. Unlike Arntz et al. (2017), we found that high unemployment rates decrease the hazard of returning full time to the previous employer, but we arrive at the same conclusion as Arntz et al. (2017) concerning GDP; higher GDP levels are associated with returns to part time work rather than full time work. Less uncertainty about the state of the economy can incite women to look for another job which can allow them to reconcile more easily family life and professional life. Conversely, poor economic conditions can lead to uncertainties in terms of sustainability of jobs and encourage mothers to return to their job quickly to prove their commitment to the company.

One of the limitations of our data is that it does not contain information on the activity or the income of the partner, if present. These elements can likely have an impact on the way women return to work, however, according to INSEE data, around 90% of one year old children live in a household with a couple (Chardon and Daguet, 2009). Also, for men aged 30-49, the employment rate is around 87% and only 4% of them work part time (Guedj, 2013). This means that women in our sample are most likely in a couple, with a partner that works full time, and assuming assortative mating, a phenomenon that exists for France, according to Frémeaux and Lefranc (2017), of similar education level and background as the mothers. These facts help mitigate the lack of information on the partner.

4.6 Conclusion

Women who end their post-natal leave (including parental leave) can make different transitions back to a job: returning to the same employer or changing employer and working on a part time or on a full time basis. Results from a competing risks model show that it is not the same women who perform each of these transitions. With this model, we show how individual, professional and legal variables affect the probability of occurrence of each transition. For instance, our model shows that women who worked in large firms tend to return to the same firms, but working part-time. This suggests that these larger firms have less problems to offer mothers the opportunity to work part-time. We have tested the effect of various other variables, however, these are variables upon which the policy maker does not have any influence. From our results, we see that the CLCA increased the probability for women to return to their previous employer. From the perspective of both the mother and the employer, this is beneficial, as it greatly reduces job-hunting as well as recruitment costs. The macroeconomic variables, which proxy for the state of the economy at large, also play a role into the decision process of young mothers. If we compare our result to the results of Arntz et al. (2017), we see that in both France and Germany, the post-birth transitions in the labour market seem to be influenced, in part, by the same variables. This does not mean, however, that the duration of post-natal leave and the proportion of consecutive births or mothers that work in part time are the same in both countries.

Results such as the ones we found might help individuals make choices that align better with their preferences, as well as help decision makers draft new policies.

Potential future research could focus on the effects of internal companies policies or branch agreements on how mothers return to employment. Understanding the different effects of these specific policies could help the state to find optimal policies for parents. Other sources of information could also be used such as survey data. This would be especially useful if these two sources, survey and administrative, could be merged. From the survey data, it would be possible to have information such as the health of the mother and of her children, commuting times, ability to choose working hours, etc. However, matching survey data with an administrative source can be very challenging, especially if a common identifying variable is missing.

4.7 Appendix

Figure 4.5 shows the hazard functions of the different risks.

FIGURE 4.5: Hazard function for each risk

Bibliography

Melanie Arntz, Stephan Dlugosz, and Ralf A. Wilke. The sorting of female careers after first birth: A competing risks analysis of maternity leave duration. *Oxford bulletin of economics and statistics*, 79(5):689–716, 2017.

Yukiko Asai. Parental leave reforms and the employment of new mothers: Quasi-experimental evidence from Japan. *Labour Economics*, 36:72–83, 2015.

Olivier Chardon and Fabienne Daguet. Enfants des couples, enfants des familles monoparentales. des diffécerences marquées pour les jeunes enfants. *Insee Première*, (1216), 2009.

Pauline Domingo and Céline Marc. Trajectoires professionnelles des mères : Quels effets des arrêts et réductions d'activité ? Technical Report 108, 2012.

Per-Anders Edin. Unemployment Duration and Competing Risks: Evidence from Sweden. *The Scandinavian Journal of Economics*, 91(4):639–653, 1989.

Jason P. Fine and Robert J. Gray. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, 94(446):496–509, 1999.

Bernd Fitzenberger, Susanne Steffes, and Anthony Strittmatter. Return-to-job during and after parental leave. *The International Journal of Human Resource Management*, 27(8):803–831, 2016.

Nicolas Frémeaux and Arnaud Lefranc. Assortative Mating and Earnings Inequality in France. IZA Discussion Papers 11084, Institute for the Study of Labor (IZA), October 2017. URL https://ideas.repec.org/p/iza/izadps/dp11084.html.

Johannes Geyer, Peter Haan, and Katharina Wrohlich. The effects of family policy on maternal labor supply: Combining evidence from a structural model and a quasi-experimental approach. *Labour Economics*, 36:84–98, 2015.

Bob Gray. *cmprsk: Subdistribution Analysis of Competing Risks*, 2014. URL https://CRAN.R-project.org/package=cmprsk. R package version 2.2-7.

Hélène Guedj. Le taux d'emploi des hommes et des femmes. Des écarts plus marqués en équivalent temps plein. *Insee Première*, (1):462, 2013.

Maia Güell and Barbara Petrongolo. How binding are legal limits? Transitions from temporary to permanent work in Spain. *Labour Economics*, 14(2):153–183, 2007.

Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

David G. Kleinbaum and Mitchel Klein. *Survival Analysis: A Self-Learning Text*. Springer Science and Business Media, LLC, 2005.

Rafael Lalive and Josef Zweimüller. How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments. *The Quarterly Jounral of Economics*, 124(3):1363–1402, August 2009.

Rafael Lalive, Analía Schlosser, Andreas Steinhauer, and Josef Zweimüller. Parental Leave and Mothers' Careers: The Relative Importance of Job Protection and Cash Benefits. *The Review of Economic Studies*, 81(1):219–265, 2014.

Arleen Leibowitz, Jacob Alex Klerman, and Linda J. Waite. Employment of New Mothers and Child Care Choice: Differences by Children's Age. *The Journal of Human Resources*, 27(1):112–133, 1992.

Brian P. McCall. Unemployment Insurance Rules, Joblessness, and Part-Time Work. *Econometrica*, 64(3):647–682, 1996.

Chiara Mussida. Unemployment duration and competing risks: A regional investigation. DISCE - Quaderni del Dipartimento di Scienze Economiche e Sociali dises0749, 2007.

Wiji Narendranathan and Mark B. Stewart. Modelling the Probability of Leaving Unemployment: Competing Risks Models with Flexible Base-Line Hazards. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 42(1):63–83, 1993.

Ariane Pailhé and Anne Solaz. Inflexions des trajectoires professionnelles des hommes et des femmes après la naissance d'enfants. *Recherches et prévisions*, 90(1):5–16, 2007.

Pedro Portugal and John T. Addison. SIX WAYS TO LEAVE UNEMPLOYMENT. *Scottish Journal of Political Economy*, 55(4):393–419, 2008.

Bruno Rodrigues and Vincent Vergnat. The impact on wages and worked hours of childbirth in France. *Working Paper BETA*, (2016-48), 2016.

Terry M. Therneau. *coxme: Mixed Effects Cox Models*, 2015a. URL https://CRAN.R-project.org/package=coxme. R package version 2.2-5.

Terry M Therneau. *A Package for Survival Analysis in S*, 2015b. URL https://CRAN.R-project.org/package=survival. version 2.38.

Sabrina Volant. Un premier enfant à 28,5 ans en 2015 : 4,5 ans plus tard qu'en 1974. *INSEE PREMIERE*, (1642), 2017.

Bingqing Zhou and Aurelien Latouche. *crrSC: Competing risks regression for Stratified and Clustered data*, 2013. URL https://CRAN.R-project.org/package=crrSC. R package version 1.1.

Chapter 5

Version control systems to facilitate research collaboration in economics

5.1 Introduction

Git is a version control system developed by Linus Torvalds around 2005 to make developing the Linux kernel easier. The Linux kernel is without a doubt the most successful piece of Free Software. The Free Software movement was officially started in 1983, with the announcement of the GNU project by Richard Stallman.[1] Before the GNU project, source code was usually distributed alongside compiled binaries, but when companies understood that by not sharing the source code of their commercial software they could have a competitive advantage over their competitors, this practice was stopped. In 1998, another movement, the Open Source Initiative, was founded by Bruce Perens and Eric S. Raymond. The Open Source Initiative argues that software for which the code is available will eventually perform better and be developed faster thanks to the so-called Linus law, named after Linus Torvalds. If enough people can look at source code and submit patches, eventually the software will be practically bugfree.

Interestingly, Free Software, or Open Source software development looks very much like scientific development. Software developers usually start working because they want to *scratch an itch*[2], they check whether someone else worked on easing this itch, and if not, they propose a solution of their own, or extend an existing solution that does not quite ease their itch. It is very much the same for science; a researcher starts with a question, reads the related literature on the matter, and if his curiosity is still not

[1] https://www.gnu.org/gnu/initial-announcement.html

[2] This expression is used by Eric S. Raymond in his book, *The Cathedral and the Bazaar* to explain the motivation software developers need to have to start working on a project.

satisfied, he will work on extending the current literature. This could not work without available literature for the researcher, or available software and source code for the software developer. In recent years, the line between the software developer and the researcher has been getting thinner; for an economist or econometrician, a lot of time writing a paper is actually spent on writing code. Cleaning data sets, computing descriptive statistics, plotting data, running regression and then putting all these results together and writing a paper using LaTeX. The amount of code an economist or econometrician writes for a paper today keeps getting bigger, more complicated and more sophisticated. As Koenker and Zeileis (2009) put it "[...] software development is no longer something that should be left to specialized commercial developers, but instead should be an integral part of the artisanal econometric research process." Authors also have been concerned with reproducibility in economics for a long time now. Dewald et al. (1986) tried to ask co-authors for data so that they could replicate their findings. Most of the polled authors were either unable, or unwilling, to share their data and source code. When authors did share their data and code, their data was usually poorly documented (sometimes completely undocumented), and the source code was for non free software that was not shareable (and the source code itself was often not very well documented either). Dewald et al. (1986) also find that errors in studies are the rule rather than the exception. Dewald et al. (1986) thus recommended to journals to urge authors to systematically share their data and source code, and lament the lack of alternatives to non-free programs used at the time for empirical research in economics. Almost thirty years later, much has changed, though. The importance of sharing the source code to a paper has been recognized by some authors that host the code to their papers on their own personal webpages. Even the Federal Reserve released the source code to its internal DSGE model; critiques of the model and patches will surely follow in the next few months, making the model more accurate and perform better. Journals also recommend, and sometimes require, the authors to share their source code. Stodden et al. (2013) gives a detailed overview of this growing trend among journals.

Tools have also evolved since Dewald et al. (1986). We now have powerful and credible free software alternatives to non-free programs. Storing and sharing data is also much easier now than in 1986, a time where floppy disks were ubiquitous but also not very reliable and the world wide web was not invented yet.

This paper will present Git, which solves most problems with collaborating with authors and with the sharing of source code to a paper. I will not argue why you should share

your code; others such as LeVeque (2013) have done it before me, and there is not much I could add to their arguments.

5.2 Scope of Git

Git is a distributed revision control system developed by Linus Torvalds in 2005 to make the collaboration of developers working on the Linux kernel easier. In 2012, the Linux foundation released a document stating that since 2005 more than 7800 developers contributed to the Linux kernel. Managing this amount of developers and patches seems impossible, and yet, the Linux kernel has been worked on for more than 20 years now. This is made possible by revision control systems such as Git.

For a lot of research teams, the current workflow is as follows: they set up a shared folder using a file hosting service, and have to take turns to change the paper. Once one team member is done, another collaborator can then change the paper and so on. If two or more people change the paper simultaneously, there will be conflicts and either some of the changes will be lost, or there will be as many versions of the paper as needed to save all the modifications. Someone then has to manually merge the modifications back to a single file. There are a range of problems with this workflow:

- Collaborators have to take turns to modify a file.
- Some modifications can easily be overwritten and lost.
- There is no simple way to revert some modifications.
- After some time, it is quite hard to know who changed what.
- There is no simple way to compare different versions of the same file.
- If a team member has a new idea while writing the paper, he has to create a new folder with the same files and modify this copy of the file.

Git (and other distributed revision control system tools) aims to solve these issues. With Git, one's work never gets lost or overwritten. As long as a team member saved his file and *committed* his changes, it simply cannot be overwritten or erased. Team members can work simultaneously on the same file: as long as the team members don't change the same lines, the modifications will be merged without problems. Should the

team members change the same lines, Git helps to solve conflict between versions and more importantly, Git still keeps every change from every team member. Reverting to an earlier version of the project is trivial as the whole history of the project is saved automatically, with precise information on who changed what. Team members can also easily check the differences between different versions of the project. Exploring new ideas is also easily possible: instead of copying the project's folder structure and modify this copy, a team member can create a new *branch* and work on this branch without cluttering his hard disk with copies of the project, and potentially lose track of his work. If the team members are satisfied with the idea, the changes can then be merged back to the main branch, or they can simply discard the experimental branch if the idea is not satisfactory.

The whole project can be hosted on different websites, for free. There are `www.github.com` and `https://bitbucket.org/` for example. The main difference is that Bitbucket allows the creation of private repositories, so the project's code is not accessible to people outside of the project. Once the team is satisfied with the state of the project, or once they published their paper, they can make the repository public (if they wish to share their code with the rest of the scientific community). There are several reasons why it is better to host code on sites like Bitbucket and Github instead of one of the team members' personal webpage:

- If the code is hosted on Bitbucket or Github, nobody on the team needs to have a personal webpage, which is often difficult to do.
- Bitbucket and Github virtually have no downtime.
- Download links do not ever expire.
- Collaboration is easier via Bitbucket or Github. People can comment on the project's code and propose patches.
- Discoverability of the code is increased. It is not always easy to find code on an author's page.
- Finding future colleague to work on projects is easier. Websites such as Bitbucket or Github are also very dynamic social networks.[3]

[3] Github hosts over one million code repositories, and has 340.000 registered authors, Dabbish et al. (2012).

FIGURE 5.1: This figure shows the commits of a project on the website bitbucket.org where the project is hosted. Two authors are working on a paper, and it is possible to know exactly who did what and when.

With code hosted on such websites, it is possible to find code and replicate a study in a matter of minutes. Suggesting modifications or submitting patches is also very simple and done very fast. Git is also useful when working on a paper alone. It allows the researcher to keep track of his changes more easily and also revert to older versions.

5.3 Basic usage of Git

Installing Git is straightforward. On GNU+Linux operating systems, Git is often pre-installed, and if not, can be easily installed via the system's package manager. To install Git on Windows, a researcher can download the installer made by the msysGit project[4] and for OSX there is also a graphical installer.[5]. Then the team members can each create a personal account on Bitbucket or Github. By default Git is used with the system's command line interface, but there are graphical interfaces for users that are not used to working with commands. Showing how Git works would take too much space here; the interested reader is invited to consult the online appendix for a tutorial on Git. In a few sentences however, the idea is that the *project leader* initializes the project using Git. This creates a folder in which he puts the files needed for the paper, such as the data, computer code, etc. The next step is to tell Git to *track* these files (so Git knows the changes that were made to the files) and then *push* the project to the repository. The other team members can then *pull* the project and the changes that were, make their own modifications and push the changes back to the server and so on. Figure 5.1 shows

[4] http://msysgit.github.io
[5] http://sourceforge.net/projects/git-osx-installer/

the commits log for a project on bitbucket.org. Using the commit's hash (the number in the second column), it is possible to revert to the state the project was at that point.

Table 5.1 summarizes the basic commands to use Git:

TABLE 5.1: Basic Git commands

Common commands	Common options	Usage[6]
`git init`	None	Initializes a repository on your local machine.
`git log`	`-p -N`	This command shows commit[7] logs. `-p` shows the differences between commits and `-N` shows the last commits.
`git status`	None	Shows the current status of your local working directory. Added files, untracked files, modified files, etc.
`git add`	`-n`	This command adds the changes to the next commit. You can specify single files by giving their exact names, or you can add every change to the next commit with `git add .`. The `-n` option does not actually add the files. This option is also called `--dry-run`.
`git commit`	`-a -m`	This command commits the changes along with a log message describing the changes. The `-a` option automatically adds every modification to the commit (thus no need to run `git add .` before). The `-m` option allows the user to add a commit message.
`git push`	`-n`	Pushes the commit to the remote repository. The `-n` option does not actually update the remote repository.

Researchers in other disciplines have already acknowledged the usefulness of version control systems such as Git (Ram 2013 and Peng 2011 for instance) and economists are also starting to use Git and Github for their papers (Aruoba and Fernández-Villaverde 2014). [8] Git has a steep learning curve, and it may seem useless to learn to master it. But the more complicated and the more researchers are invested in a project, the more Git streamlines the workflow of the team. With just the few commands from Table 5.1 it is already possible to start using Git efficiently. To learn more about Git, one can consult the online documentation[9] or read books about Git such as Loeliger and McCullough (2012).

5.4 Conclusion

Version control systems such as Git are tools that make collaboration among software developers very easy and fast. The current tools used by most research teams to collaborate are not really suited because economists need to write always more complex source code for their papers. Using file synchronization tools or e-mailing files around among co-authors can get very messy and difficult to keep track of. A tool such as Git could very well be adopted by economists (that practice applied econometrics or not) to help them collaborate with each other. Mastering Git may be hard, because of its steep learning curve. But the benefits and time savings in the long run more than make up for it. Hosting the code on a website such as Bitbucket also makes sharing the source code to a paper very simple. Once the authors are ready (or required by a journal) to share the code a paper, they can publish the repository, and anyone can download the code on their machines, and propose patches to the authors if they find bugs. This paper only presented Git, but other version control systems exist. Some examples are Mercurial and Apache Subversion.

[8] You can access the authors' Github repository here: https://github.com/jesusfv/Comparison-Programming-Languages-Economics. All the source code discussed in the paper is available to download and numerous suggestions to make the code run faster have been made by other developers, that are thanked in the current version of the paper.

[9] https://git-scm.com/documentation

Bibliography

S. Boragan Aruoba and Jesús Fernández-Villaverde. A Comparison of Programming Languages in Economics. 2014.

Laura Dabbish, Colleen Stuart, Jason Tsay, and Jim Herbsleb. Social coding in GitHub: transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 1277–1286. ACM, 2012.

William G. Dewald, Jerry G. Thursby, and Richard G. Anderson. Replication in Empirical Economics: The Journal of Money, Credit and Banking Project. *The American Economic Review*, 76(4):pp. 587–603, 1986. ISSN 00028282.

Roger Koenker and Achim Zeileis. On reproducible econometric research. *Journal of Applied Econometrics*, 24(5):833–847, 2009.

Randall J. LeVeque. Top Ten Reasons to Not Share Your Code (and why you should anyway), 2013.

Jon Loeliger and Matthew McCullough. *Version Control with Git: Powerful tools and techniques for collaborative software development.* " O'Reilly Media, Inc.", 2012.

Roger D. Peng. Reproducible Research in Computational Science. *Science*, 334(6060): 1226–1227, 2011.

Karthik Ram. Git Can Facilitate Greater Reproducibility and Increased Transparency in Science. *Source Code for Biology and Medicine*, 8(1):7, 2013. ISSN 1751-0473.

Victoria Stodden, Peixuan Guo, and Zhaokun Ma. Toward reproducible computational research: an empirical analysis of data and code policy adoption by journals. *PloS one*, 8(6):e67111, 2013.

Chapter 6

Conclusion

This short chapter will conclude the thesis. This thesis contributes to the literature on female labour supply and fertility decisions in different manners, using different data sources and different models.

The first, in Chapter 2 dealt with regional differences in number of children between French and German regions. This chapter used data from the LIS data center in Luxembourg which is very well suited for internatiol comparisons. The shortcomings of these data though is that the data are cross-sectional in nature and the dependent variable is number of children living in the household at the time of the study. However, despite these shortcoming, this chapter shed some light into wether there is something about French regions that makes TFR systematically higher than in German regions. After controlling for individual characteristics, and introducing a random intercept to a hierarchical Poisson model, I showed that that there were indeed differences between French and German regions. The random intercept is statistically significant and positive for most French regions, while not statistically significant for German regions. This study, along with others shows that it is important to consider regional variations when studying fertility decisions. Expanding this study could be done by including data measured at the country-level and thus considering a three-level hierarchical model. Individuals would be nested inside the NUTS 1 regions, and the regions nested inside countries. Data on attitudes towards children, religiosity and other such values could be included as country-level regressors. Also, extending the number of countries and thus regions would also be of interest. France and Germany were chosen as, examples of a high TFR country and a low TFR country, but it would be interesting to extend the analysis to more countries, and to more recent years. At the time of writing, 2010 is the last year

available for France, but there is data for 2013 and 2015 for Germany. Most other European countries have data for 2013 available too. The best case scenario however would be to use panel data. I am not aware of a source with information on labour supply, income (not just work income, but also capital income for instance), socio-demographic variables, region of residence, in a panel format, that is internationally comparable. The LFS data comes close to it, but income is only available in deciles, and worse, the data is not longitudinal.[1]

The second contribution[2], in Chapter 3, aimed to determine the impact of childbearing for French women on hours worked and hourly wages. The impact was studied for different education levels and maternity or paternity leave durations. The DADS-EDP data, made available by the INSEE through the CASD was used for this Chapter. These data are very rich in information, but not easy to manipulate. Women and men were split in lower educated and highly educated groups and in control and treatment groups. Then, a difference in differences procedure showed that women, not men, decrease their supplied worked hours and that the loss in wages is proportional to the education level of the mother. As in other studies, the results do not indicate a premium for men. To study subsequent births, women with one child were used as the control group (while still controlling for education and other variables) and women with two children as the treatment group. The same was done to study women with three children. Parallel trend graphs before the treatment are shown in the appendix of the Chapter. They show that before treatment, the subjects in the control and treatment groups were in most cases on the same path. However, the treatment being endogenous, we cannot state that our results are causal. However, by constructing the control and treatment groups the way we did, we feel confident that we removed a lot of sources of bias and thus the results do help us in evaluating the impact of childbearing. The results show that highly educated mothers suffer a higher penalty than lower educated ones, and thus the results raise the question of whether the compensation for maternity leave should be proportional to the pre-birth wage of the mother. However, this might not be enough to thwart the negative impact entirely. There are other non-pecuniary aspects that can impact a woman's career after she gave birth to a child, such as being the one having to pick up the child at school, and having to spend more time in general to take care of the child.

[1] This is not entirely accurate; there are 2 quarter and 5 quarter longitudinal versions of the LFS.
[2] Derived from a paper co-authored with Vincent Vergnat.

The third contribution[3], in Chapter 4 also uses the DADS-EDP data to study the time and transitions back to the labour market after giving birth. To do so, we estimated a competing risks model, where each risk stand for different ways a woman can return to a job, which may be her previous job. The results show that women that have been working with the same employer or in large firms tend to return to their previous employer, and in full-time work. The pre-birth wage also plays an important role, and employers that wish to retain competent female employees should keep this in mind. If a mother feels that the wage she was earning before giving birth is not high enough, she will most likely decide to switch jobs, or stay home (longer that she would otherwise have). The data source used, however, might not be the best for this kind of study because important information, such as the health of the mother, or complications that arose during the birth are not available. Commuting time might also play a large role, but such information is also missing from typical administrative data. I am not aware of a dataset that would combine precise information on hours worked such as the DADS-EDP and such other variables, though. However, if it would be possible to collect such information and combine it to the richness of an administrative dataset such as the DADS-EDP, the potential research that would emerge from it would be of great value.

The last contribution, in Chapter 5 does not deal with female labour supply or fertility, but deals with an important topic that affects any research topic in economics, and in any other discipline. The chapter discusses a tool for version control called Git, which can be used to streamline the production of scientific papers. As explained in Chapter 5 software development is no longer a skill that is only required of software engineers, but also from scientists, and social scientists as well. The source code of all the computations that produced the results of this thesis are available online[4], and the amount of lines of code is quite substantial. Keeping track of all this code, and being able to share it with the wider scientific community is as important as the papers that compose this document. Also, two of the chapters of this thesis are derived from co-authored articles, which were in part written after I changed institutions in 2016. Using a version control tool such as Git made the writing process much smoother.

[3] Also derived from a paper co-authored with Vincent Vergnat.
[4] https://bitbucket.org/b-rodrigues/

Résumé de thèse: Essais sur l’offre de travail et la fécondité des femmes

Chapitre 1

Cette thèse porte sur l'offre de travail et les décisions de fécondité des femmes en France.
Chaque chapitre correspond à un document de recherche; chacun de ces documents de recherche contient une revue de littérature détaillée. Le chapitre 1 quant à lui présente les différentes sources de données qui ont été utilisées pour rendre cette thèse possible. En effet, un important élément du travail empirique est évidemment la disponibilité des données. La phrase précédente semble évidente, cependant, l'accès aux données est toujours un obstacle majeur que les économistes appliqués doivent surmonter.

Les économistes et autres spécialistes des sciences sociales se heurtent à un obstacle majeur qui n'existe pas dans d'autres disciplines telles que les sciences de la vie : à savoir l’accès à des données de haute qualité pour mener des études empiriques. Dans les sciences de la vie et d'autres disciplines, comme l'apprentissage machine, la compilation de données peut être une tâche complexe, mais souvent les chercheurs ne sont pas confrontés à des obstacles administratifs.
Par exemple, un data scientist s'intéressant à l'analyse des sentiments après un débat politique, peut simplement utiliser des données de réseaux sociaux tel que Twitter. Un physicien intéressé par le comportement d'une certaine particule n'aurait qu'à observer et à utiliser des outils pour mesurer cette particule. La phrase précédente est une simplification excessive, bien sûr, car il y a aussi des obstacles juridiques et administratifs. Les économistes, cependant, et principalement les micro-économistes, sont confrontés à un problème très pratique concernant les données : les questions de protection de la vie privée.

Par exemple, un économiste du travail intéressé par l'offre de main-d'œuvre des femmes, devrait avoir des données sur des milliers de salaires, de niveaux d'éducation, de fécondité et d'autres variables similaires pour les femmes, dont la plupart pourraient être sensibles. Un économiste intéressé par l'organisation industrielle aurait besoin d'avoir accès à des données comptables sur des milliers (ou beaucoup plus) d'entreprises.

Il s'agit généralement de données très sensibles dont les organismes les possédant sont très réticents à donner accès à des tiers. Cela signifie que l'accès à ce type de données est en général réglementé et seulement possible dans le cadre d'un projet de recherche bien spécifique. Afin de mieux protéger les individus, un nouveau règlement européen, le Règlement Général sur la Protection des Données qui est entré en vigueur en Mai 2018 donne beaucoup plus de contrôle aux individus sur leurs données personnelles. Les particuliers seront également en mesure de refuser que leurs donnée soient utilisées sans leur consentement. La collecte de caractéristiques individuelles sensibles, telles que celles énumérées à l'article 9 de la réglementation ([...] origine raciale ou ethnique, opinions politiques, religieuses ou philosophiques les croyances, ou l'appartenance syndicale, et le traitement des données génétiques, des données biométriques, etc. aux fins d'identifier de manière univoque une personne physique, les données relatives à la santé ou les données concernant concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique) sera interdite.
Toutefois, ceci ne devrait pas être un frein pour la recherche, car l'article 9 stipule en outre qu'un tel traitement des informations est licite s'il est effectué à des fins d'archivage dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins de statistiques officielles.
La recherche économique est sans doute dans l'intérêt public.

Deux sources de données ont été utilisées pour rédiger les chapitres de cette thèse, à savoir les données du LIS et les données DADS-EDP.

Les données du LIS

Dans le Chapitre 2, j'estime un modèle hiérarchique en utilisant des micro-données Françaises et Allemandes. Des problèmes de comparabilité peuvent émerger, étant donné que la façon de collecter des données se fait différemment d'un pays à l'autre. Pour pallier à ce problème, j'utilise des données du LIS Cross-National Data Center qui se trouve à Esch-Belval, au Grand-Duché de Luxembourg. L'accès aux données se fait à distance.

Une fois une demande d'accès aux données acceptée, le chercheur obtient un login ainsi qu'un mot de passe. Par ce biais, le chercheur peut se connecter à LISSY ; LISSY est un logiciel écrit en JAVA par l'intermédiaire duquel il est possible d'envoyer du code informatique écrit en R, STATA ou SAS. Le code tourne sur les serveurs du LIS, et les résultats sont renvoyés au chercheur presque en temps réel. Il est toujours possible d'envoyer le code par courriel aussi, mais cette méthode est moins pratique. Par rapport à un modèle comme celui du CASD (décrit dans la prochaine section), il est possible d'obtenir d'étudier les données très rapidement, car l'accès au données se fait très vite.

L'équipe du LIS met en œuvre beaucoup d'efforts pour harmoniser les différentes sources de données et pour imputer les données manquantes. Généralement, les données du LIS ont été utilisées pour

étudier les inégalités, mais contiennent beaucoup d'autres informations qui permettent de répondre à d'autres questions empiriques.

Les données DADS-EDP

Les données DADS-EDP, accessibles via le CASD[1], sont composées de deux sources, les DADS[2] et l'EDP[3]. Les DADS sont un jeu de données administratif contenant des informations sur les personnes employées ; leur niveau de salaire, le type de contrat, leur secteur d'activité, la taille de l'entreprise dans laquelle travaille la personne, les dates de début et de fin de contrat, nombre d'heures travaillées, et bien plus encore. L'EDP est un panel avec des informations d'ordre démographique comme le statut marital, le nombre d'enfants, le niveau d'étude accompli, le lieu de résidence... Le jeu de données DADS-EDP exploité dans cette thèse ne contient pas d'informations sur les employés du secteur public, personnes n'ayant jamais travaillé ni sur les personnes travaillant à leur compte. Les Français nés à l'étranger ne sont pas inclus non plus.

L'accès à ces données demande plus de patience que pour celles du LIS, mais cela s'explique par le fait que, dans le cas des données du CASD, le chercheur peut voir les données individuelles contrairement à celles du LIS, ou le chercheur ne peut pas voir les données. En effet pour le LIS, le chercheur n'aura jamais que les résultats de son analyse. Un chercheur qui a accès aux données via le CASD peut les manipuler comme si les données étaient stockées localement sur son ordinateur personnel. Toutefois ce n'est pas le cas, cependant, les données ne sortent jamais des serveurs du CASD. Les chercheurs qui souhaitent avoir accès aux données doivent soumettre une proposition qui doit être approuvée par un comité. Une fois le projet approuvé, les chercheurs doivent se rendre à Paris pour signer d'autres papiers, et faire enregistrer une de leurs empreintes digitales. L'empreinte digitale est nécessaire pour se connecter au système. Les chercheurs reçoivent également un cours accéléré en secret statistique. Une fois cela fait, les chercheurs reçoivent un clavier spécial à l'adresse de leur institut de recherche. Une fois le clavier branché à Internet et à un écran d'ordinateur, les chercheurs peuvent se connecter au système à l'aide d'une carte-clé reçue lors de la séance de formation et en utilisant aussi l'empreinte digitale qui a été enregistré lors du passage à Paris. Une fois que cela est fait, le chercheur est connecté à une machine virtuelle Microsoft Windows qui contient les données. De plus, le chercheur a accès à des logiciels statistiques, tel que GNU R, SAS ou STATA si le chercheur en fait la demande. Les données des chapitres 2 et 3 ont été consultées de cette façon. Cette procédure peut sembler fastidieuse, mais est nécessaire pour protéger la vie privée des citoyens.

[1] Centre d'accès sécurisé aux données
[2] Déclaration annuelle des données sociales
[3] Échantillon démographique permanent

Les logiciels de gestion des versions

Ce court chapitre discute et présente les logiciels de gestion de versions et explique leur intérêt dans le cadre de la recherche en comparant le processus de recherche avec le processus de création de logiciels libres. Énormément de points communs existent qui justifient l'utilisation de ce type de logiciel afin de rendre la recherche plus transparente et reproductible.

Les chapitres de cette thèse

Chaque chapitre traite de l'offre de travail et des choix de fécondité des femmes, mais d'un point de vue différent. Le chapitre 2 vise à voir s'il y a différences entre les régions françaises et allemandes en ce qui concerne le nombre total d'enfants que les femmes décident d'avoir. Ces deux pays ont été choisis parce qu'ils se ressemblent à bien des égards, mais sont aussi très différentes lorsqu'il s'agit de la participation des femmes au marché du travail et les choix en matière de fertilité.

Le chapitre 3 étudie l'impact d'une première, d'une deuxième et d'une troisième naissance sur l'offre de travail et les salaires des femmes et des hommes. Il s'agit ici de voir dans quelle mesure ces naissances ont un impact négatif, ou peut-être dans le cas des hommes positif, sur les salaires et heures travaillées.

Le chapitre 4 étudie le temps que les femmes passent hors du travail après une naissance et la façon dont elles reviennent au travail. Les femmes peuvent décider de revenir à temps plein ou partiel, chez leur employeur précédent ou changer de travail. Les chapitres 3 et 4 n'utilisent que des données françaises.

Chapitre 2

Le chapitre 2 vise à voir s'il y a différences entre les régions Françaises et Allemandes en matière de fécondité.

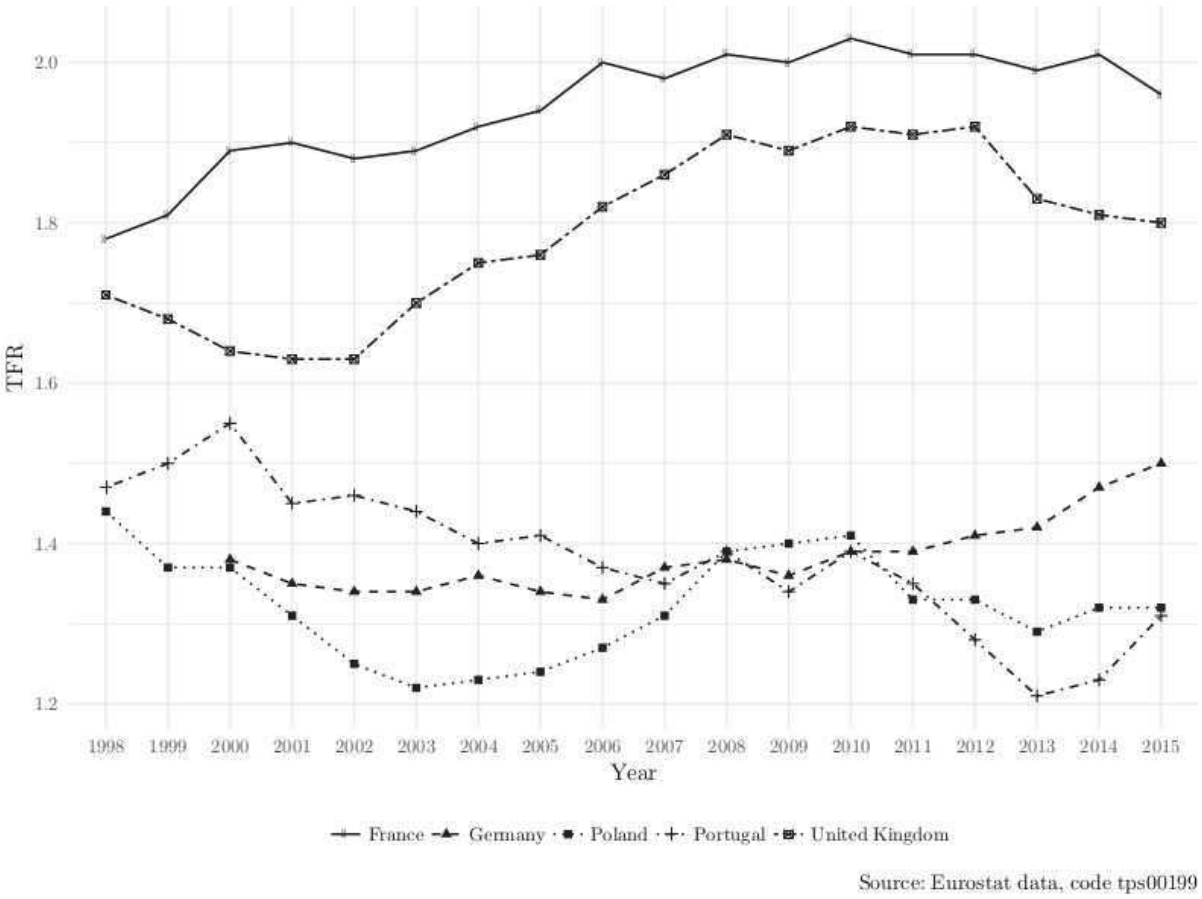


Figure 1: Taux de fécondité pour une sélection de pays européens

La figure 1 suggère qu'il existe deux types de pays concernant le taux de fécondité, des pays à taux faible et des pays à taux élevé. La France et l'Allemagne ont été choisis comme représentants de ces pays. Ces deux pays ont été choisis car ils sont similaires à bien des égards, mais ils diffèrent aussi beaucoup quand il s'agit de la participation des femmes au marché du travail et au taux de fécondité. Ce chapitre utilise les données du LIS, qui contient des observations au niveau de l'individu, dont sa région de résidence, et je rajoute des données d'Eurostat au niveau régional pour ainsi pouvoir estimer un modèle hierarchique. Les données contiennent finales utilisées pour l'estimation du modèle contiennent 15'946 observations.

Le but de l'exercice est de voir de quelle manière les régions Françaises et Allemandes différent, une fois que l'on contrôle pour des variables individuelles.
Dans la littérature, Klüsener et al. (2013) expliquent que ce type de différences ont une nature institutionelle plutôt que culturelle. Leur étude est basée sur des données d'Allemagne de l'Ouest. Toutefois, pour Goldstein et al (2003) le taux de fécondité bas en Allemagne est un phénomène culturel. Dans leur étude, ils montrent, en utilisant des données de 2001 de l'Eurobaromètre, que la taille de famille idéale en Allemagne et en Autriche est plus faible que la taille de famille idéale en France (1.7 contre 2.5). Toutefois, en pratique, la taille de famille moyenne est plus basse dans tous ces pays que la taille idéale.

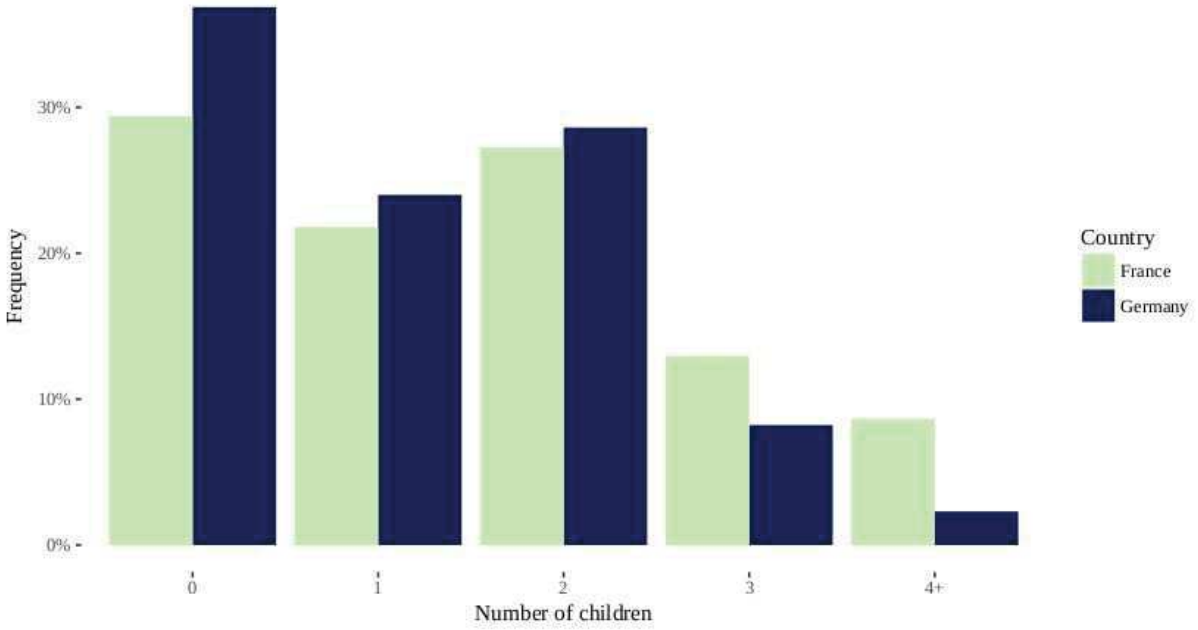


Figure 2: Distribution de la variable dépendante

Un modèle de poisson hiérarchique est estimé sur ces données, avec comme variable dépendante le nombre d'enfants vivant dans le foyer au moment de l'enquête. Les résultats de l'étude montrent des différences entre les régions significatives ; le nombre minimum d'enfants par région en France est toujours positif, mais négatif pour la région Parisienne. En Allemagne, ce nombre est toujours négatif ou n'est pas statistiquement significatif. Ce résultat semble indiquer qu'il y aurait un taux de fécondité de référence différent entre les deux pays, mais aussi au sein des pays entre les différentes régions.

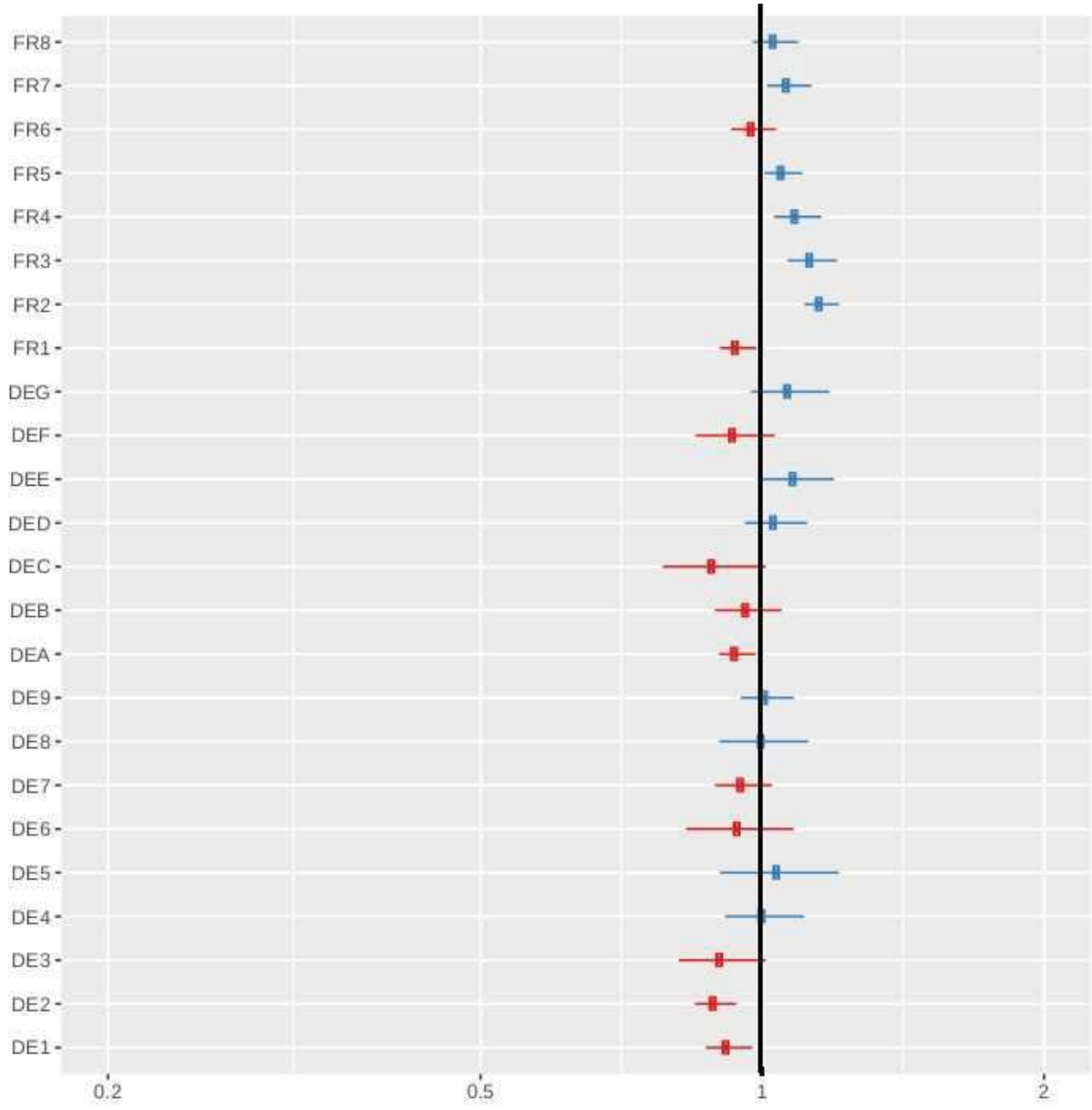


Figure 3: Incidence régionale

Chapitre 3

Ce chapitre étudie l'impact d'une première, d'une deuxième et d'une troisième naissance sur l'offre de travail et les salaires des femmes et des hommes. La question ici est de voir dans quelle mesure ces naissances ont un impact négatif, voire peut-être même positif dans le cas des hommes

Les pays industrialisés sont tous confrontés à une faible fécondité. La France semble être l'exception à cette règle cependant. En France, les taux de fécondité sont élevés par rapport aux autres pays industrialisés comme l'Allemagne, mais toujours en dessous du seuil de remplacement de 2.1.

Ce faible taux de fécondité dans les pays industrialisés peut être lié à l'émancipation de la femme. Gayle et Miller (2006) expliquent qu'aux États-Unis, la femme moyenne avait quatre
enfants au début du XXe siècle, mais ce nombre est tombé à 1,9 à la fin du siècle dernier. Cette diminution a été également accompagnée par une augmentation de l'offre de travail des femmes. Selon Gayle et Miller (2006), " la participation de toutes les épouses a augmenté de 36 % au cours des 25 dernières années, les taux de participation des mères avec des enfants de moins de trois ans a augmenté de 83 %, et de 91 % pour les femmes avec des enfants d'un an ou moins " (p. 2). Cet article se base en partie sur les travaux de Duguet et al. (2015), qui utilisent la méthode des doubles différences afin d'estimer ce family gap pour les femmes et les hommes du secteur public et privé.

Le jeu de données utilisé dans ce chapitre est le "DADS-EDP" qui est un panel composé de deux sources, à savoir le panel DADS apparié avec l'ÉDP. Ce jeu de données est mis à disposition des chercheurs par l'INSEE via la platforme CASD.

Le DADS est un jeu de données administratives comprenant des informations sur les salaires, le type de contrat de travail, le secteur d'emploi, la taille de l'entreprise, l'entreprise dans laquelle la personne travaille, les dates de début et de fin de la période de rémunération le travail, le nombre d'heures rémunérées, etc.... Chaque année, les entreprises doivent faire une déclaration pour chacun de leurs employés. Tous les travailleurs en France sont couverts par ces déclarations, à l'exception des salariés des administrations publiques, des indépendants et des salariés de l'administration française
entreprises établies à l'étranger.

L'EDP est un panel qui contient des informations sur l'état matrimonial, la fécondité, les diplômes obtenus et l'état civil, le lieu de résidence. De 1967 à 2004, les personnes nées entre le 1er et le 4e jour du Octobre sont dans l'EDP. Depuis 2004, l'ensemble de données s'est enrichi de personnes nées du 2 au 5 Janvier ainsi que ceux nés entre le 1er et le 4 Avril et Juillet. Les données sont recueillies à partir des registres de l'état civil chaque année, ainsi qu'auprès du recensement, chaque fois que cela est nécessaire pour compléter les informations provenant des registres de l'état civil.

Par exemple, pour les personnes nées entre 1989 et 1997, les informations sur leurs enfants provient exclusivement du recensement. Pour la plupart des personnes nées entre 1982 et 1989, cela est aussi souvent le cas. Avant 2004, seules les personnes vivant en France métropolitaine étaient incluses dans l'ÉDP. Depuis 2004, les personnes résidant dans les territoires français d'outre-mer sont également

incluses. Les données brutes sont dans un format qui ne se prête pas à une analyse économétrique, et donc un certain travail de préparation et nettoyage des données à été effectué. Après nettoyage, le panel comporte 409'107 personnes au total couvrant les années 1976 à 2010, dont 194'956 sont des femmes et 214'151 sont des hommes.

Afin d'appliquer la méthode des doubles différences, un groupe dit de traitement et un autre groupe dit de contrôle ont été créés. Le groupe de contrôle est composé d'individus aux caractéristiques similaires à ceux du groupe de traitement, mais avec un enfant de moins. Ces groupes sont comparés, deux, quatre, et six années après le traitement, à savoir la naissance d'un enfant.

Les résultats de ce chapitre montrent que la perte en salaire est généralement proportionnelle au niveau d'éducation de la mère. Pour les femmes ayant au plus un diplôme d'études secondaires les résultats indiquent qu'il n'y a pas de pénalité dans le salaire horaire et qu'il y a une prime quatre ans après la naissance de leur enfant, mais seulement pour les femmes qui ont pris un congé de maternité de courte durée. Pour les femmes ayant plus un diplôme d'études secondaires, nous constatons un impact négatif, mais seulement pour les femmes qui sont restées plus de 6 mois hors du marché du travail. Pour les femmes qui sont restées en dehors du marché du travail entre 6 mois et deux ans et demi, l'impact négatif de l'accouchement s'accroît au fil du temps. Les femmes qui ont arrêté de travailler moins de 6 mois ne sont pas pénalisées. Cela pourrait être la preuve d'une dépréciation du capital humain qui ne se produit que si les femmes restent en dehors du marché du travail pour de plus longues périodes de temps. Les heures de travail diminuent considérablement pour les femmes, quelque soit leur niveau d'éducation. L'impact de la naissance sur l'offre de travail ne semble pas transitoire.

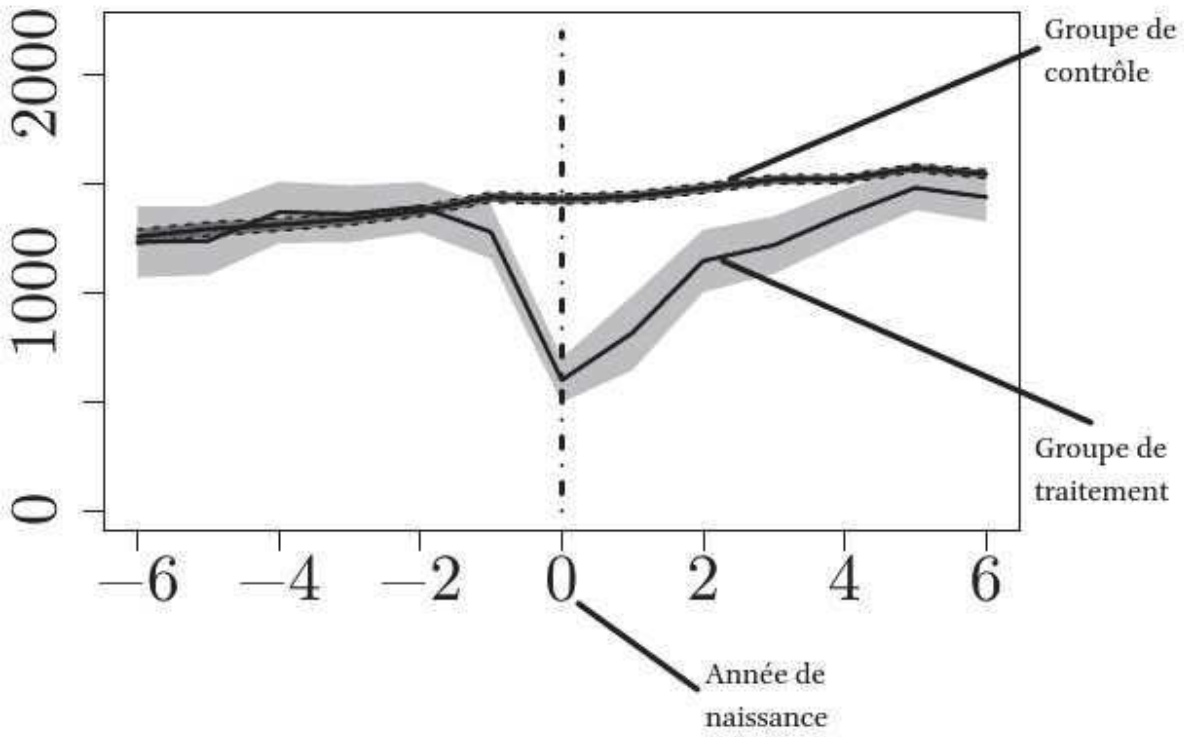


Figure 4: Exemple de l'impact d'une première naissance sur les heures travaillées pour les femmes qui ont arrêté entre 6 mois et 2 ans et demie de travailler

Pour les femmes qui ont deux, voire trois enfants, l'impact négatif sur les heures travaillées s'accroît. Pour les pères, nos résultats ne montrent pas qu'ils gagnent systématiquement plus lors de la naissance de leur enfant, résultat récurrent de la littérature, voir Duguet et al. (2015) pour un exemple. Ces résultats sont alignés avec ceux de Wilner (2016).

Chapitre 4

Le Chapitre 4 étudie la durée qu'une mère passe hors du travail après la naissance de son premier enfant, et comment elle décide de retourner en activité. Un modèle de risques concurrents, similaire à celui présenté dans Arntz et al. (2017) et présenté en détail dans Kleinbaum et Klein (2005) et basé sur les travaux de Fine et Gray (1999) est utilisé pour cela. En effet, une mère peut décider de revenir à son ancien employeur à temps plein ou à temps partiel, ou de changer d'employeur et d'y travailler à temps plein ou à temps partiel. Une mère peut aussi décider de ne plus retourner travailler, mais les données utilisées (DADS-EDP) ne permettent pas de connaître la raison de ce non-retour.

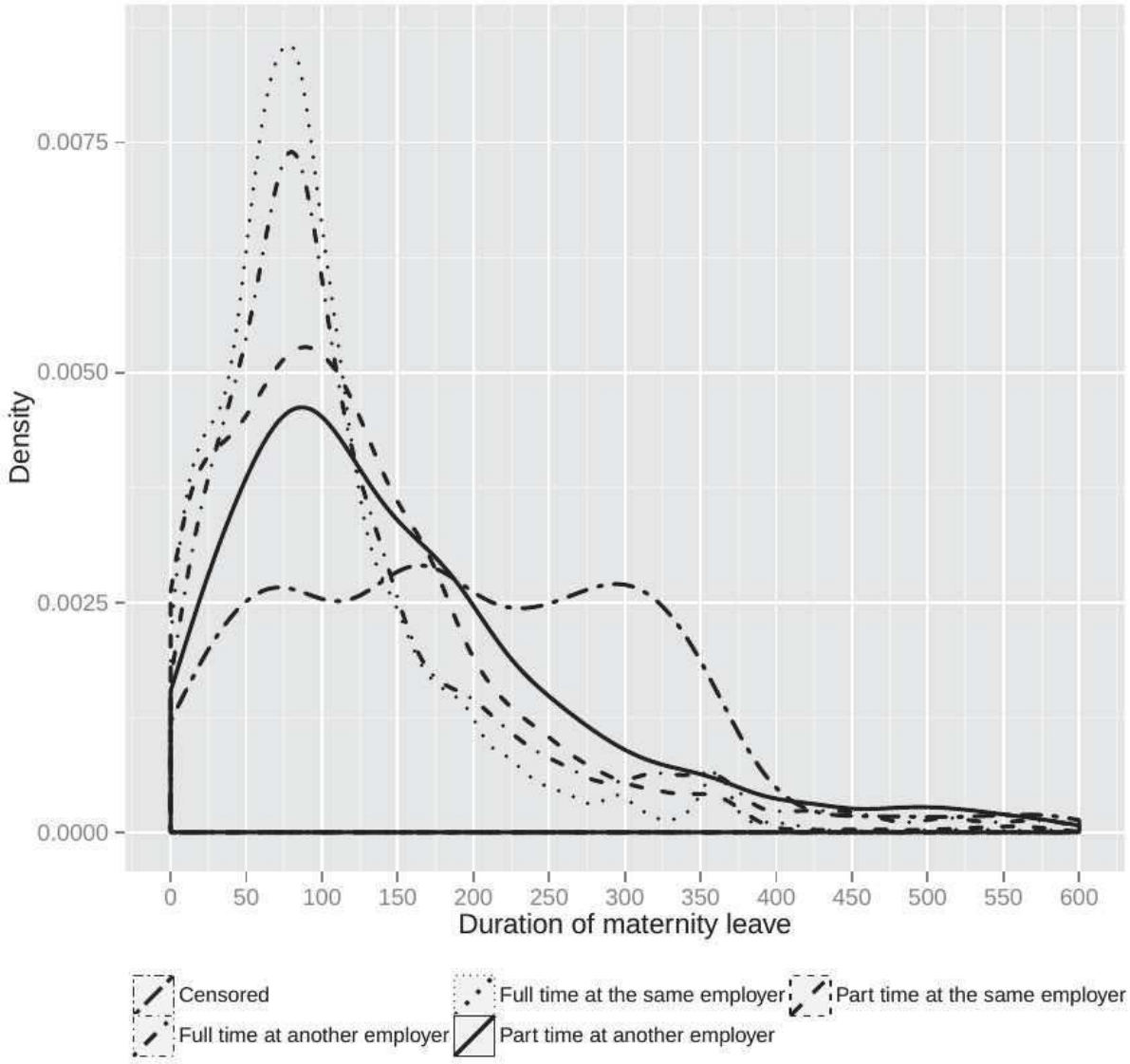


Figure 5: Distributions de la durée du congé de maternité selon le type de transition effectuée

Les résultats montrent comment les variables individuelles, professionnelles et le cadre légal influent sur la probabilité de chaque transition.

Un an d'âge supplémentaire diminue le risque des deux risques de retour à temps plein d'environ 2 à 4 %, tout en augmentant le risque de 6 % de retourner chez le même employeur à temps partiel. Les plus jeunes mères ont tendance à retourner plus souvent au travail à temps plein, peut-être parce qu'elles souhaitent limiter les pertes de salaire dues à la dépréciation du capital humain, ainsi que de signaler l'attachement à leur ancienne entreprise. Le fait d'être marié au moment de la naissance augmente le risque de retourner à temps plein chez le même employeur, ce qui reflète peut-être une meilleure répartition du temps de travail entre hommes et femmes dans les couples mariés, ce qui permet aux femmes de concilier leur vie professionnelle et leur vie privée de manière optimale. Cela peut aussi s'expliquer par le fait que les couples mariés se spécialisent avant la naissance d'un enfant. Ainsi, les femmes mariées qui ont un enfant, ont déjà décidé, au moment du mariage, d'investir dans l'éducation de leurs enfants ou dans leur carrière.
Vivre en Île-de-France, la région de la capitale, diminue le risque de retourner travailler chez le même employeur de 11 %, tout en augmentant le risque de changer d'employeur de 30 %. Ceci clairement montre que les mères qui vivent dans la région de la capitale ont beaucoup plus de possibilités d'emploi que les mères de famille dans le reste du pays.
En ce qui concerne les caractéristiques professionnelles, la titularisation augmente le risque de retour chez le même employeur d'environ 3 %, et diminue celui de changer d'employeur de près de 10 %.
Les mères qui travaillent dans la même entreprise depuis longtemps ont également tendance à revenir à leurs employeurs précédents, très probablement comme un moyen pour ces mères de signaler à leur l'attachement à leur ancien employeur (ou pour éviter d'avoir à prouver à nouveau leur valeur dans une nouvelle entreprise. Ce résultat est aussi trouvé par Arntz et al. (2017) ou Fitzenberger et al (2016). Ces femmes ont également accumulé un capital humain spécifique à l'entreprise, et changer d'entreprise rendrait ce capital humain spécifique obsolète. Peut-être de façon surprenante, l'expérience globale du marché du travail n'a pas d'impact. C'est peut-être parce que nous avons inclus l'âge et l'ancienneté en tant que variables de contrôle supplémentaires, qui sont fortement corrélées avec l'expérience du marché du travail et absorbent donc l'effet de l'expérience. Comme il fallait s'y attendre, pour les mères qui travaillaient à temps partiel dans le passé, le risque de retourner à temps plein chez le même employeur décroît, peut-être reflétant un moindre attachement de ces femmes au marché du travail.
La distance entre le domicile et le lieu de travail, bien que statistiquement significative, n'a qu'un très faible effet. Cela pourrait provenir du fait que ce n'est pas la distance en tant que telle qui a de l'importance, mais plutôt le temps qu'il faut pour se rendre au travail. Il serait très intéressant d'analyser l'impact de cette variable sur les transitions d'emploi après avoir donné les données suivantes mais une autre source de données serait nécessaire pour cela. Une possibilité serait de utiliser un système de routage tel que OSRM, mais cela n'est pas possible via le système sécurisé que nous devons utiliser pour accéder aux données.

Le salaire annuel prénatal est la variable qui a l'effet le plus important. Par rapport aux femmes dans la catégorie de référence, le risque de retourner à plein temps dans le même employeur pour les femmes gagnant entre 20 et 30'000 € par an est de 8 % plus élevé et d'environ 25% plus bas pour aller travailler à temps partiel pour un autre employeur.
Pour les femmes gagnant plus de 30'000 € annuellement, le risque de retourner à temps plein chez le même employeur est 37 % plus élevé, et le risque d'aller travailler à temps partiel pour le même employeur ou pour un autre employeur est respectivement de 31 % et 63 % plus bas. Les femmes qui gagnent des salaires plus élevés ont également tendance à retourner travailler pour leur ancien employeur à temps plein.
De plus, nous constatons que la CLCA a augmenté la probabilité pour les femmes de retourner chez leur ancien employeur. Du point de vue de la mère et de l'employeur, ceci est bénéfique, car cela permet de réduire les coûts de recherche d'emploi pour la mère, et les coûts d'embauche pour l'entreprise. Comme dans Arntz et al. (2017) nous trouvons que des nivaux de PIB élevés sont corrélés avec un retour au travail à temps partiel.

Chapitre 5

Git est un logiciel de gestion de versions développé en 2005 par Linux Torvalds afin de rendre le développement de logiciels, et notamment du noyau Linux, plus simple et transparent.
À bien des égards, la production scientifique et le développement de logiciels se ressemblent. En effet, lors d'études empiriques, un grand nombre de documents sont générés, et ce par de plus en plus de co-auteurs. Cela va du code source pour l'analyse économétrique au code source de l'article de recherche lui-même, comme expliqué dans Koenker et Zeileis (2009) :

« [...] le développement de logiciels n'est plus quelque chose qui ne concernen que le développeur commercial spécialisé, mais doit maintenant faire partie intégrante du processus de recherche en économétrie. »

Ces systèmes de gestion de versions permettent de simplifier le processus de création scientifique, mais aussi de rendre les résultats plus facilement reproductibles. Ceci est une question importante, et qui n'est pas récente. Dewald et al. (1986) ont demandé à des collègues de leur fournir les données et code sources de leurs études afin de les répliquer. La plupart de ces auteurs étaient soit dans l'impossibilité d'accéder à leur requête, soit ne voulaient pas partager leurs données et/ou code. De nos jours, beaucoup de revues demandent le code source de l'analyse, et de plus en plus de jeux de données sont en Open Access. Utiliser un système de version de contrôle comme Git permet donc de répondre aux exigences en matière de coopération et reproductibilité.

Liste des travaux publiés

Chapitre 2: «aucune publication»
Chapitre 3: Rodrigues, B., & Vergnat, V. (2016). The impact on wages and worked hours of childbirth in France (No. 2016-48). Working Papers du BETA, Université de Strasbourg
Chapitre 4: Rodrigues, B., & Vergnat, V. (2019). The time and the transitions back to work in France after maternity. Rev Econ Household, https://doi.org/10.1007/s11150-019-09442-5
Chapitre 5: Rodrigues, B. (2016). Version control systems to facilitate research collaboration in economics. Computational Economics, 48(3), 547-553.

Bibliographie

Klüsener, S., Neels, K., & Kreyenfeld, M. (2013). Family policies and the Western European fertility divide: Insights from a natural experiment in Belgium. *Population and Development Review, 39*(4), 587-610.

Goldstein, J., Lutz, W., & Testa, M. R. (2003). The emergence of sub-replacement family size ideals in Europe. *Population research and policy review, 22*(5-6), 479-496.

Gayle, G. L., & Miller, R. A. (2006). Life-cycle fertility and human capital accumulation. *Tepper School of Business. Paper, 128.*

Duguet, E., Duvivier, C., Lanfranchi, J., & Narcy, M. (2015). Identification et quantification des sources de l'écart de rémunération entre hommes et femmes au sein de la Fonction publique. *Document de travail-Centre d'études de l'emploi,(89).*

Wilner, L. (2016). Worker-firm matching and the parenthood pay gap: Evidence from linked employer-employee data. *Journal of Population Economics, 29*(4), 991-1023.

Arntz, M., Dlugosz, S., & Wilke, R. A. (2017). The sorting of female careers after first birth: A competing risks analysis of maternity leave duration. *Oxford bulletin of economics and statistics, 79*(5), 689-716.

Kleinbaum, D. G., & Klein, M. (2005). Survival analysis: a self-learning text, Springer Science & Business Media.

Fine, J. P., & Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association, 94*(446), 496-509.

Koenker, R., & Zeileis, A. (2009). On reproducible econometric research. *Journal of Applied Econometrics, 24*(5), 833-847.

Bruno André RODRIGUES COELHO

Essais sur l'offre de travail et la fécondité des femmes

Résumé

Cette thèse porte sur l'offre de travail et les décisions de fécondité des femmes en France. Dans le Chapitre 2, j'estime un modèle hierarchique en utilisant des micro-données Françaises et Allemandes. Le but étant de voir de quelle manière les régions Françaises et Allemandes différent, une fois que l'on contrôle pour des variables individuelles.
Le Chapitre 3 étudie l'impact d'une première, d'une deuxième et d'une troisième naissance sur l'offre de travail et les salaires des femmes et des hommes. La question ici est de voir dans quelle mesure ces naissances ont un impact négatif, voire peut-être même positif dans le cas des hommes.
Le Chapitre 4 étudie la durée qu'une mère passe hors du travail après la naissance de son premier enfant, et comment elle décide de retourner en activité. Un modèle de risques concurrents est utilisé pour cela.
Le chapitre 5 présente un logiciel de gestion de versions. Ce type de logiciel simplifie grandement la production scientifique.

Résumé en anglais

This thesis focuses on the labour supply and fertility decisions of women in France. In Chapter 2, I estimate a hierarchical model using French and German micro-data. The aim is to see how the French and German regions differ, once you control for individual variables.
Chapter 3 examines the impact of first, second and third births on the labour supply and wages of women and men. The question here is to see to what extent these births have a negative impact, and perhaps even a positive one in the case of men.
Chapter 4 examines how long a mother spends out of work after the birth of her first child, and how she decides to return to work. A competing risks model is used for this.
Chapter 5 presents version control software. This type of software greatly simplifies scientific production.

Maternity, Motherhood, Fertility, Labor Economics, Hierarchichal models, Difference in differences, Duration Analysis, Household Data, Microdata, Software