CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Master thesis

2019 **RADKA VLČKOVÁ**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

RADKA VLČKOVÁ

Spatial Analysis of the Motherhood Wage Gap in the Czech Republic

Master thesis

Prague 2019

Author: Radka Vlčková
Supervisor: Mgr. Barbara Pertold-Gebická M.A., Ph.D.
Academic Year: 2018/2019

Declaration of Authorship

I hereby proclaim that I wrote my diploma thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 31, 2019

Signature

Acknowledgment

I would like to express my deep gratitude to Mgr. Barbara Pertold-Gebická M.A., Ph.D., my research supervisor, for her patient guidance, encouragement and useful critiques of this research work. Her willingness to give her time so generously has been very much appreciated. I would also like to thank my friends and classmates for their advice and assistance in finishing my thesis. My grateful thanks are also extended to Roman Heber for his useful and constructive recommendations.

Finally, I wish to thank my parents and my sister for their support and encouragement throughout my study.

Abstract

Using the Czech household microdata, I analyze the regional variations of the motherhood wage gap in the Czech Republic. By this spatial analysis, I aim to detect discrepancies between metropolitan areas and rural localities. Such a study has not yet been carried out in the Czech Republic. Accounting for individual characteristics and working history of employed women, I observe both positive and negative wage gaps but only in several regions. Such findings can be explained by cultural differences, individual self-selection into higher paid jobs or by employers' approach to mothers. For the remaining regions, there is not enough evidence of the motherhood wage gap presence. The hypothesis that in socially excluded localities the wage gap is largest is rejected. Additional examination of the formal childcare provision shows that increase of the kindergartens' occupancy rate significantly decreases salaries of both mothers and childless women. This result is consistent with the previous Czech research.

Keywords

Motherhood wage gap, Spatial analysis, Czech Republic, Kindergarten

JEL classifications

D63, J13, R23

Abstrakt

Pomocí mikrodat českých domácností analyzuji regionální rozdíly mzdové diskriminace matek v České republice. Cílem této prostorové analýzy je odhalit nerovnosti mezi metropolitními oblastmi a venkovskými lokalitami. Taková studie v České republice dosud nebyla provedena. Při zohlednění individuálních charakteristik a pracovní historie zaměstnaných žen pozoruji kladné i záporné mezery ve mezdách, ale pouze v několika regionech. Taková zjištění lze vysvětlit kulturními odlišnostmi, výběrem lépe placených pracovních míst nebo přístupem zaměstnavatelů k matkám. Ve zbývajících regionech nebyl nalezen dostatek důkazů o přítomnosti mzdové diskriminace matek. Hypotéza, že v sociálně vyloučených lokalitách je rozdíl ve mezdách největší, se odmítá. Další zkoumání formálního zajištění péče o děti ukazuje, že zvýšení míry obsazenosti mateřských škol výrazně snižuje platy matek i bezdětných žen. Tento výsledek je v souladu s předchozím českým výzkumem.

Klíčová slova

Mzdová diskriminace matek, Prostorová analýza, Česká republika, Mateřské školy

JEL klasifikace

D63, J13, R23

Master thesis proposal

Author	Radka Vlčková
Supervisor	Mgr. Barbara Pertold-Gebická M.A., PhD.
Proposed topic	Spatial Analysis of the Motherhood Wage Gap in the Czech Republic

Motivation

Gender and motherhood wage gaps have been deeply discussed and examined for decades and as a social problem, the gap has been considered by governments which developed various policies and programmes to fight the salary conditions of women and mothers. In the most recent studies, the researchers find little evidence for unexplained gender wage gap, but they start to explore other gender inequalities in the labour market such as occupational bias in several sectors, persistent glass ceiling, or low representation of women in the governments.

Why is it after all important to do a research on the border of the economics, gender studies, and social policy? I name several most important reasons in the same order as above. Firstly, it has been explained that discriminating women in their wages harms the level of GDP, moreover social contributions from low lifelong income create smaller pensions for the incriminated women. It is considered that the motherhood wage gap is currently the largest barrier in the way towards gender equality in jobs; the full working capacity of the mothers is likely not used meaning some human capital is wasted. When the social aspect is considered, low-income single mothers are the most likely to be hit by poverty where parental leave – labour participation pause – causes another source of the decreased income. It should be accounted also for social exclusion of such mothers either due to their poverty or due to taking care of the children. Many of the researches may serve as policy supporting documents, many of the papers were written as governmental documents anyway.

In my thesis, I would like to focus on the regional differences in the motherhood wage gaps. The spatial view of the gender or motherhood gaps is still uncovered for the Czech Republic, though there are some papers using data from Germany and Poland, so I can gain rough idea about the expected results. I expect the regional analysis to provide insight into wage situation of mothers, as well as employment and social situation of women across the republic. Though the Czech Republic may seem small for regional comparison, such analysis may partially explain the overall social situation that resembles opening scissors.

Hypotheses

1. In rural regions compared to metropolitan areas, motherhood wage gap is larger.
   (a) In the poor, socially excluded regions the motherhood wage gapis the highest.

   These hypotheses reflect the high number of working possibilities in metropolises. High concentration of firms and people makes the labour market rational and therefore mainly the relevant qualities of the workers are taken into account.
2. In the regions with higher concentration and high availability of the preschool care and/or free-time activities for children, the motherhood wage gap is small.
3. Tertiary educated women experience wage discrimination in their fertile years.
   (a) In metropolitan areas the wage discrimination is lowered.

   That is, again, due to efficiency and rationality of the firms that can enjoy large labour market and quick worker replacement as well as due to affordability of various day care.

Methodology

In the thesis, I will work with Czech cross-sectional data set from 2013, this data set was collected mainly for the purposes of the Survey of Income and Living Conditions managed by the Eurostat. However, the European data set is not as rich for the variables as the Czech one, which makes me to target only on the Czech Republic. The greatest advantage of the Czech SILC data set is the regional coding (by "okres") of the questioned person's place of living, which is extremely important for the spatial analysis. Another supporting data sets on macroeconomic indicators, again collected for each Czech region, are downloaded or obtained from Czech statistical office.

This master thesis builds on the approach presented by L.McCall in the paper Spatial Routes to Gender Wage Inequality: Regional Restructuring and Wage Differentials by Gender and Education (1998). The author studies the gender wage gap while accounting for regional variability. The model used in this paper consists of 2 stages; first, micro analysis – the maximum likelihood method is applied on the individual data, the second one, macro level, employ the variation of wages unexplained and regional specifications of the industry and labour market.

I will use this methodology to test the hypotheses described above. Namely, how the (number of) children affect women's wages, type of employment, and how these vary among the regions or between rural and metropolitan areas.

Expected Contribution

The contribution of the proposed master thesis may be grouped into several topics:

- The thesis will contribute to the analysis of the wage and labour force situation from the geographical point of view while interdisciplinary literature will be used so the motherhood wage gap can be understood and explained in deep.

- The differences between metropolitan and rural areas in terms of motherhood wage gaps may serve for policy actions, namely the analysis deliver the information where the examined policies tend to fail. The analysis will help to understand various characteristics of the women and/or mothers and what could increase their income, and well-being. Generally, the goal is to find out how to utilize the human capital of women irrespective of the number of children they have.

- Insight to the Czech labour market will be another important contribution of my thesis. The aspects of interest are number of part time jobs (according to the employment contract). This, among the day care and other family friendly policies, play a part in the inclusion of the mothers with little children to the work life and may decreases the parental leave break, if desired, which is one of the main sources of the decreased wages afterwards.

Core Bibliography

Datta Gupta, N. and N. Smith (2002, November). Children and Career Interruptions: The Family Gap in Denmark. *Economica 69*(276), 609–629.

Felfe, C. (2012). The motherhood wage gap: What about job amenities? *Labour Economics 19*(1), 59–67.

Hirsch, B., M. König, and J. Moeller (2013, 09). Is there a gap in the gap? regional differences in the gender pay gap. *Scottish Journal of Political Economy 60*.

Mandel, H. and M. Semyonov (2005). Family policies, wage structures, and gender gaps: Sources of earnings inequality in 20 countries. *American Sociological Review 70*(6), 949–967.

McCall, L. (1998). Spatial routes to gender wage (in)equality: Regional restructuring and wage differentials by gender and education. *Economic Geography 74*(4), 379–404.

Petersen, T., A. Penner, and G. Hogsnes (2007, June). The Motherhood Wage Penalty: Sorting Versus Differential Pay. Institute for Research on Labor and Employment, Working Paper Series qt9886p84f, Institute of Industrial Relations, UC Berkeley.

Waldfogel, J. (1998). Understanding the "family gap" in pay for women with children. *The Journal of Economic Perspectives 12*(1), 137–156.

Table of contents

List of Figures iii

List of Tables iii

List of abbreviations iv

1 Introduction 1

2 Literature review 4

2.1 Gender wage gap 4
2.2 Motherhood wage gap 5
2.3 Other issues relating to motherhood wage gap . . 11
2.4 Evidence from the Czech Republic 15
2.5 Regional analysis of the motherhood wage gap . . . 17

3 Methodology 19

3.1 Methodology approaches to measure motherhood salary gap 19
3.2 Hypotheses 21
3.3 Empirical framework 21
3.3.1 Methodology on incorporating the child care 29
3.4 Data 30
3.4.1 General data set 30
3.4.2 Auxilary data sets 36
3.5 Descriptive statistics 39

4 Results 44

4.1 Motherhood salary gap - whole Czech Republic . . 44
4.2 Motherhood wage gap estimated on regions 48
4.2.1 Motherhood wage gap in population division 49
4.2.2 Motherhood wage gap in the administrative division 55

4.2.3 Motherhood wage gap in socioeconomic division 65
4.3 Effect of preschool availability on motherhood wage gap 66

5 Conclusion and Discusion **71**

Appendices **75**

Bibliography **79**

List of Figures

1 Occupancy rate of the kindergartens 31
2 Map of the administrative division - LAU 1 32
3 Map of the administrative division - NUTS 3 . . . 32
4 Histogram of the women's age 40
5 Mean gross salary in given age group 41
6 Gross salary in CZK and its confidence intervals given the number of children. 42
7 Mothers annual gross salaries based on age of their youngest children 43

List of Tables

1 List of the basic models and interaction effects models to be compared, using the unrestricted data set 29
2 EU-defined type of households 34
3 Overview of the population based variables 35
4 Overview of the regional based variables 35
5 List of the socially excluded regions based on PCA and its values 36
6 Summary statistics of several variables 37
7 Number of observations of selected dummy variables in the data set 37
8 Results of stability checks for the whole Czech Republic using the full data set. 47
9 Estimates of a mother and the regression statistics of separately run regressions on three urbanization levels 51
10 Results for the merged town levels 53
11 Motherhood wage gap estimates using interaction effects on four town levels. 54

12 Overview of the significant mother dummy in LAU 1 regions 56
13 Summary of signficant estimates of regional analysis LAU 1 division 57
14 Motherhood wage gap estimates regressed on macroeconomic indicators. The region data on LAU 1 level is used. 59
15 Results in NUTS 3 regions where motherhood dummy is found significant 61
16 Summary of signficant estimates of NUTS 3 regional analysis 62
17 Summary of macroeconomic variables collected on NUTS 3 level 63
18 Results of macroeconomic stage analysis on NUTS 3 level 64
19 Overview of motherhood dummy estimates in selected regions defined based on socially excluded localities. SEL refers to socially exluded localities. 67
20 The effect of kindergartens' occupancy rate on women's wages. 70
21 Complete results at different urbanization levels . . 75
22 Overview of base model results using full and restricted data sets and stability checks using the restricted data set - the whole Czech Republic . . . 76
23 Complete results of socially excluded localities and cities regressions. 77
24 Complete results of the occupancy rate and motherhood dummy interactions using the unrestricted data set 78

List of abbreviations

2SLS – two-stage least squares
BP – Breusch-Pagan test
CSO – Czech Statistical Office
CWD – compensating wage differential
CZK – Czech crown
EU – European Union
FGLS – feasible general least squares
HHH – head of household
LAU 1 – Local Administrative Units
MSMT – Ministry of Education, Youth and Sports
MWG – motherhood wage gap
NACE – Nomenclature of Economic Activities
NUTS 3 – Nomenclature of Units for Territorial Statistics
OLS – ordinary least squares
PC – principal component
PCA – principal component analysis
SEL – socially excluded locality
SILC – Statistics on Income and Living Conditions
UK – United Kingdom
US – United States of America

1 Introduction

Motherhood wage gaps have been extensively discussed and examined for decades and as an economic and social issue, the gap has been considered not only by mothers themselves but also by governments. They developed various policies and measures to alleviate the salary conditions of women and mothers. In recent studies, researchers found an evidence of persisting the motherhood wage gap or family wage gap. Determining factors of the lower-earning of those rasing the children are mostly work-related. For example, decreased work experience due to parental leaves, lowered number of worked hours, switch to part-time jobs. Studies also suggest that earnings of mothers cannot converge to earnings of childless women. Beyond these observable factors, greate heterogeneity of women and their unknown characteristics such as a tendency to have large families prevents from obtaining robust empirical results.

In my thesis, I focus on regional differences in motherhood wage gaps. The spatial view of motherhood gaps is still uncovered for the Czech Republic. Actually, many studies cover gender wage gaps in regional levels but it seems that no researcher targets the motherhood of family gaps. Though the Czech Republic may seem small for regional comparison, such analysis may reveal the economic and social position of Czech women and mothers. The second contribution of this thesis is analyzing the effect of childcare availability on earnings of mothers, which is also done as spatial analysis.

In this thesis, I modify the approach of McCall (1998) who examined the gender wage gap to study the motherhood wage gap. I follow her framework which forms two-stage estimation. The first stage deals with the individuals and applies microe-

conomic variables such as human characteristics and their work history. In the second stage, I take regional estimates of the motherhood wage gap and explain the variations by macroeconomic factors.
I define a basic model which includes a set of variables used in the literature. The model is used to estimate the wage gap in the whole country and then in various regions and localities. My framework consists of two approaches: estimating the model for each of the regions separately and estimating the model using regional interaction effects. I operate with the cross-sectional micro-level data set from 2013 which was constructed by the Czech Statistical Office (CSO) and provided by my supervisor. This data set was created by interviewing Czech inhabitants and each region is covered systematically, therefore, a reasonable number of observations is provided. Data sets for macroeconomic stage estimations are downloaded from CSO web pages.

I find little evidence of the motherhood wage gap on the national level as well as on regional levels using simple divisions of the country. Only in several regions, I find both positive and negative wage gaps which are economically significant and results are confirmed using both estimating approaches. In geographically larger objects that are used to divide the country, the GDP and unemployment rate of women are identified to partially explain the regional variance of the motherhood wage gap.
Other analysis reveals that there are no differences in earnings between mothers and childless women in socially excluded localities or in metropolitan areas.
Finally, I observe a negative relationship between the occupancy rate of kindergartens and earnings of mothers. In this case, the negative impact is stronger in socially excluded localities and notably larger in metropolitan areas compared to national aver-

ages. These results confirm the conclusions of Czech researchers that a hypothetical increase in the formal childcare provision would improve the position of mothers in the labour market.

The thesis is structured as follows. After the Introduction, an extensive literature review is presented. The review is systematically written to ensure readers' understandings. The methodological framework is included afterwards. This section includes empirical approaches and detailed description of used methods as well as data description and descriptive statistics. Then, in the fourth section, results are described. The conclusion and discussion complete the thesis.

2 Literature review

In the literature section, I will shortly cover a general topic of the gender wage gap, its development over time, current and past sources of the gap, and comparison of findings stemming from literature. This section will continue with the definition of the motherhood wage gap, suggested economic explanations and mainstream theories. Then I will discuss the main results of the researchers as well as their suggestions and proposed interpretation. I will follow with other relating issues such as job discrimination of women, or occupational segregation. I will conclude this section with studies about regional or spatial analysis of the wage gap.

2.1 Gender wage gap

The gender wage gap has decreased substantially over the years mainly because of an increase in the human capital of women, penetration of women to better-paid jobs, and, in some countries, because of changes in the institutional framework which were set to narrow the gender inequality. In the US, the wage gap between men and women decreased generally for all women during '80s (Waldfogel, 1998). One of the theories also suggests that employer's discrimination towards women has diminished because they are no longer afraid that women would automatically withdraw from the labour market when having children as public support has improved (Avellar and Smock, 2003). However, the rate of lowering the gender wage gap has been slowing down since the '90s which seems paradoxical, as women's level of education is almost equal to the men's. Moreover, female job attainment is more full-time than part-time so work experience is higher and total lifetime employment increased (Datta Gupta and Smith, 2002; Avellar and Smock, 2003; Misra and Murray-Close, 2014). Also, many industrialized countries implemented the equal pay legislation in the '90s, Sweden, Australia, and Norway were reported to have the most egalitarian hourly wages ranging from 90% to 87% at the beginning of '90s (Waldfogel, 1998). Still, there is a gap in earnings that cannot be explained by the human capital and work experience accumulation and researchers as a strong source of the prevailing inequality suggest so-called motherhood penalty and child-care related problems.

2.2 Motherhood wage gap

Grimshaw and Rubery (2015) define the motherhood wage gap in their working paper as follows:

> The motherhood pay gap measures the pay gap between mothers and non-mothers, the latter defined in most econometric studies as women without dependent children. It also measures the pay gap between mothers and fathers. This is different from the gender pay gap, which measures the pay gap between all women and all men in the workforce.

While Nielsen et al. (2004) use shorter definition:

> The family gap is defined as the earnings differential between mothers and childless women *ceteris paribus*.

In econometric terms, it can be stated that the average wages of women with children are lower than for women without a child even after controlling for human capital, labour market experience, and part-time status (Avellar and Smock, 2003). Women with children usually face dual responsibilities: job and family duties. The research suggests that mothers face lower job experience because of the decreased number of hours worked and they have to deal with the trade-off between high paid jobs and jobs with family-friendly amenities. Further, mothers may be less focused on job responsibilities or may be perceived as low productive workers, which is classified as discrimination. In extreme cases, women may leave the labour market completely either voluntarily or involuntarily. But even after controlling for these attributes, the motherhood wage gap is found to range between 3% to 7% depending on countries, data, and methods (Avellar and Smock, 2003; Budig and England, 2001).

There is a difference between the results found in US or UK literature, where the significant negative impact of children on women's wage is reported and the results based on data from Nordic countries which suggest nearly zero effect after controlling of unobserved heterogeneity using panel data estimator. Also, utilizing Danish data, the wage growth is slowed during the period of parental leave but the catch-up effect has been found so the mothers experience higher wage growth compared to non-mothers resulting in convergence in wages (Datta Gupta and Smith, 2002). The UK research indicates the parental leave as a source of the motherhood wage gap, i.e. the mothers who did not interrupt their jobs receive the same remuneration as childless women. Such a result is consistent with human capital theory according to which experience and education determine the wages (Joshi et al., 1997). Similarly, for US-data based literature, the greatest motherhood gap is found for women who leave the labour market completely for a certain time. The lower gap is related to women working part-time and little gap, 4% to 12% depends on the number of children. Such results are reported for mothers employed full-time (Waldfogel, 1997). Finally, the motherhood wage gap differs among some groups of mothers based on race, ethnicity, citizenship, and class positions (Misra and Murray-Close, 2014).

Human capital theory in detail

> According to the basic human capital model, the longer an individual spends in education or training the higher is the expected wage. This logic follows from the assumption at the core of neoclassical economics that marginal productivity equates with an individual's market-determined wage and that human capital serves as a good (observable) proxy of productivity.
>
> - Grimshaw and Rubery (2015)

For mothers, the time in paid work is affected by their duties at home consisting of childcare, housework, and other family responsibilities. It is important to mention that these burdens are imposed on women from traditional reasons and because of social expectations. In the majority of countries, women are allowed to take maternity and/or parental leave with maximum paid duration from 4 months in Spain to 3 years in France or the Czech Republic (Bičáková and Kalíšková, 2015).

The theories in literature propose that there may be a direct effect of children on wages, e.i. lower wages from no other reason than having a child/children in the household, e.i. after controlling for the working experience and unobserved heterogeneity. Waldfogel (1997) observed 4% direct motherhood gap if 1 child is present and 12% when 2 or more children are in the household. These results were obtained after taking into account working part-time which had not been considered yet. In the Spanish study performed by Rodriguez-Planas (2013) the authors concluded that the part-time work explains the motherhood wage gap in larger part than accumulated working experience. Different results were provided by Taniguchi (1999): little or zero evidence of the direct childbearing effect has been found using US panel data. On the other hand, a significant decrease in wage in response to carrier interruption or reduction

in hours worked are reported in the literature. According to to Budig and England (2001), motherhood penalty amounts to 7% of wages and around 2.3% can be attributed to lower job experience. Huge support for the accumulation of human capital theory was found by Lundberg and Rose (2000) using panel data. The authors concluded that married women working continually enjoy the same wage as before the childbirth.
Contrary to these results, proposing that motherhood lowers human capital of women and thus their wages, the research by Cools and Strøm (2014) suggests experience and other human capital measures are not important factors in the parental wage gap. Using Norwegian panel data the authors show that it is rather a lowered effort in work which contributes to lower wages of mothers. And this effect is particularly strong for women working full time in the private sector. The research by Ejrnæs and Kunze (2013) mitigates these results as their model, which controls for nonrandom selection of return to a job, suggest that productivity of mothers after parental leaves is underestimated. Complete opposite results of the motherhood wage gap provided Molina and Montuenga (2009) who found a bonus in salaries for mothers compared to non-mothers. His investigation implies that the superior earnings of mothers are observed in the public sector and for women with fixed salaried work and/or with permanent contracts. The negative motherhood wage gap is detected for temporary workers and single mothers.

When looking at the historic development of the earnings of mothers, Avellar and Smock (2003) concluded that the motherhood wage gap has not decreased. Such result was observed after controlling for various job experience variables included in the US data set covering years 1975-98. On the other hand, the historic decline of the gap is indicated in the evidence form Norway, as Petersen et al. (2007) suggest. These above-mentioned studies imply that various motherhood wage gap is present in abroad. The estimates may vary also due to different methods and data used, so it should not be claimed that the results are inconsistent but rather incomparable.

The first childbirth can shed a light on the effect of timing of the carrier interruption. Two theories are put in place by Taniguchi (1999), first says that delaying the childbirth while building the carrier creates higher opportunity cost of the first child and hence the larger motherhood wage gap. The second results in the opposite outcome, claiming that women can reach better parental support and benefits and that they have already overcome the critical period of carrier building. Additionally, it is believed that early childbearing, establishing family, and developing a carrier lowers employer productivity and thus the wages. Taniguchi (1999) confirmed that women having their first child between 20-27 years experienced wage decrease, while older first-mothers are not penalized. Amuedo-Dorantes and Kimmel (2005) showed that college-educated women delaying the first birth enjoy wage boost when comparing to childless counterparts, with the boost estimated almost 4%. The authors suggest unobserved heterogeneity proposing a relationship that high education and controlled childbearing correlate with "good job". Similarly, wage boost is observed for college-educated women who had a child around 30 compared to early mothers, this gap amounts to 13%. As a drawback of these estimations, it is proposed to study job self-selection of women and mothers.

Institutional background, maternal and parental leave impacts

Institutional support targeting parental wage gaps can be divided into 2 groups. First, job-related measures and leave policies and second, childcare services and support. The work-family policies have been applied from numerous reasons, let's name just several of them: reduction of family poverty, greater gender equality or increased fertility. For a more comprehensive list see (Hegewisch and Gornick, 2011, p.121). Most of the leave policies intend to induce work of women before and after the childbirth as soon as possible. But studies made in developing and developed countries show that the majority of women take the longest leave offered. Regarding childcare, the literature suggests that where a low-quality and/or costly childcare is provided, women tend to decrease their job participation, have lower wages and higher job turnover. If the childcare is of good quality and for example government-subsidized then it has the opposite effect. Overall, the findings propose that the investment info family policies and measures have positive influence on average. Further, the public investment in childcare can affect the employment rates of women with lower intensity than leave policies. Among these general forms of work-family public support, there are less common policies such as the promotion of part-time positions, flexible working arrangements, informal childcare (Hegewisch and Gornick, 2011; Luci-Greulich and Thévenon, 2013).

Schönberg and Ludsteck (2007) used German data from a period of leave legislation changes to find out how prolongation of job protection increases employment and how it affects wages. The authors concluded that the right to return to work later is used by the majority of the interviewed mothers, so they came back to the protected position at the end of the protection period. This measure did not increase long term work participation. Considering only the length of the maternal leave, it was found that every additional month of leave taken decreased the wages by 1%. Ejrnæs and Kunze (2013) reports consistently with these findings and add that the legal prolongation of the

parental leave has not improved the situation of women in the job market.
Similar findings to these of Schönberg and Ludsteck (2007) provided Ruhm (1996) who reported that shorter than common job interruption does not increase wages but longer than common leave decreases wages by 3%. The negative relationship between maternal or parental leave on wages is confirmed also by Ejrnæs and Kunze (2013). In Spain, the data indicate that nine years after the first childbirth, the wages of mothers and childless women are comparable (Rodriguez-Planas, 2013).

Contrary to Schönberg and Ludsteck (2007), Ruhm (1996) used European data and found out that women employment rate is positively affected by job-protected leave duration. But after controlling for all types of paid leaves, the results showed a smaller effect on employment and a larger negative effect on wages.

2.3 Other issues relating to motherhood wage gap

Job self-selection and amenities
This subsection focuses on motherhood and its effect on job choices where job characteristics, not just the wages, play a role. Because as Felfe (2012) argues, once being a mother, women's preferences may change from highly paid jobs to lower-paid jobs with desired family-friendly amenities. Such a phenomenon is called a compensating wage differential (CWD). Using German data, the author's first findings suggest that mothers returning after they leave, if changing an employer, work fewer hours and fewer night shifts, and experience a lower level of stress. When staying at the same employer, only reduced working hours are observed. The author finds little evidence that shift to family-

friendly jobs is the signs of CWD. The analogous conclusion suggest Rodriguez-Planas (2013), the Spanish women seem to exchange the high paid jobs with lower incomes and more family-friendly workplaces.

Nielsen et al. (2004) examine whether women select working in public and private sector when the public sector is considered family-friendly and private sector is claimed to punish mothers by low wages. The author argues that a rational woman who intend to have children choose public sector. The author used Danish data, which is important as Denmark is known for considerably lower wages in public sector complemented by high standards of working conditions, e.g. pension schemes, care-days. Applying endogenous switching models, it is found that having a child increases the probability of working in the public sector and medium-educated mothers have the highest likelihood among all mothers. In the private sector, mothers experience lower wages irrespective of education.
Concluding this short subsection, there is little evidence that women and mothers voluntarily choose less time demanding, lower-paid jobs to take care of their families (Taniguchi, 1999).

Discrimination of mothers at work

From the historical and socio-cultural point of view, women are perceived as mothers with the important task of taking care of children. Because of such believes employers tend to see all mothers as less productive with lower job commitments; it is not unlikely that young childless women are perceived the same way. Several laboratory experiments showed that people have certain role expectations ingrained. For example, female consultants with children were rated as less competent than consultants without children (Grimshaw and Rubery, 2015).
Wage discrimination of mothers is mostly noted in the literature

as a potential source of family wage gap which cannot be tested using data (Avellar and Smock, 2003; Budig and England, 2001; Felfe, 2012; Hegewisch and Gornick, 2011). As a possible source of the discrimination, apart from the cultural expectations, is the unreliability of childcare (Joshi et al., 1997).
More specifically, economists differentiate between discrimination "by taste" and statistical one. "By taste" discrimination occurs when employer, customer, or client find discomfort when employing or working with mothers. Employers discriminate statistically because of the cost of information, it is too costly to measure productivity individually, so averages of mothers performance are used and also serve as a base for wages (Budig and England, 2001).

Marital status and wages

The motherhood gap literature examines also the additional effect of marital status on the wages, apart from studying the effect of (number of) children, leave policy, or type of jobs, as discussed in above subsections. As research indicates, being married itself lowers the earnings of women and increase them for men (Datta Gupta and Smith, 2002). Budig and England (2001) found that with the growing number of children married women obtain lower wages, even after controlling for job characteristics. Contrary to these conclusions, Taniguchi (1999) summarized that no evidence in her research suggests that the wage penalty is related to child and marital status. Waldfogel (1997) found little support that being married or divorced compared to single increases wages after controlling for actual work experience (for both women and men with children). Then in her next study, Waldfogel (1998) concluded that never-married mothers had experienced decreasing wages over a time period, while childless women's wages had risen to men's.

Men's wages in relation to the family wage gap and institutions

In the previous paragraphs, I mention several times the comparison of wages with men, even though this thesis focuses on the motherhood gap. Still, it is important to put this topic into a general context which is provided in several already cited studies.

In the study by Avellar and Smock (2003), the results show that if women keep being the parent responsible for childcare, they can never reach the same level of earnings as men or childless women, holding the institutional settings constant. Similarly, Lundberg and Rose (2000) summarize that first child is related to the mother's wage decrease and father's wage increase. But for families where the mother works without interruption after giving birth, the father enjoys wage increase despite the decrease in hours worked. The authors also suggest that fertility and job choices may be endogenous because married couples earn 9% less on average even before having first child, compared to married couples without children. Regarding the change in job that is family-friendly, research tends to suggest that men are less likely than women to shift their jobs to these with lower wages (Hegewisch and Gornick, 2011).

To conclude the sections 2.2 and 2.3 it is important to say that research about the motherhood wage gap deals with the cost of child-raising. It was discussed who bear the costs, to what extent, and what institutions are in place to decrease the inequality between those who experience the costs. As a final point it should be stated that the benefits of well raised, healthy children are spread widely in society. And, as some economists argue, such young citizens can be viewed as "public good" (Avellar and Smock, 2003). In that manner, Budig and England (2001) hypothesize that costs incurred by an individual (usually a mother taking leave) can be socialized through par example family allowances, childcare, or less costly medical care.

2.4 Evidence from the Czech Republic

The following section is dedicated to literature from the Czech Republic. Because this thesis uses the Czech data, it is crucial to have a direct comparison with the local findings and legal background.
According toPytliková (2015), current median gender wage differences amount to 15%, which is the OECD average. Further analysis of the wage gap shows that median differences between men and women are higher when women are in the typical age of having children. This means that after being 40 years old, the differences decrease. During the period commonly related to having children, women participation in the labour force is significantly lower. But, as Hašková et al. (2015) point out, once a mother has older, school-age children they participate in the job market with high rate and work a standard number of hours per week. During the first three years of a child, mothers take maternity leave usually followed by parental leave. In contrast to many developed and developing countries, women in the Czech Republic do not return to work sooner than when their child reaches two years. That is a reaction to the generous social system, literally non-existent childcare for the youngest children, and limited offer of flexible or part-time jobs. These institutional settings may have a negative effect on the willingness of women to have children (Pytliková, 2015; Bičáková and Kalíšková, 2015).

The number of children in the family has an impact on the parent's wages as well, the median wage difference increases with every child, with the highest increase from 1^{st} (15.25%) to 2^{nd} (25.95%) child measured by wage per hour to account for hours worked. The study also shows that Czech mothers cannot catch up with men of the same age as their children grow up, even though the time spent in work is increased (Pytliková, 2015).

The job-protected period is three years long in the Czech Republic, which is one of the longest leave. It imposes rather high costs to employers and, as Bičáková and Kalíšková (2015) suggest, they often bypass the legislation. That is showed by the fact that there exists 30% unemployment of mothers who want to return to work when their children are 2 years old. However, most of the mothers come back to work when the child is more than 3 years old, which is the time when the protection of the position expires, and 60% of mothers are immediately unemployed. Considering that Czech women have the same level of education as men on average, long parental leave causes a long break in the labour market, the experience stagnates and human capital deteriorates which diminishes the value of such worker. From the government point of view, a mother with low income pays small amounts in social charges and income taxes. In this matter, Kalíšková et al. (2016) studied the cost-benefit analysis of place in state kindergartens with the result that providing an additional place for a child is cheaper for the government and thus beneficial. Also, they suggest a positive impact of working mothers on the Czech pension system, to higher household spending and willingness of women to have (more) children.

Leave policy and care support in the Czech Republic The Czech Republic is an interesting combination of the high employment rate of women and generous social policies targeted to families which are partly based on gender stereotypes (for example long parental leave is founded on the belief that women should take care of children and stay home with them). In 2008 a reform of maternal and parental leave was adopted and Křížková (2008) argues that the new adjustment punishes some class of women that are forced into 4-year-lasting parental leave and students and self-employed faced the new regulation unprepared. Moreover, the reform assumes functional families and parents able to agree, which is sometimes far from reality.

2.5 Regional analysis of the motherhood wage gap

The final part of the literature review is dedicated to regional analyses of the gender wage gap, employment, and fertility. I found no literature covering solely the regional differences in the motherhood wage gap but the effect of children on women's wage compared to childless women can be found also in literature with gender topic.

One of the topic researchers focus on is the urban wage premium, job participation, and the discrepancies in gender. Phimister (2005) discovers a small but significant urban participation premium for women but none for men. Further, the results suggest that urban density increases the gender wage differences since urban wage premium is larger for women. The author explains that women generally tend to follow their boyfriends or husbands to their place of job. Therefore women are less mobile concerning the labour market and so they benefit more from being in cities where labour demand is larger. A different situation is in Poland where the most urbanized regions are associated with the largest gender wage gap even though in these regions women are better educated and work in better-paid (public) sector (Majchrowska and Strawinski, 2016). These results indicate that there are gender wage gap differences across the regions of one country.

Another theory examined in the literature relevant for this thesis is decreased mobility of (married) women and mothers due to ties to their spouses and husbands, as mentioned above. It is argued that families maximize their wealth by moving, and usually, it is a man who can gain more by reallocating to higher-paying jobs. But similar logic is valid to the comparative advantage of women towards unpaid activities (housework) as they sacrifice less, e.i. lower wages and less household income (Nisic, 2017). For example, Hirsch et al. (2013) created own model based on the assumption that women have higher total commuting cost because time is more valuable for them as women have childcare and housework duties. And that, authors, assume, makes women less competitive on the labour market. Using own model, the results suggest that the gender wage gap is 10% larger in rural areas, utilizing German data. In other words, there exists an urban premium for women and similarly summarizes his results even Nisic (2017) who also used German data but different methodology approach.

Kondo (2018) found that agglomeration reduces and postpone fertility of Japan couples, a possible effect of the highly ambitious working environment, expensive rents and childcare services.
However, the most important work for this thesis is McCall (1998). The author examines regional differences in gender wage gap using microdata, macro economical indexes and industry variables. The results are presented by typical industry in a given region. The highest gender wage gap is among college-educated workers in regions with many high-tech oriented companies. More about the methodology will be presented in Section 3.3

3 Methodology

3.1 Methodology approaches to measure motherhood salary gap

To cover the main empirical approaches in the literature, below I provide a summary.
Generally, researchers use both (unbalanced) panel data and cross-sectional data, exceptionally they gather 2 cross-sections from different time periods to observe the trends in case no panel data is available. Using panel data is preferred as the techniques such as fixed effects are generally used and they enable to capture an unobserved individual characteristic. This overcomes the problem with women's propensity to have children and their working preferences which cannot be captured using cross-sections or cohort data sets. This issue is usually difficult to deal with and the results are affected by it.
As I use national individual data from 2013, I further cover the cross-sectional methods to shed more light on it and to support the application of my methods. Generally, 2 techniques with different modifications are used. In first Joshi et al. (1997) decomposed the salaries of mothers and childless women applying Oaxaca methodology suggested in Oaxaca (1973). The author compared the motherhood salary gap within 2 categories of workers: full time and part-time. Moreover, he adjusted the decomposition with a multinomial logit model which enables to simultaneously determine motherhood and employment status.
In the second case of Majchrowska and Strawiński (2015), the researchers used a similar approach using Oaxaca decomposition as Joshi et al. (1997) but the analysis is done for the gender salary gap and only for sectors which are defined as male-female mixed, e.i. the male or female predominance is less than 60%.

The salary itself is measured in numerous forms: daily, hourly, annual (gross) salary. In measuring the motherhood salary gap, Grimshaw and Rubery (2015) state that as a first step the researchers define a mother which is not always straightforward from the economic point of view. Consider a mother whose child lives elsewhere with father; this mother is clearly not involved in childcare. And from the other side, imagine a woman living with her boyfriend and his kids from the previous relationship. Additionally, many data sets lack a variable whether a woman takes care of the children or not but only whether she has any.
When estimating the motherhood salary gap, researchers include variables capturing the human capital in the model. One of the older approaches was to add years of formal education and then age minus years of schooling minus 6 as a variable of work experience. With broader data availability, the researchers could incorporate the exact length of full-time work and part-time work, as Waldfogel (1997) did.
As a next step a marital status is included which is again not a simple task since usually there are many stages of a relationship, e.g. single, cohabiting, married, divorced, separated, widowed. In the literature, the dummies are often used; 1 corresponding to married and 0 referring to single/ widowed/ divorced, or 1 if a woman has ever been married, or 1 if married/ cohabiting and 0 otherwise. Further practice is to use categorical variables and so include more than 2 possibilities of status or to include just an interaction of marital status with children.
Finally, a common practice is to account for the region, job sector, and additional job characteristics such as a number of employees or ratio of women employed in the respective sector (Grimshaw and Rubery, 2015).

3.2 Hypotheses

This thesis concentrates on the spatial analysis of the motherhood wage gap in the Czech Republic. The first and foremost expectation is that the gap is significantly present and therefore I can proceed to assume that there are regional differences. Subsequently, I focus on the relationship of preschools availability and the motherhood wage gap. The hypotheses are summarized as follows:

1. In rural regions compared to metropolitan areas, the motherhood wage gap is larger. (*This hypothesis will be refered to as H1.*)
   (a) In the poor, socially excluded regions the wage difference is the highest. (*This hypothesis will be refered to as H1a.*)
2. In the regions with higher concentration and high availability of the preschool care and/or free-time activities for children, the motherhood wage gap is small. (*This hypothesis will be refered to as H2.*)

3.3 Empirical framework

In this thesis, I use the method suggested and applied by McCall (1998) but modified from men-women salary differences to mothers-childless women. There are two subsequent levels - individual and macro - estimated by OLS or FGLS if heteroscedasticity is present and equations are run with mother being the dummy variable. Having a mother as a dummy is another deviation from the original methodology as McCall (1998) run separate regressions for men and women and then measured the gap by comparing the intercept estimates once they were clear from macroeconomic effects. The analogical approach is to use at the macro level only the estimate of the mother dummy variable. Also, the method is used to do regional analysis which is another dimension of the thesis.

Frameworks used

Equation 1 below shows the individual stage where the dependent variable is a salary[1], and the explanatory variables are on the right side. m_{ij} stands for the mother dummy variable, R_{ij} is a vector of used control variables mainly based on the human capital theory and used indicators of work arrangement, p_{ij} is a probability that a woman ever become a mother and an error term. The probability is a matching variable prohibiting the reverse relationship of salaries to motherhood. It is one of the procedures that prevent endogeneity problem, see below the detailed description of the method. The subscript i indicates individual, j stands for a region. The equation is estimated for the whole undivided Czech Republic, hence no j index in this case, and for several types of regions or areas defined based on administrative division or socio-economic characteristics. In the original work (McCall, 1998), the method was not described in much details and so it is not clear whether the equation is run separately for each region or whether the interaction terms are used. I use both approaches but the interaction is allowed only between a mother dummy variable and regional categorical variable.

$$Y_{ij} = b_{0j} + b_j m_{ij} + R_{ij} r_j + p_{1j} p_j + e_{ij} \tag{1}$$

$$b_j = g_0 + Z_j g_1 + m_j \tag{2}$$

The general issue in estimating a relation in microeconomics is that dependency is not straightforward. In other words, lin-

[1]see description below in Section 3.4

ear regression assumptions say that explanatory variables are exogenous and dependent endogenous. The fact that sometimes the names "exogenous" and "endogenous" are used to refer to the variables is self-explanatory. Nevertheless, I presume that the motherhood dummy variable determines salary. While being a mother may be purely exogenous variable in some relationships, in this case, it is not certain. Without doubt, there are some unobservable human characteristics which affect the women to be a mother and also which may determine the work arrangements and salary amounts, to some extent. When estimating the motherhood wage gap it needs to be corrected for the relationship that these unobserved aspects have on the motherhood dummy to observe unbiased results. One of the corrective measures, when no panel data is available, is to use instrument variable and 2SLS estimator. I follow a procedure of Amuedo-Dorantes and Kimmel (2005) who used the idea of matching.

The matching, in this thesis, is performed by estimating a probability, or a tendency, that a woman becomes a mother. For the estimation, I used logit model estimated using random forest regression. This method is found to be the most accurate which has been tested in the author's analysis where several parametric and non-parametric methods have been examined. The comparison of the methods was based on the results acquired on testing data set. The training data set, selected randomly, was used to estimate the several models. When the random forest was chosen, the whole data set was used to gain the fitted values which fall into interval $(0, 1)$. The values are then applied as a proxy variable in the basic regression as described in Equation 4. The equation estimated with the random forest is expressed below as Equation 3[2].

[2]The status in this regression stands for four factors: married, divorced, single and widowed. The education is divided into five levels: grammar school, high school, extension

$$mother = intercept + \beta_1 status + \beta_2 education + \beta_3 age +$$
$$\beta_4 number\, of\, family\, members\, in\, household + \beta_5 household\, income +$$
$$\beta_6 content\, rate \quad (3)$$

In Equation 2, the estimate of the motherhood wage gap is further repressed on the macroeconomic influences which prevail in the different regions. This stage of the analysis takes the regional estimates of the motherhood wage gap and aims to interpret the unexplained variation between the regions. This analysis depends strongly on the number of observations, e.g. on the number of regions for which the regressions are run separately. After this second stage, estimate g_0 is the national motherhood salary gap. It may be explained in such way: controlling for the usual variables of human capital and for the decision of becoming a mother, the gap is clear of macroeconomic regional differences and is/not significantly positive/negative.

For better comparison between different micro level models, I define a basic one. It contains the key human capital variables and work arrangment variables[3]. Equation 4 represents the regressions that are run separately for each region. When regression is carried out for whole Czech Republic without regional divisions, the subscript j is omitted.

$$salary_{ij} = intercept_{0j} + b_j mother_{ij} + r_{1j} experience_{ij} + r_{2j} experience_{ij}^2 +$$
$$r_{3j} working\, time_{ij} + r_{4j} part\, time_{ij} + r_{5j} high\, school_{ij} + r_{6j} university_{ij} +$$

study, university and doctorate. The content rate is a subjective evaluation of his/her overall satisfaction rate on the scale from 1 to 10.

[3]More details about the dependent and independent variables and their definition or units of measure can be found in Section 3.4 below.

$$r_{7j} marital\, status_{ij} + p_{1j} probablity_{ij} + e_{ij} \tag{4}$$

The model with interaction effects is presented in Equation 5. The interaction effect is performed without the intercept so each region and mother dummy effect can be reported. Also, if I would like to use a reference region, it would be difficult to select one to which the results would be compared to. From the estimation point of view, it is impossible to include solely the interaction effects between two-factor variables and hence the independent regional effect. Having regions as an explanatory variable allows distinguishing between regional effects on wages and regional motherhood wage gaps.

$$salary_i = b_0 mother_i * region_j + b_1 region_j + r_1 experience_i +$$
$$r_2 experience_i^2 + r_3 working\, time_i + r_4 part\, time_i + r_5 high\, school_i +$$
$$r_6 university_i + r_7 marital\, status_i + p_1 probablity_i + e_i \tag{5}$$

The probabilities are not presented in the tables with results as this variable is only auxilary, not explanatory.

For a variable expressing social exclusion, I conduct a principal component analysis (PCA) which is a simple, non-parametric method of extracting relevant information from confusing data sets (Shlens, 2014). I collect variables indicating negative social phenomena and by applying the scaling of the base variables I construct principal components (PC). The first PC is expected to capture enough of the variation. It is then reasonable to use only the first PC in the analysis and in the regressions.
In this stage, every region has its own value of social exclusion. I select a threshold value above which a region is defined as

socially excluded locality (SEL)[4]. Because negative social behaviour is concentrated in cities as well, there are the three largest cities, Prague, Brno, Ostrava, labelled as SEL. As the cities are different regions than the other labelled as SEL, I created extra variable called "city". This PCA is performed using the data on LAU1 levels.

Principal component analysis
PCA is dimensionality-reduction method. The PCs are linear combinations of the original variables, they are uncorrelated, and each PC explains the maximum amount of variation in the original matrix (Moronge, 2016).
GIven the matrix of p variables and n samples, the formal notation of the first PC (Y_1) is determined by the linear combination of the variables $X_i, i = 1, .., p$:

$$Y_1 = a_{11}X_1 + a_{12}X_2 + ... + a_{1p}X_p$$

In matrix:

$$Y_1 = a_1^T X$$

The first PC captures the greatest variance in the data set given that:

$$a_{11}^2 + a_{12}^2 + ... + a_{1p}^2 = 1$$

For all p variables there are p PC created and sum of the variances of the original matrix equals sum of the variances of the PCs. The first PC, having the most of the variance, expresses the most of information stored in the initial data set.
Even though the PCA is simple to use, there are a lot of assumptions which when violated the PCA performs poorly. The PCA can be then adjusted to cases when assumptions are too strict, see Shlens (2014). Finally, PC are not easily interpreted and do not have real meaning as it is artificial variable created as linear combination of the original matrix. (Holland, 2016)

Testing the models and comparing the results

The main goal of regional analysis is to compare the estimates among each other. To be sure that the average estimates are truly significantly different, I calculate the 95% confidence intervals of the estimates. If the intervals of the selected variable estimates overlap, the difference among the regions is not significant.

[4]There is no special procedure for determinign the threshold. It is a coincidence that threshold value equal to 2 resulted in ten regions with the PC value above it.

The model is tested for the stability of results on the national level, e.g. the stability checks are not performed in spatial analyses. The stability checks encompass exchange of several variables with those having similar meaning and which are expected to have similar effect on the gross salaries. I exchanged years of experience and its square with the age variable and its square. It is assumed that the value of the estimates should be equivalent and only the value of the intercept should be lower. As the second check, I exchange the variable of interest - mother dummy - with number of children below 15 years of age living in the household. That is expected to have similar effect as the mother dummy due to the mother dummy definition, see below Section 3.4.
Since the definition of mother is rather broad, even mothers of higher age are taking care of their children are labeled as mothers. In certain age when women is labled as non-mother, it is likely that she has children but does not raise them any more. And in this thesis she would be in the same category as childless women. From this reason, another results stability check is restrict the data set by age from above by 45 years included. This data set is refered to as restricted data set, for the simplicity.

In my thesis, I follow (McCall, 1998), the author used maximum likelihood estimation as "the ordinary least squares (OLS) assumption that random errors are independent and have constant variance is violated". The procedure was not further explained. I check the assumption of homoscedastic errors by Breusch-Pagan test. In case that the homoscedasticity is rejected for one of the regions on 5% level and lower, the errors are adjusted for all the regions. As the variance of the errors is not known, the heteroskedasticity-consistent estimation of the covariance matrix of the coefficient estimates, White's estimator, is used.

Before running the regressions, a correlation test between salary and motherhood dummy is conducted to obtain a brief idea about the results. Even though the test results obtained for the whole Czech Republic is sufficient, it might be interesting to see whether there are regional differences before having the results from the regressions.
The equations estimated for various regional distribution are compared using adjusted R^2. When the regional equations are estimated separately, the number of regions varies and I compare the adjusted R^2s averages over the set of regions.

Breusch-Pagan test and heteroscedastic robust standard errors method
One of the OLS assumptions is the presence of homoscedastic errors, e.i. the constant variance of the errors. This assumption is testable by White or Breusch-Pagan test. I used the latter for its great applicability. "The BP test involves regressing the squares of the OLS residuals on a set of variables in an auxiliary regression":

$$e_i^2 = d_0 + d_1 x_1 + ... + d_k x_k + v$$

The squared residuals is a consistent estimator of the error variance. If the residuals are uncorrelated with the explanatory variables, the above regression would have very small R^2 and thus the errors would be homoscedastic (Wooldridge, Wooldridge). The BP test assumes that under H_0 the errors are homoscedastic.
Heteroscedasticity robust SE are created applying White method when the variance is unknown and thus consistently estimated. This allows for heteroscedasticity robust t-statistics of the explanatory variables. The R^2 or adjusted R^2 are not affected by this procedure.

The basic models and models with interaction effects are, at the end, compared using the Vuong test. This test allows for simple comparison of nested and non-nested models and applies the Kullback–Leibler Information Criterion.

The individual observations in the data set are equipped with the weights which represent a statistical estimate of similar persons in the Czech Republic (the weights are not in terms of frac-

number	model
1	stability checks of the basic model
2	the interaction effects model with three urbanization levels
3	the interaction effects model with four town sizes
4	the interaction effects model with 27 LAU 1 regions
5	the interaction effects model with 14 NUTS 3 regions
6	the interaction effects model with soc. excluded and city localities

Table 1: List of the basic models and interaction effects models to be compared, using the unrestricted data set

tions but in terms of number of similar persons, ranging from 57 to 3141). These weights are always used in the regressions to adjust the observations to the whole population. When the commands in Rstudio do not allow for weights, the data set is adjusted to repeat each observation by its weight, rounded to an integer.

3.3.1 Methodology on incorporating the child care

For testing one of the hypotheses, I use adjusted model as presented by Equation 5. I want to test the idea that the opportunity to use government-run and subsidized kindergartens can increase mother's employment and thus her income. For that purpose, I created a variable called occupancy rate which I added as another explanatory variable into the model. The occupancy rate is defined as the capacity of the facilities in given LAU 1 region[5] divided by the number of children in the age of 3 to 6 (6 included). Usually, the children start school attendance in 6, but some at the age of 7. This is compensated with the children who join the kindergartens just before the 4th birthday. See the occupancy rate in Figure 1.

[5]-local administrative units. The regional and other divisions of the Czech Republic are defined and commented in Section 3.4.

I run the same regression as described in Section ?? but adjusted for the interaction effect between motherhood dummy and the occupancy rate, as depicted in Equation 6. The interaction effects are treated in such a way that no effect is left in the intercept. First, I analyse the effect of kindergarten's occupancy rate for the whole Czech Republic. Further, I apply the defined socially excluded localities and city regions and run regressions on data sets consisting only from these localities and data set consisting from the all other regions - in the Figure named as "common" - neither socially excluded, nor the cities. The same analysis is completed with the restricted data set.

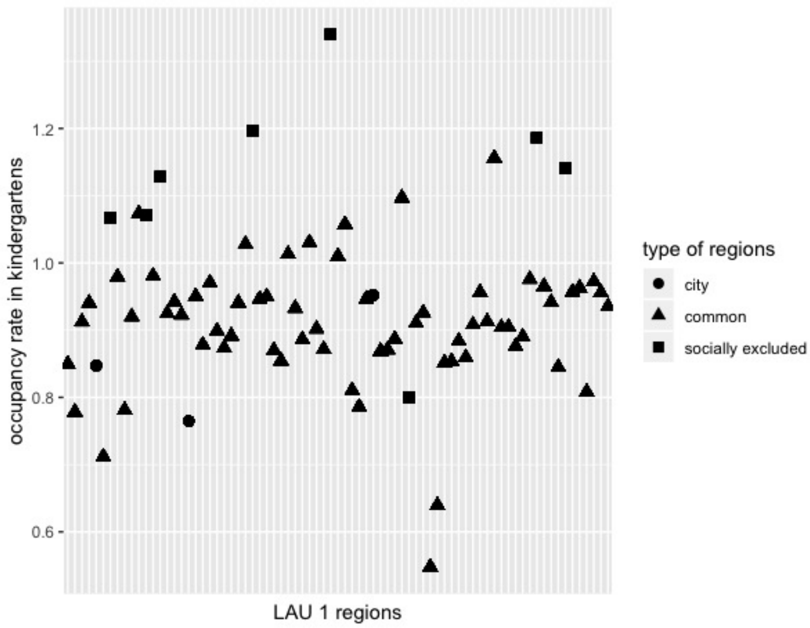
$$salary_i = intercept_{0j} + b_0 mother_i * occupancy_j + r_1 experience_i + r_2 experience_i^2 + r_3 working\,time_i + r_4 part\,time_i + r_5 high\,school_i + r_6 university_i + r_7 marital\,status_i + p_1 probablity_i + e_i \tag{6}$$

3.4 Data

This section starts with a description of the main data set, continues with defining the variables and how data is handled and it is closed with comments on the auxiliary data sets.

3.4.1 General data set

Cross-sectional data from 2013 provided by the Czech Statistical Office is used in this thesis, it is data collected annually for the Eurostat's large European database called Statistics on Income and Living (SILC). The data set is created by interviewing the Czech Republic's inhabitants, not necessarily just



This graph shows the occupancy rate of the kindergartens, in %, in given LAU 1 regions. In color, the excluded regions and 3 major cities are highlighted.

Figure 1: Occupancy rate of the kindergartens

Czechs, in a non-random manner. It contains data about households and individual members of the household above a certain age. Importantly, the data contains information about the regional location of the households - identified with the LAU 1 and NUTS 3 codes. It is equally important to note that every individual is assigned a weight which represents a number of the Czech republic inhabitants similar to that individual in the data set. So basically clusters of persons resembling each other are created. It is a way how to collect data about a sample but receive rough data about the Czech inhabitants. This makes it possible for the researchers to work with the whole population, not just a sample.

The data set used in this thesis consists of 8,276 households with 167 variables observed and 19,105 individuals with 92 variables. However, the main focus is put only on selected variables within a subsample of women, see below for more details. If not explicitly mentioned, this subsample is called the data from now on. The data contains 5,303 observations (women only) of many variables. In the Results - Section 4, it can be seen which variables are used in which analysis. See the summary statistics of several variables in Table 6.

Figure 2: Map of the administrative division - LAU 1

Figure 3: Map of the administrative division - NUTS 3

Defining a mother

Due to the characteristics of the data set, it is not possible to determine if a child in the household is the child of a mother living in the same household, but, as discussed above, it is good enough to assume that once in the same household such mother (it is known when a woman is a mother) takes care of a child and so she bears the burden of the childraising. In the thesis, it is not distinguished between a household with or without a man. The mother is defined based on several variables. Firstly, I include in my data set only women aged strictly above 14 and strictly below 60. Such women need to be either a head of household (HHH), partner or wife of a HHH or a daughter of a HHH. A mother is further defined as a woman in the household of type 6-9 or 0, as it is defined by EU, see Table 2, and she needs to be a parent. And parent is defined as "living in the same household with his or her children". This is a more strict definition than initially intended. But when the restriction of parenthood was relaxed, I encountered many observations of young women labelled as mothers because of their younger sibling. Such data set would contain inaccurate observations of mothers which would be concerning. Still, in several observations of the strictly defined mothers it is possible to observed two and more women in the same household identified as mothers so I selected only the younger one as the caring mother. This change is required as it is not likely that in the same household two women of different generations raise a child. It was inspected case by case. The final data contains 2,326 mothers and 2,977 women without a child she would take care of.

Data manipulation - producing a consistent gross salary This is probably the biggest manipulation with data that might affect the results. The original data set includes information about annual gross salary earn in 2013 (for employees) and months worked (in terms of integers). From this, the monthly salary is derived by interpolating the gross salary of workers to annually-comparable numbers, e.i. interpolating the salaries

EU type of household	Definition
0	other households with children
6	pair of adults with 1 child
7	pair of adults with 2 children
8	pair of adults with 3 or more children
9	1 adult (not necessarily a parent) with at least 1 child

Table 2: EU-defined type of households

reported as annual but consisting only of the salaries earned during a part of the year. Approximatelly 100 obsevations where a woman had positive annual gross salaries and no information about the number of months worked are omitted from the data set.

Defining administrative regions, socially excluded and urban areas

In the regional analyses, I use several types of definitions. Firstly, 2 distributions based on population are already in the data set: three levels of population density, let's say urbanization, and nine levels of a town size based on the number of residents, see Table 3. Several town levels are found to contain low number of observations and so the levels are merged into four levels. Secondly, I use official EU regional distribution: NUTS 3 regions (kraje in Czech) and LAU 1 regions (okresy in Czech). The LAU 1 regions form the NUTS 3, meaning that borders of group LAU 1 regions overlap with NUTS 3 regions. Due to the shortage of observations suitable for the analysis, I grouped the LAU 1 regions into 24 units regarding the regions proximity and unemployment rates from 2013, and Table 4.

Apart from that, three LAU 1 regions corresponding to three largest cities - Praha, Brno, Ostrava - are defined as "cities". For next, socially excluded LAU 1 regions are created from data collected annually by the CSO applying the PCA. See the full list in Table 5.

Population based					
	Town size	observ.		Level of urbanization	observ.
1	up to 999 inhab.	621	1	densly inhab.	861
2	1,000 - 4,999 inhab.	757	2	moderately inhab.	1087
3	5,000 - 49,999 inhab.	946	3	sparsely inhab.	1275
4	50,000 + inhab.	899			

Table 3: Overview of the population based variables

Regional based - administrative	
division	number of regions
LAU 1	76 + Prague
NUTS 3	14

Table 4: Overview of the regional based variables

Human capital and years of work experience

Years of schooling is identified as the highest education level achieved, including vocational training and grammar school dropouts, represented by a categorical variable. As schooling is not expressed in the year exactly, I created only dummy variables: high school graduate and university graduate. Vocational training on the level of high school without graduation and general high schools are part of the former and bachelors, masters, and higher educated are part of the latter. Years of work experience do not need to be calculated from age because the data set includes years worked. Individuals are also distinguished by working part-time or full-time but such information is available only for the that-time situation, e.i. year 2013.

Marital status
The data set separate individuals into four groups: single, married, divorced, and widowed. Based on this, marital status is equal to 1 if a person is married, and 0 otherwise.

Hours worked The data set consists of the estimate of the individual average working hours per week. Sometimes for the employed persons earning non-zero salary, this information was not included and so the zero values are rewritten with 40 hours if the person works full-time and with 20 hours if part-time.

Code	Name of the region	Value of the 1st PC
CZ0806	Ostrava-mesto	5.287
CZ0425	Most	5.115
CZ0426	Teplice	4.341
CZ0803	Karvina	4.271
CZ0427	Usti nad Labem	3.555
CZ010	Hl.m. Praha	3.449
CZ0642	Brno-mesto	3.359
CZ0422	Chomutov	3.092
CZ0323	Plzen-mesto	2.684
CZ0422	Decin	2.592
CZ0801	Bruntal	2.110

Table 5: List of the socially excluded regions based on PCA and its values

3.4.2 Auxilary data sets

Several data sets are downloaded from the website of the Czech Statistical Office. It concerns the data for the macroeconomic situation and demographic data. Data about the schools and kindergartens are taken from the Ministry of Education, Youth and Sports (MSMT). The data set is either directly downloadable from the website of MSMT or it can be filtered manually

	summary statistics					
	min	1st Q	median	mean	3rd Q	max
gross salary >0	1124	13,000	17,909	19,499	23,750	215,358
age	15.0	29.0	39.0	38.8	50.0	59.0
years of experience	0.0	11.0	20.0	19.92	29.0	44.0
hours worked p. week >0	6.0	40.0	40.0	38.66	40.0	70.0

Table 6: Summary statistics of several variables

dummy variable	true	false
mother	1548	1675
part-time	111	3112
university	562	2661
high school	2485	738

Table 7: Number of observations of selected dummy variables in the data set

from the Register of schools and school facilities. This was handled by the web crawl, a unique program created to get the data automatically.

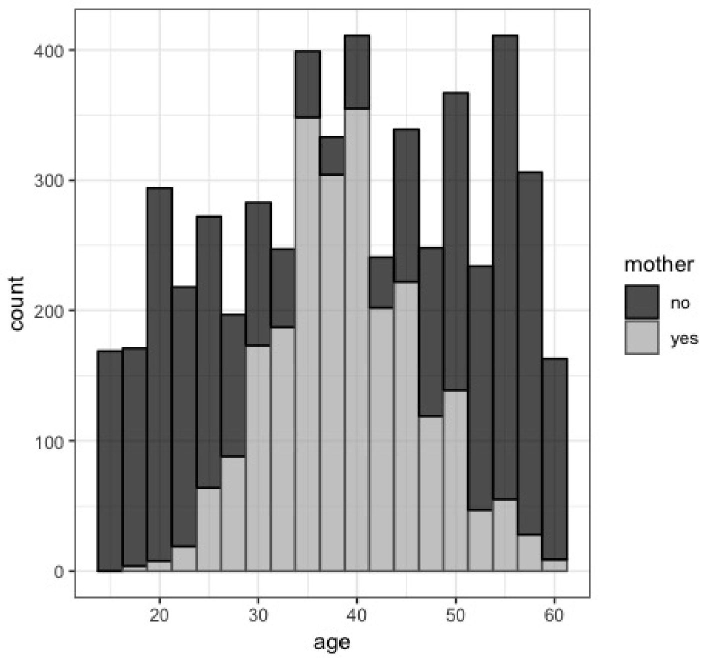
The macroeconomic data encompasses regionally collected data of unemployment, especially women's unemployment, GDP per capita, average housing prices or average monthly gross salaries These data are used in the macroeconomic stage of the econometric analysis as it is illustrated in Section 3.3. CSO does not collect as many macroeconomic variables on regional levels as one would wish so. Therefore, in the analysis using LAU 1 regions, only several of these variables are applied because other variables cannot be obtained on this level.
The demographic data collected on LAU 1 or NUTS 3 levels are used mainly for scaling variables to per capita form. I apply the age structure of the population in the calculation of childcare availability. Another data set encompasses data of people's behaviour: criminality, robberies, house robberies and government housing allowances. These data are used for principal component analysis which I perform to gain single variable determining the socially excluded localities, see again Section 3.3 for methodological details.

The last description of the data set belongs to the childcare data which are extracted from the Register of schools and school facilities - official register of MSMT. I acquire the capacity data on the kindergartens and create average regional occupancy rate which is used in the analysis as an additional explanatory variable, see 3.3.1 for the methodology of this analysis.

3.5 Descriptive statistics

The descriptive statistic is included in order to shed some light on the key variables and to have a better understanding of a mother variable definition used in this thesis. A mother is defined as a woman living in a common household with one or more children to whom she is a mother. The common household makes the definition troublesome as once a child gets older, leaves the household and thus he/she is not cared of, a woman is not considered a mother for the purpose of this analysis. This can be easily observed from Figure 4. The ratio of mothers and childless women increases with rising age of women. But the figure shows that the number of mothers, as defined here, peaks between the age of 35 and 40. This particular figure also shows the distribution of mothers in a women's population. Between 25 and 30 years, half of the women have a child or children, then between 35 and 40 majorities of women raise children. The ratio of mothers and women not raising children starts to decrease around 45 years. That is a typical age of mothers whose children leave the common household. It can be generally understood that women spend nearly half of the productive life taking care of children or having dependent children living home.

I focus on the dependent variable as well. Figure 5 shows the average annual gross salary of employed women divided into age groups (data is restricted to gross salary higher than 0) together with the respective 95% confidence intervals. After the age of 30, childless women are in their top career in terms of salary but then after 40, mothers start to earn more than the childless or women not sharing a household with their children. The biggest difference in salaries between mothers and childless women is observed for the age group 25-35, e.i. when the children are supposed to be little and when women identified as childless

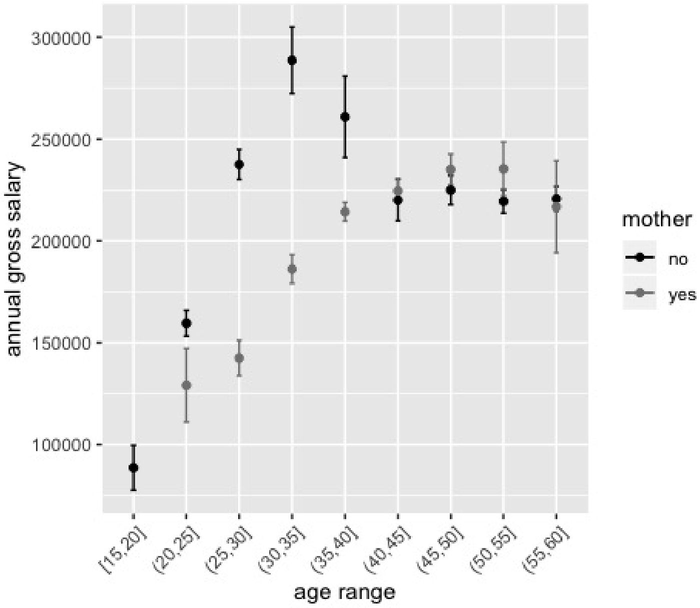


This graph shows the distribution of mothers and non-mothers in respective age groups as represented in the data set. The weights of the observations are not accounted for.

Figure 4: Histogram of the women's age

most probably do not have children. Note here, that salary is not adjusted to full-time or to 12 months working so different number of working hours may explain the biggest differences in the mentioned age groups. By the end of the working life, it seems that motherhood does not matter which is reasonable as older children do not require such care. Moreover, over the years when the children grew up, the mothers had increasingly more time for work and so they could catch up with childless women, in terms of a career.

Effect of the number of children on salary is not primarily examined in this thesis but it is a valuable insight. According to Figure 6, having a single child or two children is associated with a salary decrease by around 10,000 CZK. Information about the



This graph shows annual mean gross salary in CZK for women divided into mothers and non-mothers in the respective age group. The standard errors are shown as whiskers. A reader should mind the definition of a mother as a woman taking care of a child in the same household.

Figure 5: Mean gross salary in given age group

children's age is not recognizable from the Figure 6, for this information see Figure 7. Anyway, with the third and fourth child, we observe that the annual gross salary decreases significantly. The increasing range the confidence intervals also shows that women with three or four children may have very different income. Various family-related (or other) reasons are behind this phenomenon. The family-related justification may be low working hours, working less than the full year 2013, or working part-time. But still, women not raising any child have the highest gross salary.

Figure 7 demonstrates the mean annual gross salary of a mother whose youngest children fall into different age categories. There are five age groups of the children and for each mean gross

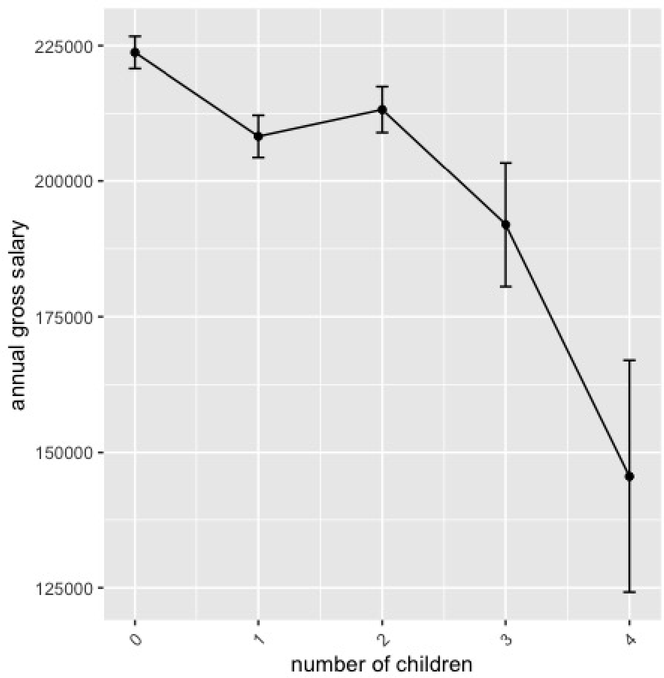
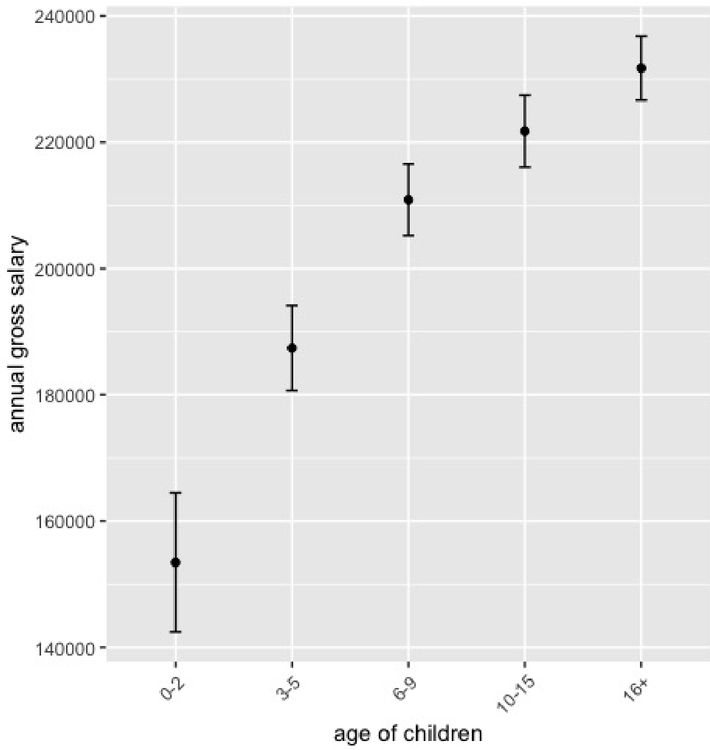


Figure 6: Gross salary in CZK and its confidence intervals given the number of children.

salary, there are the confidence intervals to show the variability in the data. As one could expect, the gross salary rises as the age of the children increases. The largest increase is between the youngest 0-2-year-old children and the 3-5-year-old ones. Majority of women take parental leaves and do not work during the first 3 years of age of the child. The salaries are not adjusted to 12 months employment, so the low amounts may be affected by working only few months of year 2013, e.g. when mothers return to a job after the parental leaves or work before giving birth or have a small part-time job while being on leave. Further, when children attend kindergarten, women are usually part-time or full-time employed, if they are not self-employed, and the annual salaries grow. The average salary for mothers with the oldest children is very similar to the average salary of every employed woman in the data set. From Figure, I can say that the annual gross salaries depending on the age of the youngest child tend to follow the logarithm curve.



This graph shows the distribution of annual gross salary in CZK of the employed women whose youngest child falls into the given age bracket.

Figure 7: Mothers annual gross salaries based on age of their youngest children

4 Results

In this section, results of motherhood wage gap estimations are presented in a structured way to ensure understanding of the reader. It starts with the presentation of results of the regression of geographically large object, whole Czech Republic, and continues with results of the simple breakdown of the Czech Republic based on population variables, town size and density of inhabitants. Administrative-based, NUTS 3 and LAU 1, regional division results follow. Each of the following parts has the same format: summary of estimates of motherhood wage gap - control dummy, detailed description of regional results and information about the fit to data at the end. Analysis using regions defined as NUTS 3 and LAU 1 also continue with the macroeconomic stage where regional differences between estimates are explained.

4.1 Motherhood salary gap - whole Czech Republic

The econometric analysis starts with the simplest model and largest geographic entity, the whole Czech Republic, as reported in Table 8.

There is little evidence supporting the existence of the motherhood wage gap in the Czech Republic as a whole. This result is stable even when basic variables are exchanged with variables having similar meaning and effect on salaries, for example, age with work experience. Second check focusing rather on the homogeneity of the data set is running the regressions and the stability checks on the restricted data set. These analyses return an insignificant estimate of motherhood dummy.

When tested for a correlation between salaries and mother dummy, the test cannot reject the hypothesis that true correlation is zero. This is the first hint that the motherhood salary gap is either very small or non-existent. When salary is tested to be dependent on an increasing number of selected explanatory variables, the mother dummy is not significant in any case. On the other hand, each of the remaining variables is significant on the lowest significance level, i.e. 1%. I started this test with a model consisting of education and mother dummy, continued with adding weekly average working time, then dummy of part-time employment, and finally, I added years of experience and its squared value that allows for a decrease in salary for older women. In this short analysis, I used only job-related characteristics. One of the notable results is that the effect of education on salary is not lowered or otherwise shifted when other variables are put into regressions. Adjusted R^2 increases from 0.138 to 0.216 as variables are added to the model. This analysis shows that even when I let salary to be dependent only on education, age and work time, not experience, there is no motherhood gap. This is an important result as it shows that on average childless women and mothers have *ceteris paribus* similar wages, at least on the national level.

Set of stability checks are performed by exchanging the basic set of explanatory variables with those having a similar meaning. First, I change variable expressing the number of children in a household which are below 15 years of age with mother dummy, see "stability check 1" column of Table 8. The exchanged variable remains insignificant which is expected. The rest of the variables stay truly unaffected by this change in terms of the significance and size of the estimates. It also shows how the definition of a mother is close to just having children below 15 years in a household, the variation in data appears to be similar. Further, I replace years of experience and its squared with age and its squared term, the results are presented in the third column of Table 8. The magnitude of the age variable is increased by 44% compared to experience. This happens because age is numerically shifted in comparison to years of experience, e.g. in 25 years of age, one is expected to have 0-8 years of experience. Again, the estimates of other variables in the model have almost identical value as the basic model, except for the constant. In both modified regressions the explained variance expressed as adjusted R^2 remains at around 0.21. These checks clarify that the results of a basic model using the full data set are reliable and stable.

These stability checks are performed also on data set that restricts women's age to 45 at maximum, complete results are appended as Table 22 on page 76. The regressions, basic one and the stability checks, obtained using this restricted data set all have much less evidence of heteroscedastic errors, measured by BP test. So, data is found to be more homogenous. Still, robust errors are created so the results can be compared with the numbers in Table 8. Motherhood dummy is not significant using this restricted data set. But variables such as marital status, average weekly working time, part-time contract are amplified. Positive estimates are higher and negative lower than when the full data set was used. For example, the estimate of a marital status dummy is about 800CZK lower when restricted data set is used than for base one, which is itself negative. That may imply that for the younger wives in the data set, salaries are more affected by a marriage, negatively. If we suppose that married women are likely to have children then this may mean that married women focus more on family/marriage issues than their careers resulting in a lower salary. But the interpretation should be very careful. The causal effect can operate in the opposite direction. Low earnings of a woman may be an (another) reason for her marriage provided that common marital income makes her better off and a woman benefits. Moreover and more importantly, the reasons for marriage should not be

	Dependent variable:		
	gross salary		
	basic	stability check 1	stability check 2
mother	34.172		−38.710
	(506.657)		(529.208)
number of children <15		229.267	
		(284.028)	
experience	461.114***	460.531***	
	(58.630)	(58.481)	
experiecnce ^{2}	−9.355***	−9.173***	
	(1.485)	(1.374)	
age			663.302***
			(145.590)
age^2			−7.080***
			(1.841)
marital status	−1,022.804**	−1,099.396**	−965.724**
	(471.398)	(454.862)	(478.421)
worktime	307.795***	309.903***	313.743***
	(34.245)	(34.318)	(34.368)
university	14,548.930***	14,604.000***	14,294.760***
	(804.351)	(801.774)	(789.313)
high school	4,219.685***	4,251.356***	4,549.536***
	(403.656)	(404.710)	(402.751)
part-time	−3,751.994***	−3,775.461***	−3,846.881***
	(837.159)	(835.291)	(840.515)
constant	−1,717.839	−1,974.097	−12,409.010***
	(1,444.614)	(1,470.815)	(3,004.148)
Observations	3223	3117	3117
R^2	0.209	0.220	0.214
Adjusted R^2	0.207	0.218	0.212
			*p<0.1; **p<0.05; ***p<0.01

Table 8: Results of stability checks for the whole Czech Republic using the full data set.

reduced only to an economic rationale.
Adjusted R^2 of the basic model and stability checks using restricted data set is 0.18.

From above-described results, I obtained an overview of the general condition of the motherhood wage gap in the Czech Republic. I showed that when salary is explained only by basic human capital variables, education and experience, there is no effect of motherhood dummy on salary, considering average effects. It might be expected that even if there were any motherhood salary gap, the effect would become smaller as other explanatory variables were added to the model. For example, it is reasonable to assume that a mother with very small children earn less as she works less. If working hours were not accounted for, the motherhood dummy might capture some of the variance indirectly connected to this effect. Or, as stated above, married women may be assumed to have children and therefore the effect of motherhood may be captured in the marital status dummy. Nevertheless, all these variables themselves explain variations in salaries and cannot be omitted from the model.
Also, the data set does not provide information about which person in a household is mother or father to which children and therefore I was forced to make a lot of assumptions regarding the definition of a mother. This might contribute to results.
Provided results are stable the standard errors are heteroscedasticity robust. Table 22 on page 76 reports results using restricted data set.

4.2 Motherhood wage gap estimated on regions

Results of motherhood wage gap estimation and other regional findings are presented and commented in sections based on the individual divisions of the Czech Republic. I perform individual, micro-level analysis on different regions independently as well as using interaction effects[6]. Very little evidence supporting the hypothesis of motherhood salary gap existence is found. Despite that, I proceed to analyze regional differences of the motherhood wage gap which is macroeconomic part of the proposed empirical framework. This stage is meant to adjust the estimates of effects such as regional unemployment rate, average housing prices or regional GDP. As these regressions are based on the number of observations equal to the number of regions, only results on NUTS 3 and LAU 1 regions are used in the macroeconomic stage.

4.2.1 Motherhood wage gap in population division

The first two divisions of the Czech Republic have their foundation in the Czech population. Variables used to split the country are based on the predefined variables included in data set which I called urbanization and town size[7]. This division, especially the urbanization, is quite rough but on the other hand, the large number of observations on each level ensure stable and efficient estimates. Only basic model is used in these divisions[8].

For urbanization division, the mother dummy is not significant in regressions run separately on each of the three levels or using interaction effects, even after treatment of heteroscedasticity. All the variables in the model are highly significant except for marital status which is not significant. There are differences in estimates of the same variables across the urbanization levels. Results indicate that in densely populated areas, university-educated women earn by 16,000CZK more than their counter-

[6]As stated in Methodology, see Section 3.

[7]See Table 3 in Section 3.4 for specifications.

[8]See Equations 4 and 5 in Section 3

parts with grammar schools. On the other hand, in the other two urbanization levels, this estimate is around 13,000CZK. Non-overlapping confidence intervals of estimates confirm that this difference is significant. Being a high school graduate has a similar effect on salary at each urbanization level. In the least inhabited areas, the years of experience bring the lowest increase in salary, 225CZK per year. In mid densely populated areas, it is almost 475CZK and in high-density areas, the rise is nearly 650 CZK. However, only the difference between the first two urban levels is significant. Further, the square of experience is the highest there as well so women earn rapidly more each year but from some point in their working life, many experiences (or higher age) leads to lower salary. Complete results can be seen in Table 21 on page 75 in Appendix. Variance explained by the model, measured by adjusted R^2 remains at around 0.21 (mean and median values).

The regression using the interaction effect of mother dummy and urbanization confirms that the mother is not a significant variable in any of the urbanization levels. Individual estimates are quite similar to the basic model results. Salaries are generally higher in the densely inhabited areas; in mid urbanization, salary is lower by 2,200CZK and in the last level, salaries are decreased by another 1,000CZK. Inclusion of the urbanization effect increased adjusted R^2 to 0.23.

The second population-based division variable is town size[9]. In towns of different sizes base on the number of inhabitants, I do not find enough evidence that supports the motherhood wage gap presence. The mother dummy variable turns to be insignificant after controlling for the basic set of variables and after correcting for heteroscedasticity in all regressions, separate

[9]See Table 3 in Section 3.4 for overview of the variable and definition.

	Dependent variable:		
	gross salary		
	low urbanization	mid urbanization	high urbanization
mother	569.18	−891.33	671.09
	(499.35)	(819.75)	(1,228.67)
experience	225.53***	476.91***	650.57***
	(83.90)	(96.70)	(119.75)
experience^2	−4.10**	−9.86***	−13.92***
	(2.06)	(2.62)	(2.83)
university	12,748.12***	13,205.88***	15,952.23***
	(1,064.57)	(974.16)	(1,883.57)
high school	4,321.58***	3,868.89***	4,384.26***
	(641.53)	(582.64)	(950.17)
Observations	861	1275	1087
R^2	0.224	0.221	0.215
Adjusted R^2	0.218	0.215	0.207
			*p<0.1; **p<0.05; ***p<0.01

Table 9: Estimates of a mother and the regression statistics of separately run regressions on three urbanization levels

or with interactions.
Results of individually run regressions vary a bit in terms of magnitude and significance, see Table 10. Some patterns regarding the town size and estimates can be observed. The estimate of work experience variable rises with the town size. Salary increase relating to the additional year of work experience is highest it the largest cities, the estimate is almost double compared to remaining results. Nonetheless, the differences are not statistically significant as their CI overlap in large parts of intervals. For other estimates, the pattern is not so strong. Work time and university education estimates are rather similar in regressions of smaller and mid-sized town. Estimates of these variables are, however, much higher in the large cities. The difference in salary rise for increased work time between cities and other town levels is statistically significant. There is no obvious pattern regarding the town size for the rest of the basic variables such as part-time or high school dummies. Marital status and part-time variables are not even significant in all of the regressions. Average adjusted R^2 is 0.215, the median is 0.217.

The interaction effects between the mother dummy and town sizes bring very similar results as the basic model or as the urbanization interaction effect. The motherhood wage gap is not confirmed at any of the town levels, as depicted in Table 11. The size of town does not influence women's salaries except for the largest town size. In other words, working in a larger town does not mean that salaries are higher if everything else is held equal. Nevertheless, this is not true for the largest towns and cities. In cities with 50,000 and more inhabitants, the average salary is higher by 3,200CZK, compared to small towns. When I use restricted data set[10] for the same analysis, the mother-

[10]The basic data set restricts the age of women. The restricted data set includes only women strictly younger than 46 years.

	Dependent variable:			
	gross salary			
	small towns	mid towns 1	mid towns 2	cities
mother	−273.2	806.4	−848.8	1,046.6
	(735.1)	(723.4)	(869.8)	(1,207.3)
experience	230.8*	375.6***	430.3***	638.2***
	(120.3)	(118.8)	(102.2)	(111.9)
experience^2	−4.0	−7.2**	−9.2***	−13.6***
	(3.0)	(3.0)	(2.8)	(2.7)
marital status	233.1	−1,315.2*	−379.7	−1,674.1
	(659.5)	(793.4)	(665.7)	(1,197.0)
work time	230.2***	261.6***	223.5***	457.0***
	(49.2)	(88.5)	(52.1)	(75.6)
university	13,535.0***	13,696.2***	12,539.8***	15,741.9***
	(1,402.4)	(1,577.9)	(992.8)	(1,743.6)
high school	3,748.8***	4,413.5***	4,141.8***	4,469.8***
	(861.5)	(878.0)	(562.2)	(875.2)
part-time	−5,258.0***	−3,837.1*	−4,367.3***	−2,370.4
	(1,468.2)	(2,017.3)	(1,252.9)	(1,737.7)
Observations	621	757	946	899
R^2	0.232	0.237	0.207	0.222
Adjusted R^2	0.221	0.228	0.199	0.214
				*p<0.1; **p<0.05; ***p<0.01

Table 10: Results for the merged town levels

hood dummy remains insignificant. Women earn by 3,200CZK more in the largest cities than women in the smallest towns. This estimate is not different when older women are omitted from the data set. Women in the restricted data set seem to have a bit higher salary bonus if they are university educated. On the other hand, their salary decreases more when they work part-time, compared to the average salary decrease when all age groups are included. Errors are not found to be heteroscedastic as in the case of the whole data set. The restricted data set is thus more homogenous. The explanatory power of the model with town size interaction effects is similar to the urbanization model, e.i. 0.22 in terms of adjusted R^2.

	Dependent variable:	
	gross salary	
	whole data set	restricted data set
mother*small towns	452.25	−544.52
	(716.81)	(1,050.34)
mother*mid towns 1	−529.91	−588.37
	(652.51)	(957.23)
mother*mid towns 2	−597.78	−1,860.07
	(673.24)	(1,142.88)
mother*cities	847.65	615.89
	(1,016.76)	(1,489.92)
Observations	3,223	1,901
R^2	0.239	0.203
Adjusted R^2	0.235	0.197
		*p<0.1; **p<0.05; ***p<0.01

Table 11: Motherhood wage gap estimates using interaction effects on four town levels.

4.2.2 Motherhood wage gap in the administrative division

The administrative division encompasses the "okres" regions and "kraj" regions. Eurostat refers to the former as LAU 1 regions and the latter as NUTS 3. I had to merge some of the LAU 1 to ensure that each region contains enough observations, resulting in 25 LAU 1 regions. Still, after this grouping, LAU 1 is the finest division that I use in this thesis. For both of the administrative divisions, I examine the basic model using full data set and for interaction effects on NUTS 3 level[11]. I also do these analyses with the restricted data set. Interaction effects are not examined on LAU 1 as results would be highly fragmented and resulting estimates might be biased.

I observe little evidence of the motherhood salary gap present in LAU 1 regions. There are four merged regions for which the estimate of motherhood dummy is significant on 5% and 10% significance level: northern regions (Liberec, Jablonec n. N. and Ceska Lipa), Southmoravian regions (Hodonin, Znojmo and Breclav), regions around Prague (Mlada Boleslav, Praha-vychod and Praha-zapad) and central Bohemian regions (Havlickuv Brod, Kolin, Kutna Hora and Nymburk)[12]. These named regions are merged into four nameless groups, so there are only four significant estimates. Three of the estimates are positive which conflicts with the initial hypothesis that the motherhood wage gap is negative. In these three regions mothers earn by lower thousands of CZK more than women not raising children, even after controlling for the basic set of variables: education, years of experience or average working time etc. The result is, however, not statistically significant as confidence intervals of the estimates overlap in great amount. Also, results are based on 155

[11]For the methodology of interaction effects see Section 3.

[12]See the map of regions presented in Figure 2 in Section 3.4.

observations so one should interpret the estimated relationship with precaution.

	Dependent variable:			
	gross salary			
	northern	Southmorav.	around Prague	central Bohem.
mother	2809.07**	1885.72*	-2891.67*	3803.03*
	(1363.20)	(1043.29)	(1742.59)	(2224.10)
Observations	91	130	123	113
R^2	0.234	0.235	0.165	0.121
Adjusted R^2	0.234	0. 0.243	0.165	0.121

Table 12: Overview of the significant mother dummy in LAU 1 regions

Interaction effects do not support this result, each of the interaction with the mother dummy is insignificant. I suppose the model is quite fragmented and interactions become insignificant with comparison to important variables such as education.

Results of separately run regressions reveal that university education is the most important determinant of a women's salary. Respective dummy variables are found significant in 24 regions out of 25 and the average increase in salary of university graduates is 13,600CZK compared to women with grammar schools only. The highest difference in salaries based on university education is detected in Prague and surrounding regions, graduates earn by 20,000CZK more than women with grammar schools. On the other hand, the smallest salary increase due to university education, 7,000CZK is observed for women in western regions: Karlovy Vary, Chomutov and Sokolov. Low demand for university graduates may cause low pay detected. High school graduates earn by nearly 6,600CZK more on average (5,500 is the median value) than their less-educated counterparts. This is

true for 12 regions, high school education does not significantly increase salary in remaining parts. The highest salary difference is found inregions surrounding Prague, the same as for the highest pay for university-educated. It seems that in these regions educated women are highly demanded but people seem offer their workforce rather in Prague.
Working on part-time contract decreases the salary of women in only two regions, in other regions this variable is insignificant. Years of work experience are found significant in seven regions. On average, an additional year of experience increases salary by 830CZK, the median value is a bit lower, 715CZK. In Hradec Kralove and Pardubice, an additional year in work grants a woman a 1,500CZK increase in monthly salary. In 12 regions out of 25, one working hour rise per week increases the monthly salary by 380CZK, which gives roughly 90CZK per additional hour of work. The highest bonus for working additional hours is paid in Prague and Western regions: Karlovy Vary, Chomutov and Sokolov. Together with the low pay for university graduates in these regions, results suggest that there is a demand for not so educated workers and employers are willing to pay for overtime.

Statistic	Min	Median	Mean	Max
experience	397.2	715.1	832.7	1,607.3
university	7,057	12,097	13,537	20,776
high school	2,437	5,465	6,559	13,633
work time	222.4	353.2	379.7	700.6
part-time	-11,158	-11,046	-11,046	-10,934

Table 13: Summary of signficant estimates of regional analysis LAU 1 division

The variance explained by this basic model ranges from 0.052 to 0.398 in terms of the adjusted R^2, mean and median values are the same, 0.233.

Interaction effects regressions do not indicate any significant mother dummy variable except for one positive estimate significant on 10% level in the northern regions. This result is consistent with individual regressions but the estimate is 3,400CZK. This model with 25 regional interaction effects seems quite fragmented as at least 50 variables concern only the regions.
The other estimates are not considerably different from the average values commented above in Table 13 or from the results of the basic model presented in Table 8. A moderate discrepancy between the national results and results of regional interaction model is in education where the change is in terms of +/- 10%. This effect is balanced by general regional differences in the salaries; in Prague and its surrounding regions, further in Plzen, Hradec Kralove and Pardubice, the salary is higher by approx. 5,000CZK on average. It is natural that in larger cities the salary is higher on average and it seems that there is not enough evidence for other regions to significantly differ in women's salaries. The adjusted R^2 of this regression is 0.27. The results are not detected to be heteroscedastic, so the unadjusted standard errors are used.
Interaction effects analysis is performed on the restricted data set as well. No motherhood dummy is found significant. The data is found homogenous and thus no heteroscedastic errors are observed.

Macroeconomic stage of the model using LAU 1 motherhood wage gap estimates
There is a lack of data collected on the level of LAU 1 regions. Moreover, the estimates themselves are not significant in the majority of regions and thus I use only 2 variables as explanatory: an unemployment rate of women and occupancy rate of the kindergartens[13]. Neither of the variables is significant and

[13]This variable is included to test the hypothesis regarding the childcare availability.

truly do not explain the variation of the estimates.

	Dependent variable:		
	motherhood wage gap estimates		
	model 1	model 2	model 3
occupancy	11.15		−7.61
	(40.60)		(68.15)
unemployment of women		95.38	128.21
		(219.82)	(369.87)
Observations	25	25	25
R^2	0.003	0.01	0.01
Adjusted R^2	−0.04	−0.04	−0.08
	*p<0.1; **p<0.05; ***p<0.01		

Table 14: Motherhood wage gap estimates regressed on macroeconomic indicators. The region data on LAU 1 level is used.

The analysis of administrative based regions proceeds to fourteen NUTS 3 regions. The mother dummy is significant in three of these regions: Liberecky, Jihomoravsky and Vysocina. Estimates show that mothers salaries differ from salaries of childless women in both directions: positively and negatively. In Vysocina and Liberecky region the mothers, on average, earn more by 4,800CZK and 3,200CZK, respectively. In Jihomoravsky region, mothers earn by 2,220CZK less than their childless counterparts. All three estimations are based on a rather low number of observations: 122, 376 and 186 (in Liberecky, Jihomoravsky and Vysocina, respectively). Positive estimates are again in contrast with the initial hypothesis that childbearing and family work affects mothers' work and thus salaries. On the other hand,

See the methodological procedure of how this variable is created in Section 3.4.

results, positive estimates, show that in the two regions, mothers have superior salaries to childless women. Complete results are presented in Table 15.

There are of course differences among NUTS 3 regions in other variables, the summary is provided in Table 16. University education significantly increases salaries in each of the regions. The highest salary increase, over 26,000CZK, have university graduates in Plzensky region where women also obtain the highest reward for an additional year of work experience, nearly 1,200CZK. In this regions, there is also distinctly high cut in salaries if women work part-time, -13,000CZK, which is another reduction on top of the decreased working hours. Differences in salaries for university graduates are, though, statistically significant only with comparison to Karlovarsky and Moravskoslezsky region where the smallest salary increase for university-educated women is identified. This result is consistent with the LAU 1 analysis. Other estimates found for Plzensky region are not significantly different from the rest of the regions. Further, the highest bonus for working additional hours is found in Jihocesky region, which is significantly higher than in the other four regions. Remaining variables do not have such a wide range of estimates among the regions.
Adjusted R^2 ranges between 0.11 and 0.33, with mean and median values equal to 0.22.

Interaction effects results show that in three regions the mother dummy is significant: Vysocina and Jihomoravsky. As in individual regressions, the motherhood estimate in Vysocina is positive and it is negative in Jihomoravsky region. In Vysocina, mothers' salaries are on average by 3,300CZK higher than salaries of their childless counterparts and in Jihomoravsky region, this difference is estimated to be -1,800CZK. These 2 estimates are

	Dependent variable:		
	gross salary		
	Liberecky	Vysocina	Jihomoravsky
mother	3,248.4**	4,872.1**	−2,191.0*
	(1,424.6)	(1,976.9)	(1,177.9)
experience	−235.3	334.8	711.7***
	(216.1)	(208.5)	(222.4)
experience^2	8.7	−4.2	−16.7***
	(5.3)	(4.9)	(5.5)
marital status	−1,980.0	−3,406.3*	−3,126.8***
	(1,697.7)	(1,994.9)	(1,169.0)
work time	211.2	330.9*	313.8***
	(132.3)	(175.1)	(97.8)
university	16,165.6***	12,345.1***	16,885.3***
	(2,583.8)	(2,711.0)	(1,815.6)
high school	7,658.6***	4,072.7***	5,268.0***
	(1,625.4)	(1,482.6)	(1,041.3)
part-time	−7,492.0**	1,432.4	−2,920.1
	(3,630.5)	(3,919.2)	(2,179.9)
Observations	122	376	186
R^2	0.340	0.210	0.298
Adjusted R^2	0.287	0.170	0.280
			*p<0.1; **p<0.05; ***p<0.01

Table 15: Results in NUTS 3 regions where motherhood dummy is found significant

Statistic	Min	Median	Mean	Max
experience	355	547	630	1,172
university	6,829	12,861	13,857	26,431
high school	2,379	5,268	5,200	7,876
work time	218	275	315	667
part-time	-13,261	-7,200	-7,542	-4,514

Table 16: Summary of signficant estimates of NUTS 3 regional analysis

significant different[14]. Adjusted R^2 of this regression is 0.25. Estimates of other explanatory variables are similar to the median values reported in Table 16. There are some differences in the estimates of high school and part-time variables. These are lower in absolute values, e.i. closer to zero. This effect might be captured by regions themselves as in nine of them salaries of women are lower than salaries in Prague.
Interaction effects model run with the restricted data set produces significant estimates of mother dummy in the same regions as the model run with the whole data set. The negative effect of motherhood dummy in Jihomoravsky region is amplified to -4,000CZK and the estimate in Vysocina is lower: 3,000CZK. The model explains 23% of the variance, measured by the adjusted R^2.

Macroeconomic stage of the model This macroeconomic stage suffers, unlike the LAU 1 analysis, from the low number of observations. There are 14 NUTS 3 regions and as macroeconomic variables for Prague differ in terms of orders, I created second data set with 13 observations - omitting Prague. Apart from that, the CSO[15] collects a great amount of macroeconomic variables on NUTS 3 level. Variables I use are summarized in

[14]Confidence intervals of the two significant estimates of the mother dummy do not overlap.
[15]Czech Statistical Office

Table 17.

Statistic	Min	Median	Mean	Max
unemployment rate/women/	5.09	7.96	8.06	11.50
occupancy rate	76.49	91.33	91.95	110.19
GDP	272,500	329,200	361,100	812,500
average of monthly wages	22,010	23,300	24,210	32,730
average flat price[16]	8,918	17,213	18,569	43,968
average house price[9]	1772	2368	2762	8061

Table 17: Summary of macroeconomic variables collected on NUTS 3 level

I tested the correlation between individual variables and for GDP, house price and an unemployment rate of women the test shows significant correlation effects on 10% level. I use these variables in univariate, bivariate and multivariate regressions in all possible combinations. The testing using the Vuong test shows that models are indistinguishable in terms of fit to the data. Comparing adjusted R^2, which from definition punishes adding more variables, reveals that univariate models with women unemployment rate and GDP and bivariate model with both of these variables perform the best. Regressions have adjusted R^2 around 0.16.

Regarding individual estimate values, unemployment rate is positively related to the mother dummy estimate: 1 p.p. increase in unemployment rate increases the motherhood wage gap by 6CZK. GDP is in negative relationship to the mother dummy: 1,000CZK increase in GDP lowers the mother dummy estimate by 8CZK. See the results in Table 18.

All these results are valid only when Prague is included. Once the capital is omitted in the data set, tests find no correlation at all and the same models' adjusted R^2 are sometimes negative. This result is understandable in the case of GDP or house price

[16]per m^2 for flats and per m^3 for houses

as Prague perform like an outlier. But the unemployment rate is not any different in Prague, yet the regression performs very poorly in comparison with the data set containing all fourteen regions.

	Dependent variable:		
	motherhood wage gap estimate		
	model 1	model 2	model 3
unemployment rate of women	6.110*	4.082	
	(3.156)	(3.830)	
GDP		−4.983	−8.140*
		(5.279)	(4.395)
constant	−4,351.061	−968.572	3,430.375*
	(2,589.367)	(4,427.670)	(1,613.917)
Observations	14	14	14
R^2	0.238	0.295	0.222
Adjusted R^2	0.175	0.167	0.158

*p<0.1; **p<0.05; ***p<0.01

Table 18: Results of macroeconomic stage analysis on NUTS 3 level

These macroeconomic stage results should be taken conservatively as there are many concerning issues. First is the low number of observations, second is the low number of significant estimates of motherhood dummy from the microeconomic stage. Last but not least, regressions produce quite uneven results when data sets with and without Prague are used.
The analysis using the occupancy rate of kindergartens is not performed in this macroeconomic stage using the data of NUTS 3 levels. It does not have any meaning as women do not typically

consider putting their child too far from the place of living. Therefore the average occupancy rate of the NUTS 3 region is useless information regarding salaries of women.

4.2.3 Motherhood wage gap in socioeconomic division

The last division of the Czech Republic is formed by two dummy variables defining city localities and socially excluded localities. The identification of socially excluded localities is based on several characteristics such as criminality, robberies or the total amount of housing benefits. The principal component analysis created a single variable for social exclusion. Due to typical aspects of cities, the three biggest cities of the Czech Republic are found socially excluded as well. Needless to say, cities have different characteristics than true localities consisting of smaller towns and therefore I created the variable "cities". Cities are thus part of the excluded localities and, on top of that, they constitute own variable. To analyse the motherhood salary gap, I use interaction effects on the whole and age-restricted data sets: first all excluded regions, then only cities and third model encompasses both effects.

Basic regressions with interaction effects seem to find no significant estimate of the mother dummy, hence no motherhood wage gap. No significant effect is found for the restricted data set as well. Apart from the mother dummy, variable of interest, women in cities have higher salary by almost 5,600CZK on average, compared to salaries in the rest of the Czech Republic. That is the average effect of Prague, Brno and Ostrava on women's salaries. Compared to this finding, women in socially excluded regions do not have lower salary compared to remaining regions as it is found in the model with both dummy effects[17] The excluded variable in the model containing only the SEL dummy is even positive as it captures the effect of cities. Very same results are obtained using restricted variables. Remaining explanatory variables have nearly the same estimates as the base regression results presented in Table 8, the same statement is valid for the restricted data set results and their basic model counterpart. This division based on the social exclusion increases the amount of variance explained by a small amount, adjusted R^2 is 0.24 for the SEL & cities model (compared to 0.22 of the basic model estimated for the whole undivided country). Complete results are appended as Table 23 on page 77.

4.3 Effect of preschool availability on motherhood wage gap

This last analysis aims to complete the analysis on the discussed issue of whether capacity in kindergartens brings additional income to the family. I have data for the capacity of every officially approved kindergarten which can be divided into LAU 1 regions and then I obtained data for age structure of Czech inhabitants from which I am able to calculate regional occupancy rate of kindergartens. In Section 4.2.2 above, the occupancy rate is introduced as the explanatory variable in the macroeconomic stage. It is found insignificant. In the following analysis, the occupancy rate is added as an interaction effect with the mother dummy. The model is run on whole and restricted data sets. To uncover any regional variations, data sets are further restricted to only cities, only SEL and the rest of LAU 1 region - named as "common".

[17] E.g. city and SEL dummies, see the fourth column of Table 19 - SEL & cities.

	Dependent variable:		
	gross salary		
	SEL	cities	SEL & cities
mother:SEL	30.323		−1,068.253
	(1,073.694)		(833.168)
mother:cities		1,096.607	2,136.664
		(1,622.254)	(1,749.058)
SEL	2,139.564***		−350.241
	(592.494)		(654.891)
cities		3,808.945***	4,061.989***
		(802.881)	(979.956)
Observations	3,223	3,223	3,223
R^2	0.228	0.242	0.243
Adjusted R^2	0.225	0.239	0.240
			*p<0.1; **p<0.05; ***p<0.01

Table 19: Overview of motherhood dummy estimates in selected regions defined based on socially excluded localities. SEL refers to socially exluded localities.

Obtained negative estimates of the relationship of the occupancy rate on mothers' earnings confirm the hypothesis. The increased occupancy rate of kindergartens, e.i. lowered availability of free places, negatively affect earnings of mothers. What is surprising, though, that very similar estimates are observed for childless women. Results are displayed in Table 20. And this result prevails even on the age-restricted data set. So it seems that not only does the increased occupancy rate reduce salaries of mothers but also of childless women and the reduction is very similar for both. Moreover, the average effect of 1 p.p. increase in occupancy rate is quite large: -130CZK for childless or -128CZK for mother when whole Czech data set is used. The decrease in wages is larger in cities where the effect is estimated to -380CZK, this difference compared to national average estimate is statistically significant. Lower decrease in wage experience women, mothers and childless, in common regions, e.i. neither socially excluded localities nor cities. Additionally, when the data set is further restricted in age[18], estimates in each defined localities are amplified, hence lower.

The highest adjusted R^2 is observed for socially excluded localities, nearly 0.28, the lowest value is found for common regions. In each regression, the estimate of the occupancy rate significantly negative. Results indicate that the higher the occupancy of kindergartens, the lower salaries of women, see the overview in Table 20. The mildest effect is found in common localities - that are neither characterized as socially excluded nor as cities. In these regions, the increase in the occupancy by 10 p.p. lowers monthly salaries by 1,050CZK. It suggests that the kindergarten and their availability are not key factors affectingsalaries of women or rather mothers. On the other hand, in cities, the increase in the occupancy and thus decrease in the availability

[18]As in every other analysis this concerns restriction in age variable by 45 years of age, from above.

of the pre-schools determines salaries quite distinctly. 10 p.p. increase in occupancy rate reduce salaries by almost 3,700CZK on average. It is reasonable to expect that women with older children are not affected by pre-schools as a mother with small children. Therefore it can be concluded that the availability reduces salaries even more for such women who want to place their children into a kindergarten.

	Dependent variable:							
	gross salary							
	age-unrestricted data set				age-restricted data set			
	Cz. Rep.	common	cities	SEL	Cz. Rep.	common	cities	SEL
mother0:								
occupancy	−130.5***	−106.2***	−378.2***	−158.4***	−144.7***	−116.1***	−375.6***	−185.5***
	(15.1)	(19.8)	(78.9)	(24.0)	(15.1)	(26.4)	(141.5)	(38.1)
mother1:								
occupancy	−128.6***	−104.0***	−355.1***	−155.8***	−146.5***	−121.6***	−341.2***	−174.8***
	(14.8)	(21.0)	(68.2)	(20.7)	(14.8)	(28.4)	(113.5)	(29.3)
Observations	3,223	2,362	485	861	1,901	1,415	271	486
R^2	0.243	0.227	0.236	0.278	0.209	0.187	0.238	0.268
Adjusted R^2	0.241	0.224	0.220	0.269	0.205	0.181	0.208	0.253

*p<0.1; **p<0.05; ***p<0.01

Table 20: The effect of kindergartens' occupancy rate on women's wages.

5 Conclusion and Discusion

In this thesis, I have analyzed the motherhood wage gap and its spatial variation in the Czech Republic using the least-squares method. Results show that the motherhood wage gap is generally not present in the Czech Republic as a whole. However, the wage gap is found to be positive in two out of three NUTS 3 regions, which are geographically medium-large objects. Contrary to the initial hypothesis, the wage is found positive in two of these three regions. The difference in salaries of mothers and childless women is found in four LAU 1 regions, a fines division of the country used in the thesis. Also, in this case, the gap is negative only in a single case and positive in the remaining three regions.

I do not detect the wage gap in the first part of the analysis when I compare localities with different urbanization levels and then different town sizes among each other. The wage gap is not identified even when some of the explanatory variables are excluded from the model which would, as the literature indicates, typically increase the estimated gap.

The lack of evidence of the motherhood wage gap on the national level as well as in variously defined regions is in strong contrast with the findings in the literature. Previous studies indicate that mothers earn less on average than childless women, with variations among race, age or educational groups. Some works using data from Nordic countries suggest that the motherhood wage gap is very low, mainly due to extensive family policies as indicated by authors. Moreover, the observed gap could be attributed to the lowered working effort of the mothers. This is not tested in the thesis but the specification of the model may explain insignificant results. The model used in this thesis incorporates the marital status and three work characteristics: experience, part-time employment contract and an average number of hours worked per week. All of these variables could capture childraising and family-related activities affecting the mothers' salaries. On the other hand, many researchers included the same variables and still, the motherhood wage gap was detected.

The positive motherhood wage gaps found using LAU 1 and NUTS 3 divisions, which are economically significant, are in line with the study by Molina and Montuenga (2009). He found mothers with a permanent contract and/or fixed salary experience salary bonuses. As my data set includes only women with gross salaries, results may be compared. The higher gross salaries of mothers in several regions can be explained by mothers attaining higher-paid jobs. It would mean that Czech women do not select or shift to jobs with lower pay and better work arrangements suitable for mothers. In fact, they may even select jobs with higher pay. Such explanation suggest Petersen et al. (2007) who also found marital premiums. Another explanation involves the employers themselves, they may give salary benefits to mothers simply because they are mothers and they need to take care of their children. Childless women may also want to have children later in their lives and therefore employers can refuse to promote them or offer higher pay. That would suggest negative discrimination of childless women.

The negative motherhood wage gap is identified in Jihomoravsky region only. It is possible that the Moravian culture and stronger religion affiliation creates an environment where mothers may prefer taking care of their families than having great careers. Also, the higher unemployment rate of women may contribute to the worse condition of mothers compared to childless women.

The variation of the motherhood wage gap among the regions is further explained by macroeconomic variables. The unemployment rate of women and GDP on regional levels are found to affect regional variations of the motherhood wage gap. Higher unemployment rate increases the gap and higher GDP it reduces. Both of the effects seem quite natural.

The thesis aims to respond to the stated hypotheses. I assumed that there is a larger motherhood wage gap in rural and socially excluded localities compared to metropolitan areas. To be able to answer the hypotheses H1 and H1a, I performed the principal component analysis to find out which localities are socially excluded. Dummy variables marking the metropolitan areas and excluded localities were added in the model as the interaction effects. The result implies that the motherhood wage gap is not present in neither of the defined localities. In socially excluded localities as well as in cities, mothers seem to earn on average the same salary as childless women. Therefore, the hypothesis H1 and H1a are rejected.

The final analysis of the thesis contributes to the debate of formal childcare and its effect on mothers' wages. I found that the 1 p.p. increase in the occupancy rate of kindergartens lowers the mothers' gross salaries on average by 130CZK in the whole country. That is in line with the Czech studies done by Pytliková (2015) or Kalíšková et al. (2016). Even childless women experience the same effect of childcare provision on wages, which may be a surprising result.
Further research shows that for women and mothers in regions, which are not socially excluded and do not contain the largest cities, the effect of the kindergartens' occupancy is diminished to 105CZK. Compared to the large negative effect in cities, the difference is statistically significant. It implies that in the cities and the socially excluded areas, the mothers' salaries are largely affected by the lack of available places in the kindergartens. That is another evidence supporting the idea that creating additional kindergartens would be beneficial for mothers and also

for the government income, as Kalíšková et al. (2016) suggest.

Regional findings of the motherhood wage gap cannot be compared to another Czech study as the spatial analysis of any kind has not been conducted so far.
The recommendation for the future researchers in this study area using Czech data is to further separate the data according to age groups or type of industry sectors. Currently, such analysis would not be possible on the LAU 1 regions due to the insufficient number of observations, assuming a similar data set is used. Further, results can be improved if the data set included better information about the mothers, such as the number of own children. As the data set used in this thesis is collected for Eurostat applying the EU methodology, it is not likely that the definition of variables would change. But it should be possible to obtain a panel data set which would make the whole analysis more precise as unobserved characteristics could be captured using fixed effects.

Appendices

	Dependent variable:		
	gross salary		
	low urbanization	mid urbanization	high urbanization
mother	569.18	−891.33	671.09
	(499.35)	(819.75)	(1,228.67)
experience	225.53***	476.91***	650.57***
	(83.90)	(96.70)	(119.75)
experience^2	−4.10**	−9.86***	−13.92***
	(2.06)	(2.62)	(2.83)
marital status	−523.51	−379.33	−1,554.41
	(490.70)	(647.34)	(1,245.96)
work time	263.55***	229.74***	439.61***
	(49.85)	(52.22)	(75.88)
university	12,748.12***	13,205.88***	15,952.23***
	(1,064.57)	(974.16)	(1,883.57)
high school	4,321.58***	3,868.89***	4,384.26***
	(641.53)	(582.64)	(950.17)
part-time	−3,795.53***	−4,715.91***	−2,660.50
	(1,280.06)	(1,137.94)	(2,039.68)
Observations	861	1275	1087
R^2	0.224	0.221	0.215
Adjusted R^2	0.218	0.215	0.207

*p<0.1; **p<0.05; ***p<0.01

Table 21: Complete results at different urbanization levels

	Dependent variable:			
	gross salary			
		restricted data		whole data set
	base	stab. check	stab. check	base
mother	−730.49		−857.09	34.17
	(840.94)		(909.60)	(506.66)
number of children <15		−179.29		
		(330.89)		
marital status	−1,019.61	−1,172.39*	−977.69	−1,022.80**
	(779.85)	(676.20)	(775.98)	(471.40)
experience	1,003.34***	966.76***		461.11***
	(146.48)	(150.56)		(58.63)
experience^2	−30.09***	−29.23***		−9.35***
	(5.34)	(5.47)		(1.48)
age			1,995.15***	
			(374.91)	
age^2			−26.58***	
			(5.48)	
work time	281.63***	285.99***	285.92***	307.79***
	(45.53)	(44.59)	(45.84)	(34.25)
university	13,705.60***	13,796.96***	12,562.38***	14,548.93***
	(1,241.28)	(1,200.74)	(1,211.01)	(804.35)
high school	3,740.69***	3,820.07***	3,867.93***	4,219.69***
	(672.19)	(670.69)	(684.00)	(403.66)
part-time	−3,160.43***	−3,091.37***	−3,424.40***	−3,751.99***
	(1,069.52)	(1,064.35)	(1,087.87)	(837.16)
Observations	1,901	1,901	1,901	3,223
R^2	0.183	0.183	0.173	0.220
Adjusted R^2	0.180	0.179	0.169	0.218
				*p<0.1; **p<0.05; ***p<0.01

Table 22: Overview of base model results using full and restricted data sets and stability checks using the restricted data set - the whole Czech Republic

	Dependent variable:		
	gross salary		
	SEL	cities	SEL & cities
mother1: SEL	30.3		−1,068.3
	(1,073.7)		(833.2)
mother1: cities		1,096.6	2,136.7
		(1,622.3)	(1,749.1)
experience	458.5***	440.8***	447.4***
	(55.3)	(54.1)	(54.2)
experience^2	−9.4***	−9.1***	−9.3***
	(1.3)	(1.3)	(1.3)
marital status	−881.2**	−898.4**	−896.5**
	(443.5)	(436.5)	(439.1)
SEL	2,139.6***		−350.2
	(592.5)		(654.9)
cities		3,808.9***	4,062.0***
		(802.9)	(980.0)
work time	301.0***	299.0***	299.0***
	(34.4)	(34.0)	(34.0)
university	14,309.9***	13,724.5***	13,628.7***
	(777.8)	(727.8)	(729.2)
high school	4,170.6***	3,883.3***	3,804.4***
	(401.5)	(402.0)	(405.4)
part-time	−3,811.9***	−3,935.6***	−3,977.8***
	(842.2)	(838.7)	(840.2)
Observations	3,223	3,223	3,223
R^2	0.228	0.242	0.243
Adjusted R^2	0.225	0.239	0.240
			*p<0.1; **p<0.05; ***p<0.01

Table 23: Complete results of socially excluded localities and cities regressions.

	Dependent variable:			
	gross salary			
	Cz. Rep.	common	cities	SEL
mother0: occupancy	−130.51***	−106.18***	−378.23***	−158.40***
	(15.08)	(19.77)	(78.89)	(23.97)
mother1: occupancy	−128.59***	−104.03***	−355.08***	−155.77***
	(14.82)	(20.96)	(68.19)	(20.65)
experience	450.98***	372.17***	607.08***	631.91***
	(57.75)	(63.08)	(179.50)	(122.06)
experience^2	−9.23***	−7.46***	−14.24***	−13.82***
	(1.45)	(1.64)	(4.17)	(2.92)
marital status	−1,082.63**	−434.45	−2,530.16	−2,348.67*
	(455.86)	(411.42)	(2,055.58)	(1,231.84)
work time	300.02***	245.07***	556.89***	416.28***
	(33.91)	(35.87)	(115.61)	(76.11)
university	13,477.34***	12,576.50***	14,899.80***	14,321.26***
	(791.48)	(743.94)	(3,129.53)	(1,965.41)
high school	3,669.54***	4,036.08***	3,054.26	2,532.28**
	(449.34)	(477.19)	(2,014.27)	(1,068.32)
part-time	−3,711.38***	−3,896.71***	−4,387.80	−3,906.83*
	(834.28)	(877.65)	(2,831.91)	(2,090.58)
Observations	3,223	2,362	485	861
R^2	0.243	0.227	0.236	0.278
Adjusted R^2	0.241	0.224	0.220	0.269
				*p<0.1; **p<0.05; ***p<0.01

Table 24: Complete results of the occupancy rate and motherhood dummy interactions using the unrestricted data set

Bibliography

Amuedo-Dorantes, C. and J. Kimmel (2005). The motherhood wage gap for women in the united states: The importance of college and fertility delay. *Review of Economics of the Household 3*(1), 17–48.

Avellar, S. and P. J. Smock (2003). Has the price of motherhood declined over time? a cross-cohort comparison of the motherhood wage penalty. *Journal of Marriage and Family 65*(3), 597–607.

Bičáková, A. and K. Kalíšková (2015). From maternity to unemployment: Women with young children returning to the labour market.

Budig, M. J. and P. England (2001). The wage penalty for motherhood. *American Sociological Review 66*(2), 204–225.

Cools, S. and M. Strøm (2014, 03). Parenthood wage penalties in a double income society. *Review of Economics of the Household*.

Datta Gupta, N. and N. Smith (2002, November). Children and Career Interruptions: The Family Gap in Denmark. *Economica 69*(276), 609–629.

Ejrnæs, M. and A. Kunze (2013). Work and wage dynamics around childbirth. *The Scandinavian Journal of Economics 115*(3), 856–877.

Felfe, C. (2012). The motherhood wage gap: What about job amenities? *Labour Economics 19*(1), 59–67.

Grimshaw, D. and J. Rubery (2015). The motherhood pay gap: a review of the issues, theory and international evidence. Ilo working papers, International Labour Organization.

Hašková, H., A. Křížková, and R. Dudová (2015). The economic costs of motherhood: what the responsibility for care of a child(ren) means for women in terms of employment and economic status.

Hegewisch, A. and J. C. Gornick (2011). The impact of work-family policies on women's employment: a review of research from oecd countries. *Community, Work & Family 14*(2), 119–138.

Hirsch, B., M. König, and J. Moeller (2013, 09). Is there a gap in the gap? regional differences in the gender pay gap. *Scottish Journal of Political Economy 60*.

Holland, S. M. (2016, 07). Introduction/doing prcomp() more efficiently.

Joshi, H., J. Waldfogel, and P. Paci (1997). The wages of motherhood : better or worse. Bibliography: p. 25-28.

Kalíšková, K., D. Münich, and F. Pertold (2016). Veřejná podpora míst ve školkách se vyplatí: analýza výnosů a nákladů.

Kondo, K. (2018). Does agglomeration discourage fertility? evidence from the japanese general social survey 2000–2010. *Journal of Economic Geography*, lbx048.

Křížková, A. (2008). *Práce a péče : proměny "rodičovské" v České republice a kontext rodinné politiky Evropské unie.*

Luci-Greulich, A. and O. Thévenon (2013, Nov). The impact of family policies on fertility trends in developed countries. *European Journal of Population / Revue européenne de Démographie 29*(4), 387–416.

Lundberg, S. and E. Rose (2000). Parenthood and the earnings of married men and women. *Labour Economics 7*(6), 689 – 710.

Majchrowska, A. and P. Strawinski (2016). Regional differences in gender wage gaps in poland: New estimates based on harmonized data for wages. *Central European Journal of Economic Modelling and Econometrics 8*(2), 115–141.

Majchrowska, A. and P. Strawiński (2015). Regional differences in gender wage gaps in Poland. Technical report.

McCall, L. (1998). Spatial routes to gender wage (in)equality: Regional restructuring and wage differentials by gender and education. *Economic Geography 74*(4), 379–404.

Misra, J. and M. Murray-Close (2014). The gender wage gap in the united states and cross nationally. *Sociology Compass 8*(11), 1281–1295.

Molina, J. A. and V. M. Montuenga (2009, Sep). The motherhood wage penalty in spain. *Journal of Family and Economic Issues 30*(3), 237–251.

Moronge, J. M. (2016, 07). Principal components analysis (pca).

Nielsen, H. S., M. Simonsen, and M. Verner (2004). Does the gap in family-friendly policies drive the family gap?*. *The Scandinavian Journal of Economics 106*(4), 721–744.

Nisic, N. (2017). Smaller differences in bigger cities? assessing the regional dimension of the gender wage gap. *European Sociological Review 33*(2), 292–3044.

Oaxaca, R. (1973). Male-female wage differentials in urban labor markets. *International Economic Review 14*(3), 693–709.

Petersen, T., A. Penner, and G. Hogsnes (2007). The Motherhood Wage Penalty: Sorting Versus Differential Pay. Institute for Research on Labor and Employment, Working Paper Series qt9886p84f, Institute of Industrial Relations, UC Berkeley.

Phimister, E. (2005). Urban effects on participation and wages: Are there gender differences? *Journal of Urban Economics 58*(3), 513 – 536.

Pytliková, M. (2015). Wage differences related to motherhood and children in the family.

Rodriguez-Planas, N. (2013, 03). The motherhood earnings dip: Evidence from administrative data. *The Journal of Human Resources 48*, 169–197.

Ruhm, C. J. (1996, July). The economic consequences of parental leave mandates: Lessons from europe. Working Paper 5688, National Bureau of Economic Research.

Schönberg, U. and J. Ludsteck (2007, March). Maternity Leave Legislation, Female Labor Supply, and the Family Wage Gap. IZA Discussion Papers 2699, Institute for the Study of Labor (IZA).

Shlens, J. (2014, 04). A tutorial on principal component analysis. *Educational 51*.

Taniguchi, H. (1999). The timing of childbearing and women's wages. *Journal of Marriage and Family 61*(4), 1008–1019.

Waldfogel, J. (1997). The effect of children on women's wages. *American Sociological Review 62*(2), 209–217.

Waldfogel, J. (1998). Understanding the "family gap" in pay for women with children. *The Journal of Economic Perspectives 12*(1), 137–156.

Wooldridge, J. M. Chapter 8.