European University Institute

DEPARTMENT OF ECONOMICS

Essays on Family and Urban Economics

Ana Moreno Maldonado

Thesis submitted for assessment with a view to obtaining the degree of Doctor of Economics of the European University Institute

Florence, 10 September 2020

European University Institute
Department of Economics

Essays on Family and Urban Economics

Ana Moreno Maldonado

Thesis submitted for assessment with a view to obtaining the degree of
Doctor of Economics of the European University Institute

Examining Board

Prof. Árpád Ábrahám, EUI and University of Bristol, Supervisor
Prof. Juan J. Dolado, University Carlos III Madrid
Prof. Lídia Farré, University of Barcelona
Prof. Nezih Guner, CEMFI

European University Institute

DEPARTMENT OF ECONOMICS

Researcher declaration to accompany the submission of written work

I Ana Moreno Maldonado certify that I am the author of the work "Essays on Family and Urban Economics" I have presented for examination for the PhD thesis at the European University Institute. I also certify that this is solely my own original work, other than where I have clearly indicated, in this declaration and in the thesis, that it is the work of others.

I warrant that I have obtained all the permissions required for using any material from other copyrighted publications.

I certify that this work complies with the *Code of Ethics in Academic Research* issued by the European University Institute (IUE 332/2/10 (CA 297).

I confirm that chapter 2 was jointly co-authored with Ms Clara Santamaria and I contributed 50% of the work.

I confirm that chapter 3 was jointly co-authored with Mr Mathijs Janssen and I contributed 50% of the work.

Signature and Date:
Ana Moreno Maldonado

9th of July 2020

Acknowledgments

I want to start by thanking my advisors, this dissertation would not have been possible without your immense dedication. You found the exact balance between freedom and guidance, and were extremely supportive every time that freedom resulted in me panicking. Thanks to Arpad Abraham for your continuous encouragement, for being a true role model and, specially, for all the good jokes. Thanks to Juan J. Dolado for believing in me from the very first day, for reading my dissertation a million times, and for being family, I would never have become an economist without your help.

I am also incredibly grateful to my wonderful co-authors, who made me enjoy research a lot more. I want to thank Mathijs for teaching me how to think about theory and for being a great friend, playing boardgames with you was a lot of fun, and to Clara, for being so helpful and kind, and for sharing so many good moments together.

Along my journey, I have always felt very lucky for the extremely friendly environment at the EUI's Economics Department, for the Departmental BBQs, and for all the great Christmas parties. Big thanks to the kind and approachable staff at Villa la Fonte and Villa San Paolo who made these and many other things possible: Sarah, Lucia, Rosella, Maurizio, Lori, Sonia, and Antonella, you were great! And thanks to Thomas for his dedication and his competence. I also want to thank the attendants of the Macro Working Group at the EUI, my research greatly benefited from their comments and from Ramon Marimon's right-on-the-spot questions.

These years in Florence I was surrounded by amazing friends. Special mention to my PhD cohort, we fought this battle together and you made it as fun as it can get. Thanks to Carolina, my blond twin and best paella maker ever. It was great to live with you and do so many fun things together, I miss having to change clothes in the morning to avoid wearing the exact same outfit to class.

To Matteo, best work and party mate, thank you for all the walks in La Fonte and for the infinite hours together in the office. To Ale, for providing me with chaos and joy in equal shares and for staying in my life well beyond our time in Florence. I am also thankful to my friends Carlos, David, Gonzalo, Iacopo, and Martina, for reviving our teens together dancing all around Florence. I cannot thank enough my *compagni* from boxing at the CPA, who accepted me as one of their own from the very beginning. My heart will always remain a bit Florentine thanks to you.

Part of this dissertation was written during my stay at Universitat Pompeu Fabra. I am indebted to my mentor there, Jaume Ventura, your invaluable advice gave rise to the first chapter of this dissertation. I am also thankful to Giacomo Ponzetto and David Nagy, your course on Economic Geography has proved crucial for this dissertation, as well as your advice. Thanks to Libertad Gonzalez, who always made time for my presentations at UPF, to Marta Santamaria, for becoming a great friend in such a short time and for your enthusiastic view on Economics, and to Francisco Queiros, for being a great office mate and friend.

Towards the end of this dissertation I was given the opportunity to visit Universidad Carlos III de Madrid. I am extremely grateful for the warm welcome by faculty and PhD students, it really felt home. Thanks to Andres Erosa and Luisa Fuster, their advice greatly improved my dissertation. I am also thankful to Alan Crawford, Antonia Diaz, Belen Jerez, Matthias Kredler, and Evi Pappa, for their insightful comments about my research. Moreover, I want to thank my friends Tati and Alexandro, working on Saturdays was much nicer with you.

This thesis benefited from the advice and inspiration of many great economists. I am specially grateful to Joan Monras, working for you was a terrific experience. Thank you for always putting my PhD first, for making it a great opportunity to learn, and for being so kind. I want to thank as well Nezih Guner and Lidia Farre for their insightful comments as members of my Thesis Committee, and my friend Milena for always being available for an emergency Skype and for all the great reunions in Madrid.

All these years I felt extremely lucky for having the best partner in life, David. I could never thank you enough for coming to Florence and sharing the PhD experience with me, it made it one million times better. Thank you for always keeping our feet on the ground and for always being there for me.

Lastly, this dissertation is dedicated to my family. I am immensely thankful for having the funniest and most loving family. To my parents, for always believing in me, for working so hard to provide me with the best opportunities, and for making me feel capable of anything. To my sisters, for all the laughters and shared confidences, I could not imagine my life without you. I also want to thank my grandma for her joy and her generosity. We spent the last months of my PhD confined together and she not only took care of everything, enabling me to work all the hours in the day, but also made me laugh constantly.

Abstract

This dissertation analyses how the geographical sorting of individuals and households affects labour markets as well as gender and spatial inequality.

In the first chapter, I show that labour force participation increases with city size for all demographic groups except for women with children, for whom it decreases, a phenomenon that I label Big City Child Penalty (BCCP). Both by means of empirical evidence and a quantitative spatial model of households, I show that the BCCP can be explained by commuting times, wages, and child-care price differentials between small and big cities as well as for unobserved heterogeneity in preferences for a stay-home parent.

The second chapter of this dissertation highlights the role of delayed childbearing as an important driver of gentrification. While downtowns provide shorter commuting times and more consumption amenities, limited housing space and schools' worse quality reduce the value of this location choice when children are born. We exploit exogenous variation in the cost of postponing childbearing to obtain causal estimates of the impact of delayed maternity on gentrification. We find that enhanced access to assisted reproductive technologies in the state increases income downtown by 5.4% relative to the suburbs.

The third chapter studies the relationship between trade and migration. Coinciding with a period of increasing trade integration, the educational composition of migrants within the European Union changed towards high-skilled workers. We build a two-country, two-sector general equilibrium model in which countries only differ in the productivity of high-tech workers. While price equalization, induced by trade integration, equalizes the real wages of non-educated workers, differences in the real wages of educated workers remain, since the latter are more productive in the most advanced country. As a consequence, factor mobility is needed to exhaust differences in real wages, leading to high-skilled emigration towards the most advanced country.

Contents

Chapter 1

Mums and the City: Female Labour Supply and City Size

1.1 Introduction

The rise in female labour force participation (LFP) has been one of the most salient changes experienced by the labour markets in developed countries since the second half of the 20th century. Moreover, as regards the US, the marked upward trend observed for female participation has been mainly driven by the trend among married women, which increased by almost 35 percentage points (p.p.) between 1955 and 1995, despite stagnating later (see Killingsworth and Heckman (1986)). Thus, the incorporation of this group of women to paid work has drawn plenty of attention in the literature (see Goldin (2014) for an overview). However, somewhat surprisingly, spatial differences in female LFP have remained understudied. Using data for the US, I contribute to fill this gap by uncovering a novel fact that hereinafter is labeled as the *Big City Child Penalty* (hereafter BCCP in short): while the LFP of married women with no children and of all men increases with city size, the LFP of married women with children is almost 5 p.p. lower in big cities than in small cities. This difference is large enough to deserve further attention in the literature on female outcomes in the labour market. Indeed, if the LFP rate of women with children living in big cities was the same than in small cities, the overall female LFP rate in the US would be 3.4% higher.[1]

[1]Regarding the intensive margin, differences are smaller but they operate in a similar direction. For ease of exposition I focus on the extensive margin and ignore the intensive margin. Yet, I include city size differences in the probability of working full time in Table

Despite the large convergence in labour supply between men and women, recent decades have witnessed female LFP rates reaching a plateau. In this context, unveiling the main drivers behind geographical differences in female labour supply can also help identify the remaining hurdles hindering gender equality at the aggregate level. In other words, we can use spatial variation on characteristics that relate to the decision to participate, such as average commuting times, to evaluate their contribution to the gender gap in labour supply. One caveat to this approach is that individuals choose residence based on their preferences. If women in big cities have different attitudes towards market work than women in small cities, the empirical analysis would likely overestimate the contribution of these factors at curtailing female LFP. For this reason, I combine regression analysis with a quantitative spatial model of households that allows to account for the endogenous location choices. More concretely, geographical differences in LFP can be rationalized in two main ways. First, one could argue that some characteristics that are intrinsic to living in big cities (such as long commuting times) make it hard to reconcile family and work life in these locations. Second, women living in big cities might be different from those in small cities in some dimensions which cannot be observed in the data, e.g. couples with higher preferences for a stay-home parent may be more likely to locate in big cities. I will refer to the latter as geographical sorting on unobservables.

In the empirical section of this paper I outline the key differences between small and big cities driving the observed geographical variation in LFP. In particular, I document that, relative to small cities, larger metropolitan areas are characterized by: (i) longer commuting times, (ii) higher real wages, (iii) higher concentration of occupations with a high return on working long hours (such as finance or managerial positions), and (iv) more expensive childcare. It should be noticed that all these features favour specialization within the household, leading one partner to take up market work while the other partner undertakes childcare. On the one hand, longer working hours and longer commuting times constrain couples' available time to look after their children or to enjoy leisure, making it difficult for both household members to participate in the labour market. On the other hand, a higher partner's wage relaxes the budget constraint, facilitating that one of the parents opts out of the labour force. Likewise, more expensive childcare increases incentives for a parent to stay out of the labour force and thus avoid the expense. Given that

A.2 in Appendix A.

traditionally women are more likely to bear a higher share of home production (Bertrand et al. 2010; Erosa et al. 2017), intra-household specialization mostly reduces female labour supply. Indeed, exploiting cross city variation, I find that the LFP of women with children is lower in cities with longer commuting times, larger shares of total employment in long hours occupations, and more expensive childcare.

However, it should be noticed that the outlined city characteristics could also induce the sorting of women with low labour attachment into big cities, something that cannot be accounted for with cross-sectional data. Therefore, I build a quantitative spatial model of households in which endogenous location and households' idiosyncratic preferences for LFP that facilitates placing geographical sorting on unobservables at the center of the BCCP. Moreover, the model allows me to account for taxation as an additional determinant of female LFP, as well as to improve our understanding on how all these different features operate in the modelling setup. Lastly, I use the model to evaluate the consequences of two policy changes aimed at raising female LFP.

In the proposed model, I consider two representative cities, small and big, that differ in occupation-specific productivity, commuting times, and housing costs. In order to capture the fact that occupations with a high return on working long hours are concentrated in big cities, I introduce two occupations that differ in the rate at which hours worked are converted into efficient units of labour: (i) a *long-hours* occupation in which this rate of conversion is a convex function of working hours, and (ii) a more flexible occupation with a linear rate of conversion.[2] I assume that the big city has a comparative advantage in the long-hours occupation so that it clusters there. In addition, I set commuting times in each city equal to the city average commuting time observed in the data, as a result of which individuals working in the big city bear a higher commuting cost than those in small cities. Finally, in line with the empirical evidence, I assume that housing rents are more substantial in the big city. Given that the price of childcare is modelled as a function of the wage and the price of housing in the city, higher housing costs in the big city also entail a higher childcare cost in this location.

The economy is populated by couples that are heterogeneous in productivity, parenthood, and preferences. In particular, households have idiosyncratic

[2]Notice that this means that, wages being equal, optimal hours in the convex occupation are higher, as otherwise the individual would be better off choosing the linear occupation.

preferences for cities, women's LFP, and occupations. They decide: a) which city to live in; b) wife's LFP; c) each partner's occupation; d) how many hours they work; and e) how to take care of the children, a task which is modelled as a time cost.[3] In addition, I assume that households earnings are taxed, given that female labour supply has been shown to respond strongly to changes on tax rates (Guner et al. (2012); Alesina et al. (2011)).[4]

I calibrate the model parameters to match the most salient features of labour markets in big and small cities in the US. In line with what is observed in the data, the calibrated model predicts a higher LFP rate of women without children and a lower LFP rate of women with children in the big city as compared to the small city. For both types of couples, the decision to participate depends on how they value time use at home relative to the additional consumption they would enjoy if the wife participated. First, couples' time devoted to family care is more constrained in the big city due to longer working hours and longer commuting times, which discourages female LFP, the more so among couples with children. Second, two opposite forces emerge from better pay in this location: while a higher wage encourages LFP because it leads to a larger increase in consumption than in the small city, the rise in husband's income reduces the marginal value of household's consumption, thus discouraging LFP. Third, the relatively more expensive childcare in the big city also reduces the net gain of participation for women with children.

To assess the relative contribution of each channel, I shut them down one by one and compute the corresponding equilibrium outcomes. These counterfactual simulation exercises reveal that the most relevant channels in explaining BCCP are commuting times, taxation, and differences in the price of childcare. Yet, on their own, they are only able to explain less than a third of the observed size of the BCCP, while more than two thirds are explained by households' sorting on unobservables. I obtain this measure by computing the equilibrium under the assumption that the underlying distribution of preferences is the same in both cities and suppressing household's choice on which city to live in. Measuring the degree of geographical sorting is particularly important from

[3]That is, I assume that the husband cannot opt out of the labour market.

[4]In particular, Guner et al. (2012) show that the US tax system imposes a large marginal tax rate on the first dollar earned by a partner considering to enter the labour force, thus deterring participation. This large marginal tax rate is the result of two known features of the US tax system: joint filing and progressivity. Since wages are larger in the big city than in the small city, taxation would be expected to discourage female participation to a greater extent in the former.

a policy perspective. A greater role of unobserved preferences in driving the BCCP highlights the importance of focusing on more general policies, as local policies would mostly shift people around different locations.

Finally, I use the model to evaluate the impact of two policy reforms: i) a childcare subsidy and ii) a reduction in commuting times. A key advantage of this framework is that it enables one to evaluate how the implementation of any of these policies in one of the two cities spills over the whole economy, as well as to assess the role of sorting on idiosyncratic preferences in carrying over the impact of a policy reform in a city to the other city. Furthermore, this framework is useful to compare how coordinated policies, that can be thought of policies implemented at the federal level, fare relatively to local policies that affect a single city. Therefore, I first compare the economic impact of implementing childcare subsidies in both cities to the case in which only one city subsidizes this expenditure. Next, I use the model to evaluate the impact of an infrastructure investment that reduces the commuting time in the big city but does not affect the travelling time to work in the small city.

The first exercise reveals that childcare subsidies are very effective at raising female LFP. In particular, the aggregate participation rate increases by 9 p.p. when childcare subsidies are implemented in both cities, and by almost 5 p.p. when only the big city subsidizes childcare. In the latter case, geographical sorting on unobserved preferences gives rise to large differences between the participation rates in each city. This calls for some caution when evaluating the impact of local policies by comparing a location to surrounding ones, since the large increase in the big city's participation rate comes at the cost of a large drop in the small city. In addition, childcare subsidies reduce average welfare due to the tax rise that is required to finance their cost. More surprisingly, and despite the large increase in female LFP, childcare subsidies cause a 4% drop in output per capita. This decline is due to changes in the intensive margin, since households' market work becomes more evenly shared among partners when subsidies are in place. Given that average productivity increases substantially with hours worked in the long hours occupation, the decline in working hours translates into a large output cost.

In contrast, a decline in the big city's commuting time has a positive impact on this location's female LFP, hours worked, output per capita and welfare. In particular, this policy results in a 2% increase in output per capita, which leaves some room for financing the cost of the infrastructure. However, most of the increase in LFP is driven by the sorting into the big city of couples in which both partners participate in the labour market so that the policy has a very limited impact on the aggregate participation rate in the economy.

The rest of the paper is organized as follows. Section 1.2 reviews the related literature. Section 1.3 documents the BCCP and other empirical regularities. Section 1.4 describes the theoretical framework. Section 1.5 contains the calibration of the model and the main quantitative results. Section 1.6 shows the implications of different policy changes. Section 1.7 provides an extension in which heterogeneity in skills is introduced. Finally, Section 1.8 concludes.

1.2 Related Literature

As mentioned above, despite the existence of a large body of literature analyzing women's labour supply (Killingsworth and Heckman 1986; Blundell et al. 2007 summarize the key insights of the literature), spatial differences in women's LFP have been overlooked. Some exceptions are Ward and Dale (1992) and Odland and Ellis (1998), who study geographical variation in female labour supply in the US. However, the former study only focuses on the impact of part-time and full-time geographical variation in female LFP, while the latter only analyzes the largest metropolitan areas in the US.

More recently, Phimister (2005) compares the LFP of men and women in urban as opposed to rural areas in the UK. While no significant spatial differences are found as regards male LFP, there is a LFP premium for women in urban areas. It is argued that this premium should result from the fact that larger and denser markets are particularly beneficial for women, since they often restrict their job search to areas that are close to their residence. Unlike this paper, I restrict my attention to urban areas and focus on comparisons between small and big cities. In addition, I make a further distinction among women with and without children, since motherhood is the main determinant of women's LFP. Together, these findings suggest that the relationship between population density and female LFP is non monotonic.

My work is also related to another line of research that studies the sorting of couples into different city sizes. For instance, Costa and Kahn (2000) predict a higher propensity to choose a large city as a residence among *power couples*– couples in which both spouses have college degrees and participate in the labour market–since denser labour markets are more likely to offer both partners a good job match. However, Compton and Pollack (2004) find no empirical support for their hypothesis, since their results show that couples' propensity to locate in large metropolitan areas is solely driven by husband's educational attainment. According to these authors, the concentration of power couples in larger metropolitan areas is due to higher rates of power couple formation in these locations. Consistent with these findings, I show that childless couples with high labour attachment (my model's equivalent of power couples) sort into large cities, which drives the city size participation premium observed for this group. However, their findings are at odds with the geographical sorting patterns I observe among couples with children. One possible explanation is that couples underestimate the cost of having children (as in Kuziemko et al. (2018)). This, together with the fact that location choices are fairly persistent, would imply that power couples that moved to large metropolitan areas in order to benefit from the better working opportunities adjust to the large cost of having children in these locations by adjusting the wife's labour supply instead of relocating to a smaller city.

Some recent work has focused on the impact of commuting times on LFP, a characteristic that is highly correlated with city size. Black et al. (2014) study differences in the LFP of married women across the 50 largest US cities and find that lower LFP rates are related to longer commutes. Likewise, Carta and De Philippis (2018) find that commuting times increase intra-household specialization in Germany, increasing the husband's working time and reducing the probability that the wife participates in the labour market. Lastly, Farré et al. (2019) obtain a negative estimate for the causal effect of commuting times on female LFP in the US, using city shape as an instrument for commuting times. Although commuting times are a key determinant of female labour supply, my contribution here is to show the importance of analyzing them in a general equilibrium setting and in conjunction with the occupational structure of the city.

Regarding the different occupational structure across city sizes, Rossi-Hansberg et al. (2019) show that 'cognitive' and 'non-routine' occupations, such as those carried out by lawyers, computer scientists, or researchers, are disproportionately represented in larger cities while, by contrast, occupations that do not primarily reflect cognitive non-routine tasks are increasingly located in smaller cities. Similarly, Santamaria (2019) documents a smaller team size in firms located in big cities due to technological differences with respect to small cities, which results in a greater proportion of managers in big cities. Moreover, Rosenthal and Strange (2008) find that, while professionals' working hours increase with the density of workers in the same occupation, nonprofessionals' working hours decline. Since the centers of large cities are very densely populated, their findings are consistent with the evidence provided in this paper. Finally, Erosa et al. (2017) show that occupational choices are influenced by differences in the reward to working long hours across occupations and asymmetries in domestic chores across gender. Accordingly, I add this feature to the model to explain the observed spatial differences in women's LFP.

1.3 Empirical Evidence

1.3.1 Data

I combine the American Community Survey for the years 2005-2010, the US Census 2000, and the 1990 1% metro sample, all of them available at IPUMS (Ruggles et al., 2018).[5]

Regarding the definition of the city, I conduct my analysis at the metropolitan statistical area (MSA) level, which is widely used in the literature to study local labour markets. The MSA is defined as "*a region consisting of a large urban core together with surrounding communities that have a high degree of economic and social integration with the urban core*" .

Throughout the paper, I classify MSAs into three (equally populated) categories according to size (small, medium, and big).[6] I choose to compare small and big cities because they are intrinsically different, which allows one to consider different industry composition or urban structure. For instance, since big cities typically host the headquarters of big companies and public institutions, this will result on a higher labour demand for managerial and professional occupations which are associated in general to longer working hours.

Lastly, I exclude from my sample individuals who are less than 25 years old

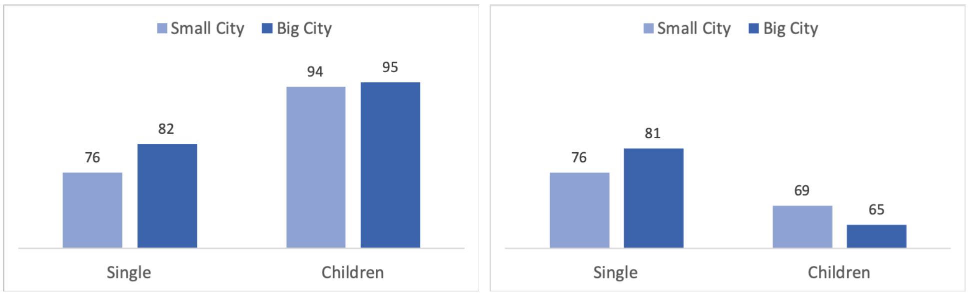
[5]The American Community Survey does not contain information on MSA in the years 2001-2004, so I cannot use them. The same circumstance applies to the 2010 Census.

[6]This results on the following thresholds: small MSAs are those below 1 million inhabitants while big MSAs are those with at least 4 million inhabitants. The smallest big city according to this classification is Atlanta while the largest small city is Richmond. All results in this paper are robust to alternative classifications of city size

and may still be undertaking education and individuals over 55 to avoid non-participation related to pre-retirement.

1.3.2 Empirical Regularities

Figure 1.1 shows the LFP rates of men and women by city size for two different groups: single individuals with no children and married (or cohabiting) individuals with children younger than 12 years old.[7] The light blue bars show the participation rate of the corresponding demographic group in small cities, while the dark blue bars refer to big cities. As can be observed, the LFP rates increase with city size for all demographic groups, except for women with children, whose participation rate is higher in small cities than in big cities. In Figure A.1 in Appendix A, I show that the lower LFP of women with children in big cities relative to small cities has been a persistent feature of the US labour markets over the last decades.



(a) Men **(b)** Women

Figure 1.1 – Differences in participation across city size

Notes: This figure shows the LFP rate of individuals aged 25 to 55 by gender, parenthood, and city size. Source: ACS 2010 available in IPUMS, own elaboration.

It is well known that the characteristics of the population vary with city size. Big cities host younger and more skilled people (Eeckhout et al., 2014). In addition, the proportion of migrants is higher in big cities (Albert and Monras, 2018). This raises the question on whether the BCCP may result from a different composition of the population in terms of age, nativity, and skill level across city size. Controlling for these characteristics in a regression framework, the results in Table 1.1 show that observable characteristics explain about half of the observed gap in participation.

[7]Throughout the analysis, I do not distinguish between married and cohabitating individuals. I refer to both groups as "married couples" for simplicity.

Table 1.1 – The Big City Child Penalty

	(1) LFP Women Children	(2) LFP Women Children	(3) LFP Women Children	(4) LFP Women Children
Big City	-0.057***	-0.038***	-0.038***	-0.032***
	(0.001)	(0.002)	(0.002)	(0.002)
Year FE	Yes	Yes	Yes	Yes
State FE	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Spouse	No	No	No	Yes
R-squared	0.004	0.012	0.073	0.086
Observations	745138	745138	745138	745138

Notes: This table displays differences in the likelihood to participate among married women with children across city size, according to a linear probability model. The dependent variable is a dummy variable equal to one if the woman participates in the labour market. Big city is an indicator variable equal to one if the woman resides in a big city. The first column includes only year fixed effects while in the second column I also add state fixed effects. In the third column I control for individual characteristics: age, age squared, dummies for race, dummies for 5 different education levels, a dummy equal to one if the individual is born outside the US, the number of children, and the age of the youngest child. In the fourth column I also control for spouse's characteristics. The sample is restricted to married women with children under 12 who live in metropolitan areas. I also exclude women below 25 or over 55 years old. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

More generally, a married woman's decision to participate in the labour market will depend mainly on three factors: (i) the expected wage that she would receive in the market; (ii) other available income/wealth; and, (iii) the number of children she has and their age. A classical problem in assessing the role of wages on labour force participation is that we only observe the wages of women that have chosen to participate and hence are likely to be a non-random sample, with the selection problem becoming worse as female LFP rates are lower. Therefore, in the regressions I control for observable characteristics that are likely to influence women's expected wage, like age, education, race, and nativity. An additional problem regarding this regression is that wealth is not reported in the data, with only husband's income being available. However, husband's income is likely to be endogenous since men whose spouses do not work in the market are more likely to choose high-paying jobs. In order to address this problem, I control for husband's predetermined observable characteristics that should proxy well for his income. Table 1.1 shows the estimates of a linear probability model where an indicator variable equal to one if the woman participates in the labour market is regressed on two dummies for each size category, medium and big (notice that the reference category is thus small size cities). Throughout my analysis I only focus on the coefficient of the big city size category, which corresponds to the BCCP. Estimates regarding the LFP of women in medium size cities lie in between the small and big cities, and are not reported here in the benefit of clarity. Columns (1) shows that the likelihood of participating in the labour market is almost 6 p.p. lower in a big city than in a small city, though this difference shrinks to around 4 p.p. after introducing state fixed effects. In column (3) I control for the woman's observable characteristics and the number of children and the age of the youngest and the estimate of the BCCP remains unaltered. Lastly, in column (4) I also control for husband's observables, yielding a 3.3 pp. estimate.[8]

As outlined earlier, the remaining BCCP may be explained by: (i) city characteristics that affect incentives to participate in the labour market, and (ii) unobserved heterogeneity, such as differences in preferences. In other words, to the extent that labour force attachment for women with children, and thus preferences for a stay-home parent in the household, may not be too correlated with observable characteristics, geographical sorting on this dimension will also be part of the remaining BCCP.

In what follows, I show which city characteristics are the most relevant factors explaining the BCCP in my data. Moreover, in Section 1.4, I construct a quantitative model in which agents differ on idiosyncratic preferences for a stay-home mother, which allows me to take into account the role of unobserved heterogeneity. For completeness, Table 1.2 documents differences in the likelihood to participate across city size for all demographic groups controlling for observables showing once more that larger cities have higher LFP rates than smaller cities for all demographic groups except for women with children.[9]

[8]Table A.1 in Appendix A reports the estimated coefficients of most of the included controls in the regression. All coefficients have the expected sign.

[9]In Section A in Appendix A I show that results do not change if I run a pooled regression for all individuals and include interactions of the impact of population with dummies for sex, marital status, and children. My preferred specification is the one reported in the main text because it allows me to control for spouse characteristics and its interpretation is more straightforward.

Table 1.2 – LFP and City Size

	(1) LFP Single Men	(2) LFP Married Men No Children	(3) LFP Married Men Children	(4) LFP Single Women	(5) LFP Married Women No Children	(6) LFP Married Women Children
Big City	0.056***	0.007***	-0.000	0.031***	0.007***	-0.032***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.094	0.037	0.042	0.104	0.077	0.087
Observations	1360098	869874	1409777	880421	846583	1279518

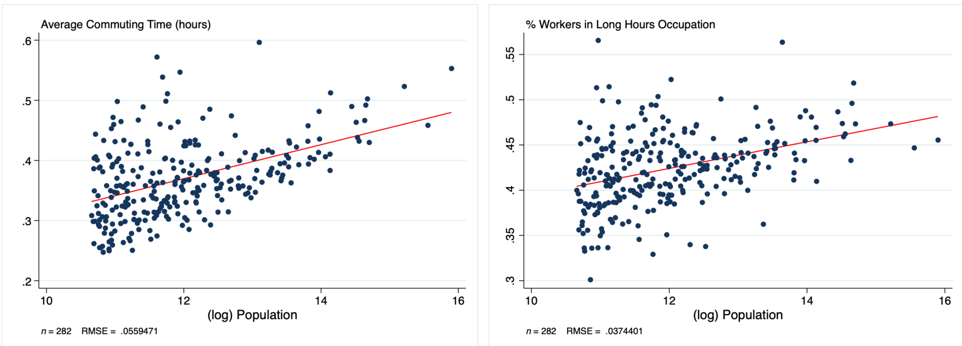
Notes: This table displays the results of regressing a dummy equal to one if the person participates in the labour market on a big city size dummy according to a linear probability model. All regressions include year and state fixed effects and control for observable characteristics: age, age squared, race, education, and a dummy equal to 1 if the person is foreign-born. For married individuals, regressions include controls for spouse's observable characteristics as well. For individuals with children, I also control for the number of children and the age of the youngest child. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

Commuting times and Occupational Structure across City Size

In this subsection I first summarize the main differences between small and big cities that are at the core of my proposed mechanism earlier: commuting times, occupational structure, and the price of childcare. Next, I provide evidence suggesting that they are important drivers of the BCCP.

As several studies have shown, commuting times to work are an important determinant of female labour supply (Black et al. 2014; Carta and De Philippis 2018; Farré et al. 2019). Since commuting times are a time fixed cost, they have a direct impact on the likelihood to participate in the labour market, and this effect is more likely to matter for women with children. Furthermore, average commuting times are likely to have an additional effect among this group of women since they are a proxy for how difficult taking the children to school or to extracurricular activities may be. The left panel of Figure 1.3 shows that more populated cities involve longer commuting times, since they are typically more extensive and more densely populated, which creates congestion. The average commuting time in each city size is displayed in Table A.5 in Appendix A.

To assess differences in occupational structure across city size, I construct two occupation categories following Erosa et al. (2017). Thus, I rank occupations by the average hours worked by single individuals at the national level and split them into two groups. I will refer to the group with above median average hours of work as the *long-hours* occupation. Table A.5 in Appendix A displays some summary statistics for occupations across city size. Individuals working in the long hours occupation are more educated on average and enjoy higher wages. Some examples of occupations in this group are lawyers, judges, managers, and other professional occupations. I use the share of single individuals employed in the long-hours occupation as a proxy for the comparative advantage of a city in these occupations. The right panel of Figure 1.3 shows that larger cities have a comparative advantage in long hours occupations. This is consistent with Rosenthal and Strange (2008), who show that working hours for high-skilled individuals increase with population density, which is higher in large cities; and with Rossi-Hansberg et al. (2019) and Santamaria (2019), who show that big cities host a greater proportion of cognitive non-routine occupations and managers, respectively.



(a) Commuting times **(b)** Occupations

Figure 1.3 – Correlation between commuting times and the comparative advantage of a city in the long hours occupation with city size

Notes: These graphs show the correlation between the (log) population of a city and: the share of single workers employed in occupation 1 in a city (panel a) and the average commuting times of men (panel b). Occupations are classified as follows: I first compute the mean hours worked for single individuals in each occupation, rank all occupations by the level of mean hours, and separate them into two groups that are equal in size. Occupation 1 is the occupation with longer working hours. Source: ACS 2000 available in IPUMS, own elaboration.

The price of childcare

Another potential driver of differences in LFP across city size could be differences in the price of childcare. Unfortunately, there is no data available at the city level on this issue. The non-profit organization Child Care Aware of America provides data on the price of childcare at the state level. Since the production of childcare can be thought as a function of land and wages, I can use this data to obtain estimates of the relative importance of each factor and then produce fitted values at the city level that allows to account for the price of childcare in my regressions. In Table 1.3 I regress the price of center-provided infant childcare at the state level on the state average price of housing (that should be very correlated with the price of land) and the median wage in the state.[10] These two covariates explain up to 76.5% of the price of childcare.

Table 1.3 – The Price of Childcare

	(1) (log) Price Childcare	(2) (log) Price Childcare
(log) Housing Price Index	0.249**	0.242**
	(0.077)	(0.077)
(log) Median Hourly Wage	1.737***	1.786***
	(0.143)	(0.145)
Year FE	No	Yes
R-squared	0.761	0.765
Observations	192	192

Notes: This table displays the result of regressing the (log) price of childcare at the state level on a state (log) housing price index and the (log) median hourly wage in the state. The second column adds year fixed effects to the regression. Standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

Which factors explain the BCCP?

Both longer commuting times and a higher concentration of occupations with a high return on working long hours may increase incentives to specialize within the household, with one partner working in the market and the other partner, typically the wife, taking on home production and especially childcare. On the

[10] I perform hedonic regressions and then compute a Housing Price Index for each city. The hedonic regression includes the number of rooms, the year the house was built, and the type of building. Therefore, this approach considers differences in housing characteristics across different cities.

one hand, commuting times impose a fixed time cost to work, which may be particularly important for couples with children, since these families are more time constrained. On the other hand, a higher reward to working long hours creates incentives to increase one of the partners' market hours and collect this return at the expense of housework hours. The higher earnings from working long hours relax the household's budget constraint while the longer working hours tighten the time constraint. Therefore, the other partner's incentives to participate in the labour market are lower.

Table 1.4 – The Big City Child Penalty (BCCP)

	(1) LFP Women Children	(2) LFP Women Children	(3) LFP Women Children	(4) LFP Women Children	(5) LFP Women Children	(6) LFP Women Children
Big City	-0.032***	-0.005*	-0.029***	-0.020***	-0.019***	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Avg Commute		-0.264***				-0.255***
		(0.014)				(0.015)
% Workers LH			-0.083***			-0.037
			(0.018)			(0.020)
(log) Housing				-0.034***		-0.008
Price Index				(0.004)		(0.005)
Avg Wage					-0.085***	0.002
					(0.010)	(0.012)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Spouse	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.086	0.087	0.087	0.087	0.087	0.087
Observations	745138	745138	745138	745138	744685	744685

Notes: This table shows the impact of several city characteristics on the BCCP. The dependent variable is a dummy equal to one if the woman participates in the labour market. All columns include year and state fixed effects and control for individual's observables and partner's as in Table 1.1. The sample is restricted to married women with children under 12 who live in metropolitan areas. I also exclude women below 25 or over 55 years old. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

I assess the role of each of the above mentioned channels in Table 1.4. In column (1), I show again the BCCP after controlling for individual and spouse's observable characteristics (those reported in the last column in Table 1.1). In the remaining columns, I add: (i) the average commuting time in the city; (ii) the share of workers employed in the long hours occupation in the city; (iii) the (log) housing price index of the city, and (iv) the average wage level in the city. The last two variables control for differences in the price of childcare.

I choose to introduce them directly because they may affect LFP for other reasons that do not relate to childcare itself, as it will be made more clear with the model.[11]

Focusing on the BCCP coefficient (the coefficient on Big City), we can see that accounting for the average commuting time reduces the BCCP the most. However, this result is driven by low-skilled women. In Section 1.7 , I show that the explanations for the BCCP vary largely with the female skill levels. Indeed, the BCCP of high-skilled women is better explained by a higher share of workers in the long hours occupation and more expensive childcare in big cities. In column (5) I include all city characteristics and show that together they fully account for the size of the BCCP.

1.4 The Model

In this section I build a quantitative spatial model that that helps assess the relative impact of the channels outlined in the empirical section and to account for some additional determinants like the roles of geographical sorting on unobservables and of taxation. Moreover, in Section 1.6 I use the model to evaluate the consequences of two policy changes aimed at raising female LFP.

1.4.1 Environment

As in the previous discussion. there are two representative cities in this economy, small and big, indexed by $s \in \{S, B\}$. While the model could be extended to include a wider set of city sizes, I choose to focus on these two groups for ease of exposition. One way to think about these cities is to consider them as two groups of multiple identical cities, a small city group and a big city group. Cities differ in productivity, A, commuting times, τ, and housing costs, H.

Agents choose among two occupations, indexed by $j \in \{1, 2\}$, which differ in the reward to working long hours. Let n_j denote efficient units of labour in occupation j, h, hours worked, and $f^j(h) : N \to N$ be the rate at which hours worked are converted to efficient units. I assume that labour services are increasing in the amount of hours worked $f_h^j > 0$, and that it increases at

[11] While average unemployment rates may seem important, I show in Table A.7 in Appendix A that they are not part of the explanation.

a convex rate in occupation 1, $f^2_{hh} > 0$, while this rate is linear in occupation 2, $f^2_{hh} = 0$. The difference in the return to working long hours implies that individuals will only choose to work in occupation 1 if they are willing to supply enough hours to the market. Otherwise, the return to working hours would be larger in the linear occupation. Hence, average hours worked in occupation 1 will be longer than in occupation 2, so that occupation 1 will be considered in the sequel as the long- hours occupation.

Consistent with the data, I assume that the big city has absolute advantage, $A^B_j > A^S_j \; \forall j$, and comparative advantage in the long- hours occupation $A^B_1 > A^S_1$. In addition, the big city will be subject to longer commuting times $\tau^B > \tau^S$, and higher housing rents, $H_B > H_S$. In the model, productivity advantages act as an agglomeration force, while commuting times and rent differentials (together with decreasing returns to labour) play the role of congestion forces.

The economy is populated by couples composed by a man (m) and a woman (f), each of them endowed with one unit of time. Some couples have children, who are modelled as a time cost ($\kappa \geq 0$) for the couple. I assume that men always work in the market while women may not participate in the labour market.

Each household consumes one unit of housing at a given city price. This implies that the price of housing is exogenous in the model and does not respond to the city housing demand. While this assumption may seem restrictive, it is not central for the mechanisms at play. Moreover, the price of housing affects the price of childcare in the city, which is computed as a combination of housing and the wage in the linear occupation in the city according to the coefficients estimated in the regressions for the price of childcare (Table 1.3).

Since progressive taxation plays an important role in female LFP (Guner et al., 2012), I introduce taxes in the model and assume that government spending is wasteful.

Lastly, households have idiosyncratic preferences for the different combination of occupations and cities that are distributed according to a Generalized Extreme Value distribution. Let $o \in \{1, ..., O\}$ denote each combination of female occupation, male occupation, and city. I partition this set into 6 non-overlapping nests, denoted B_k, which depend on the pair of female and male occupations chosen ($k \in \{j_m \in \{1, 2\} \times j_f \in \{0, 1, 2\}\}$), where 0 denotes the case in which the wife does not participate in the labour market. The utility of household i choosing option o, U_{io}, can be decomposed into a component that is common to all individuals, V_o, and some unobserved heterogeneity that reflects household i's preferences for option o, ε_{io}, implying that, $U_{io} = V_o + \epsilon_{io}$. Let the unobserved utility $\{\varepsilon_{io}, ..., \varepsilon_{iO}\}$, have the following cumulative distribution:

$$F(\varepsilon) = \exp\left(-\sum_{k=1}^{K}\left(\sum_{j\in B_K} e^{-\frac{\varepsilon_{io}}{\lambda_k}}\right)^{\lambda_k}\right).$$

This distribution gives rise to a nested model for idiosyncratic preferences with occupation choices in the upper nest and location choices in the bottom level. Figure 1.5 provides a graphic representation of the model for idiosyncratic preferences. The main feature of this distribution is that it allows for correlation of options within the same nest, that is, amongst the unobserved utility component between small and big cities within each occupational nest. In this fashion, parameter λ in the expression above can be interpreted as a migration elasticity.

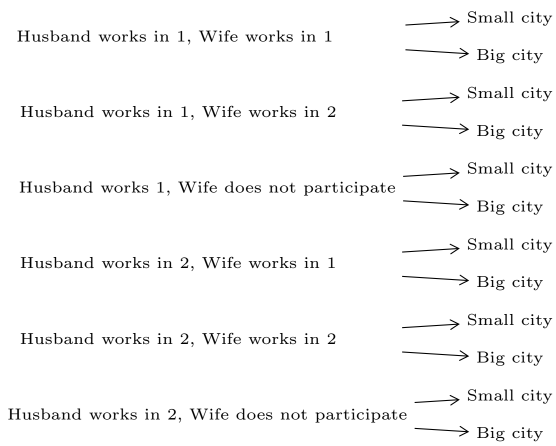


Figure 1.5 – Nested Model for Idiosyncratic Preferences

1.4.2 Households' problem

Couples decide in which city to live and work, $s \in \{S, B\}$, male's occupation $j_m \in \{1, 2\}$, female's LFP and occupation $j_f \in \{0, 1, 2\}$, where $j_f = 0$ denotes the case in which the woman does not participate in the labour market. For a given choice of o ($o \in \{j_m \times j_f \times s\}$), the amount of hours supplied to the labour market, and each partner's time spent in home produced childcare become the solution to the following optimization problem:

$$\begin{aligned}
&\max_{c,h_f,h_m,t_m,t_f,k} U_o(c,h_f,h_m,t_f,t_m,\varepsilon_{io}) \\
\text{s.t.} \quad & c + p_k^s k + H^s \leq (1-t_p)w_{o,m}^s n_{o,m} + (1-t_p)w_{o,f}^s n_{o,f} - T(I,k) \\
& T \geq t_m + \tau^s + h_{o,m} \\
& T \geq t_f + \mathbb{I}_{j_f \neq 0}\tau^s + h_{o,f} \\
& \kappa = g(k,t_f,t_m) \\
& n_j = f^j(h_j),
\end{aligned}$$

where c denotes consumption; k, childcare bought in the market at price p_k; h_m, male hours worked; h_f female hours worked; t_m, time spent with the children by the father; t_f, time spent with the children by the mother; $w_{o,m}$ ($w_{o,f}$) the wage per efficient unit in the chosen occupation for the husband (wife), κ is the time cost of children; and τ^s denotes the commuting time in city s and is only paid if the individual supplies a strictly positive number of hours. Recall that n_j denotes efficient units of labour, which are an occupation-specific function of hours worked.

Furthermore, couples are subject to two forms of taxation: proportional payroll taxes, t_p, and progressive income taxes, $T(I,\kappa)$, where I denotes income net of payroll taxes. I consider different tax rates depending on whether the couple has children or not. I assume that the utility function has standard properties, $U_c > 0, U_h < 0, U_t > 0, U_{cc} < 0, U_{hh} > 0, U_{tt} < 0$ and that couples share equal consumption. The family has access to a technology that combines both partner's time (t_f and t_m) and market childcare k. The solution of the problem must also fulfill the following restrictions: $t_i \geq 0$, and $h_i \geq 0$, which avoids choosing negative levels of time in childcare or working hours.

Given the solution to this problem, I can compute the indirect utility value of household i for a given city s and combination of occupations j_m and j_f, which I denote by $\tilde{V}_{io}$. This indirect utility can be decomposed in one part that is common to all individuals and another part capturing the individual preference: $\tilde{V}_{io} = V_o + \varepsilon_{io}$. Using the properties of the specific form of generalized extreme distribution, the probability that a household i chooses option o, π_o

is given by:

$$\pi_{io} = \frac{e^{\frac{V_{io}}{\lambda_k}} \left(\sum_{o \in B_k} e^{\frac{V_{io}}{\lambda_k}} \right)^{\lambda_k - 1}}{\sum_{l=1}^{K} \left(\sum_{o \in B_l} e^{\frac{V_{io}}{\lambda_l}} \right)^{\lambda_l}}. \tag{1.1}$$

Finally, by the law of large numbers, this probability translates in the share of households that choose option o in the model economy.

1.4.3 Firms

For simplicity, I assume that there is a single consumption good that is tradable and that cities are small open economies. As a result, the price of consumption is the same in all cities and is normalized to 1, that is, $p^B = p^S = 1$.

Output in each city and occupation is produced using only labour and is subject to decreasing returns to scale:

$$F(N_j^s) = A_j^s (N_j^s)^\alpha \qquad 0 < \alpha < 1.$$

For simplicity it is assumed that firms belong to non-modelled owners. The labour demand in each occupation for a given city s:

$$N_j^d = \left(\frac{\alpha A_j^s}{w_j^s} \right)^{\frac{1}{1-\alpha}}.$$

Thus, the labour demand in occupation j increases with city-specific productivity, A_j^s.

1.4.4 Aggregate Labour Supply

Let $n_{sj_mj_f}(\kappa)$ denote optimal hours expressed in efficient units supplied by an individual with children equal to $\kappa \in \{0, \bar{\kappa}\}$ that works in occupation j_m, lives in city s and whose partner works in j_f. Then, the aggregate labour supply of individuals in city s and occupation 1 by parenthood is given by:

$$N_1^s(\kappa) = \sum_{j_f} \pi_{s1j_f}(\kappa) n_{s1j_f}(\kappa) + \sum_{j_m} \pi_{sj_m1}(\kappa) n_{sj_m1}(\kappa),$$

where the first term adds up the fraction of males that work in occupation 1 in city s, for a given female occupation, and the second terms adds up the fraction of females that work in occupation 1 in city s, for a given male occupation. Lastly, aggregate labour supply in a city s and occupation 1 in efficient units is given by adding up the labour supply of individuals with children and individuals without children:

$$N^s_{1s} = s_\kappa N^s_1(\kappa > 0) + (1 - s_\kappa)N^s_1(\kappa = 0)$$

where s_κ is the proportion of individuals with children in the economy.

1.4.5 Childcare

Couple's childcare demand come as the solution of the household problem, k^*. Aggregating across all households in a city I obtain the city demand for childcare as a function of its price, $K^{sd}(p^s_k)$. Regarding the supply of childcare, I consider that it is perfectly elastic at $p^s_k = \exp(\beta_0 + \beta_1 \log H^s + \beta_2 \log w^s_2)$, where β_1 and β_2 are given by the regression of the price of childcare on the average housing price and the median wage in Table 1.3. Lastly, the intercept β_0 is set to match the observed average expenditure in childcare in the US.

1.4.6 Equilibrium

An equilibrium in this context is given by set of wages per efficient unit of labour $w^B_1, w^B_2, w^S_1, w^S_2$ such that agents maximize their utility and labour markets clear. Notice that decreasing returns to labour in each occupation and city ensures a unique equilibrium in which both occupations are present in the two cities.

1.5 Quantitative Results

In this section I explain in detail the parameterization of the model as well as its calibration to reproduce the most salient features of labour markets across city size in the US.

1.5.1 Functional Forms

I assume that household utility is represented by the following function:

$$U(c, h_m, h_f, t_m, t_f, \varepsilon) = \log c - \psi(t_m + h_m + \tau)^{1+\frac{1}{\gamma}} - \psi(t_f + h_f + \mathbf{1}_{h_f>0}\tau)^{1+\frac{1}{\gamma}} \\ + \nu_m \log(t_m) + \nu_f \log(t_f) + \varepsilon.$$

Thus, households derive utility from common consumption and from spending time with their children. Since mothers usually spend more time with their children, I let ν vary with gender, which allows me to reproduce this fact in the model. Moreover, individuals bear a disutility from working, commuting, and spending the time with the children, as these activities reduce time for leisure.

Households have access to the following technology to take care of the children: they can combine each partner's time with the children to childcare bought in the market and I assume that all inputs are perfect substitutes.

$$g(k, t_f, t_m) = k + t_m + t_f.$$

Notice that, given that households derive utility from time spent with the children, they will never outsource the full-time cost of children. In contrast, if childcare becomes too expensive, couples may prefer to take care of the children themselves and avoid buying any childcare at all. While this is plausible when only one partner participates, taking on full childcare needs when both partners work is not reasonable, as working schedules tend to overlap and children younger than 6 do not attend to school. For this reason, I assume that, whenever both partners participate in the labour market, they necessarily need to pay for some childcare, k_{min}. According to Urban Institute calculations from the 1997 National Survey of America's Families, more than 80% of families in which both parents work hire at least some hours of childcare, showing that this assumption is not at odds with reality (Capizzano, 2000).

Regarding the relationship between hours worked and efficient units of labour I follow Erosa et al. (2017) in choosing the following functional form:

$$n_1 = h^{1+\theta}, \quad \theta > 0$$
$$n_2 = h_2$$

As already mentioned, this implies that the rate of conversion of working hours in efficient units of labour is convex in occupation 1 and linear in occupation 2. Notice that, as also discussed earlier, the convexity in occupation 1 will induce the sorting of individuals with a greater availability to work longer hours in that occupation. Absent wage differences across occupations, the return in occupation 1 is only higher above a threshold in hours worked. Moreover, wage differences across occupations only change the exact location of the threshold. Hence, the choice of occupation 1 as the long-hours occupation and occupation 2 as the flexible occupation.

Lastly, I use (Guner et al., 2012)'s taxation functional form given by:

$$T(I,\kappa) = [t_{1,\kappa} + t_{2,\kappa} * \log(I)] * I,$$

where $\quad T_I > 0$ and $T_{II} > 0$.

1.5.2 Parameters

Table 1.5 displays the value of non-calibrated parameters, which are taken from the data or from the literature.

Time. I endow each individual with 14 hours of time, in line with the findings of Aguiar and Hurst (2007).[12] I set the time cost of children (κ) equal to the average time spent with kids by women out of the labour force using data from the American Time Use Survey. Similarly, I take the (2-ways) average commuting time in each city size (τ_s) from the 2000 US Census.

Occupations. I assume that both occupations bear the same degree of returns to scale (α) and set it equal to the labour share. Since there is no capital in my model, the labour share can be calculated by simply subtracting 1 from the profit share, which I take from Barkai (2016).

There exist several estimates of the return to long working hours in the literature. While definitive estimates are not available, I follow Erosa et al. (2017) and set $\theta = 0.6$. A careful discussion of the estimates in the literature can be found in their paper. Their choice aims at capturing both the static and dynamic effects of longer hours on earnings.

[12] However, time constraints are never binding in equilibrium. The reason is that, given the calibration for ψ, the disutility from working and spending time with the children is too large at $h + t + \tau = T$.

Demographics. I use 2000 US Census to compute the proportion of couples with children, s_κ.

Table 1.5 – Non Calibrated Parameters

Parameter	Description	Source	Value
Time			
κ	Time cost of children	ATUS	7
τ_B	Avg commuting time in B (hours)	2000 US Census	1.1
τ_S	Avg commuting time in S (hours)	2000 US Census	0.8
T	Total available time (hours)	Aguiar and Hurst (2007)	14
Occupations			
α	1-Profit Share	Barkai (2016)	0.85
θ	Degree of convexity	Erosa et al. (2017)	0.6
Demographics			
s_κ	% couples with children	2000 US Census	0.7
Preferences			
λ	Variance idiosyncratic taste shock	Diamond (2016)	0.3
γ	Intertemporal elasticity labour supply	Domeij and Floden (2006)	0.4
Price of Childcare			
β_1	Price of childcare: Housing	ACS & Childcare Aware	0.25
β_2	Price of childcare: Wage	ACS & Childcare Aware	1.7
Taxes			
t_p	Payroll tax rate	Guner et al. (2012)	0.086
$t_{1,\kappa=0}$	Tax function no children	Guner et al. (2012)	0.113
$t_{2,\kappa=0}$	Tax function no children	Guner et al. (2012)	0.073
$t_{1,\kappa>0}$	Tax function children	Guner et al. (2012)	0.084
$t_{2,\kappa>0}$	Tax function children	Guner et al. (2012)	0.09

Notes: This table summarizes the values and sources for all non calibrated parameters. HS (LS) denotes high-skilled (low-skilled) which are defined as those who hold at least a bachelor degree.

Idiosyncratic Preferences. The parameter λ_k in the distribution of idiosyncratic preferences governs the variance of the unobserved utility component ε_{ik} within each nest k. As this variance increases, larger differences in the common utility component, V_k, are needed for an individual to switch across options. As pointed out earlier, since I consider the city size choice to be in the last nest, λ_k determines the migration elasticity with respect to changes in V_k, for a given occupational choice k. Hence, a larger λ_k implies lower mobility across city size. As a consequence, the degree of sorting on unobservables that the model produces depends crucially on the value of this parameter.

I assume that this migration elasticity is the same across occupations (that is, in all nests). Diamond (2016) estimates a separate migration elasticity for each of the two head of household's education levels considered with respect to wages, housing rents, and amenities. Given this author's specification, the migration elasticity estimates with respect to wages identify the variance of the unobserved utility component for each education group.[13] In this section, I set λ equal to a weighted average of the low- and high-skilled estimates, according to their weight in the population. In Section 1.7, I will let this parameter vary with education.

Price of childcare. As explained in the empirical section, the price of childcare is not available at the city level. However, I show that the average price of housing and the median wage are very good predictors of this price at the state level, for which data are available. In the quantitative model, I use the estimated coefficients on the median wage and the average price of housing to set the price of childcare in each city. In particular, I consider the price to be a function of the calibrated housing price and the equilibrium wage in the flexible occupation and use the estimated coefficients to weigh each factor by its relative importance. Moreover, as mentioned earlier, I add an intercept term that I calibrate to match the average expenditure on childcare in the data. The precise interpretation of this constant term is deferred to the next section.

Taxes. I set the parameters in the taxing function ($t_{j,\kappa}$) to those estimated by Guner et al. (2012). They construct income tax functions by marital status and parenthood, which suits my framework perfectly. More concretely, they retrieve average income tax rates from the amount of effective taxes paid by each type of household and their reported income. I also include their estimate for the payroll tax (t_p), which is obtained using data on social security contributions.

1.5.3 Calibration

The rest of parameters are calibrated to match the most salient features of labour markets in small and big metropolitan areas. Table 1.6 summarizes the calibration exercise. Column (2) displays the resulting parameter values, while column (3) describes the targeted moment. Lastly, columns (4) and (5) compare the targeted moments in the data and in the model, respectively.

[13] Diamond (2016) divides the indirect utility function by the variance of the idiosyncratic shock. Given that the coefficient on wage is normalized to one, the estimated coefficient corresponds to the inverse of the variance. Thus, I set λ as the inverse of her reported coefficients.

Though the calibration exercise is done jointly, in what follows I discuss this exercise by groups of parameters that are closely linked to the chosen targets.

Preferences. I calibrate the parameters governing the weight of the disutility from working and spending time with the children, ψ, and the utility from spending time with the children, ν_m and ν_f, to match average hours worked and average hours spent with the children by each parent.

Table 1.6 – Calibrated Parameters

Parameter	Value	Targeted Moment	Data	Model
Preferences				
ψ	0.00012	Average hours worked	8.37	8.35
ν_m	0.08	Male average hours with children	1	1
ν_f	0.1	Female average hours with children	1.6	1.6
Productivity				
A_2^B	1.1	LH occupation wage premium in B	1.30	1.30
A_1^B	1.35	Big city wage premium in LH occupation	1.60	1.59
A_1^S	1.05	LH occupation wage premium in S	1.40	1.40
Childcare				
β_0	-0.4	Average childcare expenditure	0.1	0.1
k_{min}	4	Average hours of childcare	4.5	4.5
Housing				
H_B	3.2	Average housing expenditure	0.24	0.24
H_S	2.2	Ratio of rents B/S	1.46	1.46

Notes: This table summarizes the values and targeted moments for all calibrated parameters. LH stands for long-hours.

Productivity. I calibrate total factor productivity in each city and occupation to match the different wage premia across cities and occupations. Thus, I normalize the productivity parameter in the linear occupation in the small city, A_2^S, and set A_2^B, A_1^B, and A_1^S such that relative wages are the same in the model and in the data.

Childcare. I set the intercept (β_0) in the function determining the price of childcare to match average childcare expenditure in the US according to Childcare Aware. Moreover, I set the minimum hours of childcare that couples need to buy if they both work (k_{min}) to match average hours of childcare hired by two-earner couples according to Capizzano (2000).

Housing costs. Housing costs in each city are calibrated to match average household expenditure in housing in Davis and Ortalo-Magné (2011) and the ratio of rents across city size that I compute using ACS data.[14]

1.5.4 Non-Targeted Moments: Participation Rates

Given the calibration outlined before, I compute LFP rates for women with children and for women with no children in the model. Table 1.7 displays the participation rate in each city size for each group of women in the data (first row) and in the model (second row). As can be inspected, the model predicts a lower participation rate of women with children and a higher participation rate of women without children in the big city compared to the small city. In addition, the model predicts gaps in participation between small and big cities that are reasonably close to those in the data.

Table 1.7 – Non-Targeted Moments

	Participation Rates (%)					
	Women with children			Women w/o children		
	Big	*Small*	*Gap*	*Big*	*Small*	*Gap*
Data	64.3	68.9	-4.6	81.7	81.0	0.7
Model	64.0	69.1	-5.1	80	76.8	3.2

Notes: This the participation rates of women with children in the data (first row) and in the model (second row) by city size and maternity status. Gaps are calculated by subtracting the participation rate in the small city to the big city.

In the model, wages are higher in the big city, especially for workers in the long-hours occupation. Higher wages induce two effects on female LFP. On the one hand, there is a substitution effect by which a higher wage increases incentives to participate; on the other hand, there is an income effect through the household budget constraint, since the higher husband's wage relaxes this constraint. Which effect dominates depends on the value of the marginal utility of consumption relative to the marginal disutility from labour. Since couples with children experience a higher disutility from productive time due to the time spent with their children, I expect the income effect to dominate for this

[14] I use house price indexes computed by performing hedonic regressions so that differences in the quality and size of housing are taken care of.

group. In addition, given that a larger share of workers in the big city choose the long hours occupation, which is better paid, I expect this effect to be more important in the big city.

Another channel through which big city's comparative advantage in the long hours occupation helps explain the BCCP arises from the fact that males employed in this occupation have less time left to take care of the children, reducing further incentives to participate in the market for women with children. Notice that, given that couples without children do not face any housework, this channel does not operate among them.

Similar to the impact of higher wages, taxation has two implications for female LFP: 1) it reduces husband's earnings, which increases women's incentives to participate; 2) it reduces the payoff from participating. Given the progressivity of taxes and the fact that couples in the US file their taxes jointly, the marginal tax rate that the second earner experiences is high, making the second effect more likely to dominate, and hence discouraging female participation (recall that men always participate). Lastly, notice that this effect will be more important in the big city. As husband's income is larger there, the wife experiences a greater marginal tax rate on the first dollar earnt.

The lower LFP of women with children in big cities is also explained by differences in commuting times. Commuting times act as a fixed cost to participation, hence, the larger they are, the more they will discourage participation. However, their impact on the decision to participate depends on the marginal disutility from productive time, so that more time constrained couples (couples with children) will suffer higher utility costs from commuting times.

Lastly, differences in housing costs across city size also affect the decision to participate. In particular, the higher housing cost born by households in the big city reduces household consumption, which in turn increases incentives to participate. While this effect is present in both couples with children and without children, notice that the former incur the additional cost of childcare if the wife participates. Therefore, and given that the price of childcare is larger in big cities, the positive impact of housing in the likelihood to participate is attenuated for the case of women with children.

1.5.5 The role of sorting

In my model, couples have idiosyncratic preferences for cities, female LFP, and occupations. Cities' characteristics induce sorting of couples with different preferences, which can be thought as types, into each location. Therefore, as couples choose where to live, the resulting distributions of types in each city differ. For instance, couples that have strong preferences for a stay-home parent prefer to live in big cities. There, the husband's salary is larger, and if the wife stays at home, they avoid two big city's disadvantages: the longer female commuting time cost and the more expensive childcare. On the contrary, couples in which both partners prefer to work in the flexible hours occupation find hard to live in a big city since commuting times are longer and both childcare and housing are more expensive there than in the small city.

Measuring the degree of sorting behind the BCCP is particularly useful from a policy perspective. The aggregate LFP in my model is driven by the value of *all* options (understood as combinations of female participation, female occupation, and male occupation) in which the wife does not participate relative to the options in which she works in the market. In contrast, differences in participation across city size are the result of a) cities' differences in the value of staying at home; and b) the value of staying at home relative to other options *within* the city. In the extreme case in which mobility is not allowed, only the latter matters, and differences in LFP highlight how much worse big city's characteristic are for participation, as in this case, sorting is simply not happening at all. In this sense, women who are stuck with long commuting times, more expensive childcare, and husband's long working hours, need to decide whether they would join the labour force taking those as given. However, with geographical mobility, women willing to participate can avoid these hurdles by locating in the small city.[15] In addition, women with low labour attachment, that would not participate regardless of where they live, may locate in big cities because the value of not participating in this city is higher than in the small city. Thus, the fact that these women locate in one city or the other, despite its effect on the BCCP, is not very informative from a policy perspective because their lack of participation is due to their preferences and not to

[15] A small nuance should be added to this statement. Given that couples also have preferences for a given city size, mobility is governed by the migration elasticity between small and big cities. Thus, for couples with strong preferences for the big city, some women may not be able to avoid these hurdles, as they prefer not moving. Moreover, this highlights the role of the overall migration elasticity, λ, in governing the intensity of sorting across city size.

city characteristics.[16] Therefore, the greater the role of sorting in driving the BCCP, the more important is to focus on policies that tackle women's general obstacles to participation, as local policies would most likely only shift people around.

To illustrate this point further, I reproduce here the expression describing the share of households that choose an option o in which the wife does not participate in the labour market (denoted by $j_f = 0$) in a given city s and for a given male occupation j_m (which is a particular case of equation 1.4.2):

$$\pi_{s,j_m,0} = \frac{e^{\frac{V_{s,j_m,0}}{\lambda}}}{V_{j_m,0}} \frac{V_{j_m,0}^{\lambda}}{V} \tag{1.2}$$

where $V_{j_m,0} \equiv e^{\frac{V_{B,j_m,0}}{\lambda}} + e^{\frac{V_{S,j_m,0}}{\lambda}}$ summarizes the value of options in which the woman does not participate and the husband works in occupation j_m, irrespectively of the city, and $V \equiv \sum_{l=1}^{K} \left(e^{\frac{V_{B,j_m,j_f}}{\lambda}} + e^{\frac{V_{S,j_m,j_f}}{\lambda}} \right)^{\lambda}$ captures the aggregate value of all options in the economy. It is clear from the first fraction in this expression that geographical differences in the value of an option result in a higher share of households choosing that option in the city where it is more beneficial. Hence, we can think of this term as reflecting geographical sorting on idiosyncratic preferences. Moreover, from the second fraction of the expression, we can see that the overall value of an option (that is, irrespectively of the city) relative to the rest of options determines the size of the mass of households that choose it. Therefore, this term governs the aggregate female LFP in the economy. To see why, notice that the participation rate of women in city (PR_s) is given by the sum of the shares of households that choose options in which the woman does participate, divided by the whole population of women in city s, fem_s:

$$PR_s = \frac{\pi_{s,1,1} + \pi_{s,2,1} + \pi_{s,1,2} + \pi_{s,2,2}}{\pi_{s,1,0} + \pi_{s,2,0} + \pi_{s,1,1} + \pi_{s,2,1} + \pi_{s,1,2} + \pi_{s,2,2}} = \frac{\pi_{s,1,1} + \pi_{s,2,1} + \pi_{s,1,2} + \pi_{s,2,2}}{fem_s}$$

while the aggregate participation rate (PR) is given by:

$$PR = \frac{\pi_{B,1,1} + \pi_{B,2,1} + \pi_{B,1,2} + \pi_{B,2,2} + \pi_{S,1,1} + \pi_{S,2,1} + \pi_{S,1,2} + \pi_{S,2,2}}{fem_B + fem_S}$$

Therefore, with substantial mobility, geographical differences in the value of

[16] Of course, this is the other extreme. The mass of women that do not participate regardless of their location is reduced as the value of working on the market raises. This is explained more carefully below.

an option mostly affect the gap in participation rates of women in big and small cities but do not have a significant impact in the aggregate participation rate.

In order to establish the importance of sorting on unobservables, I assume that unobserved utilities have the same underlying distribution in each city and suppress geographical mobility.[17] The third row of Table 1.8 shows the resulting LFP rates for each group of women. In this case, the predicted BCCP decreases by 4 p.p. relative to the BCCP in the baseline. This means that sorting on unobservables explains above two thirds of the BCCP in the model. Moreover, unobserved heterogeneity is also crucial among women without children, though it operates in the opposite direction, namely, in this case couples with greater preferences for female participation are those who sort into the big city.

Table 1.8 – No Sorting

Participation Rates (%)					
Panel A: Women with children					
	Total	*Big City*	*Small City*	*Gap*	*% Gap change*
Data	66.6	64.3	68.9	-4.6	-
Model	66.6	64.0	69.1	-5.1	-
No Sorting	66.8	66.3	67.2	-0.9	-82
Panel B: Women no children					
	Total	*Big City*	*Small City*	*Gap*	*% Gap change*
Data	81.4	81.7	81.0	+0.7	-
Model	78.4	80.0	76.8	+3.2	-
No Sorting	78.7	78.2	79.2	+1	-68

Notes: This the participation rates of women with children in the data (first row), in the model (second row), and in the counterfactual scenario in which the distribution of idiosyncratic preferences is the same in both cities. Gaps are calculated by subtracting the participation rate in the small city to the big city.

To understand why sorting on unobservables widens the BCCP is important to understand which household types are attracted to the big city when the location choice is available to them. There are two types of couples that benefit the most from the big city's comparative advantage in the long hours occupation: (i) workers in which both partners prefer to work in this occupation, and (ii)

[17] In practice, the nested logit model for idiosyncratic preferences becomes a logit model in each city by suppressing the bottom nest, which incorporated the city choice, since this is no longer an available choice for households.

families that prefer the husband to work in the long hours occupation while the wife stays at home. In the former case, earnings are relatively high so that the couple can afford big city's expensive childcare. In the latter, couples earnings need not be as high to ensure a satisfactory level of consumption because they avoid the cost of childcare. Since the BCCP becomes lower when geographical sorting is absent, it must be the case that the benefit from locating in the big city is larger for couples in which the wife prefers to stay at home.

1.5.6 Decomposition

In this section, I perform a series of counterfactuals to evaluate the contribution of different features in my framework to the total participation rate and to the BCCP. As explained in detail in the previous section, participation rates at the city level incorporate the effect of sorting on preferences while the aggregate LFP reflects the overall impact of a given factor in female labour supply, which is more interesting from a policy perspective. Nonetheless, looking at changes in geographical differences in female LFP can serve that purpose as well. In particular, the fact that model's predictions regarding the importance of each element in driving the BCCP are in line with the empirical evidence presented in Section 1.3 is reassuring and increases the credibility of the model to predict adequately the response of the aggregate participation rate.

Table 1.9 summarizes the results of this exercise. The first and second rows of this table show the participation rate of each group of women by city size in the data and in the model. The next rows show the result of performing each counterfactual at a time, while the last row shows the result of introducing all changes to the baseline economy simultaneously. Notice that in each exercise the first column displays the aggregate LFP rate in the economy while the second column displays the percentage change in LFP with respect to the baseline. Similarly, the two last columns show the resulting BCCP of that counterfactual simulation and the percentage change in the BCCP as a result of shutting down a given channel in the model. A negative sign means that the BCCP is reduced, that is, the gap in participation between the small and the big city is closing. Moreover, figures above 100 p.p. mean that the gap in participation is reversed, namely, the resulting LFP rate in the big city is higher than in the small city. In what follows, I first describe how I perform each counterfactual, and next discuss the main implications they have for LFP rates.

Table 1.9 – Decomposition

Participation Rates Women with Children (%)						
	Total	% Δ	Big City	Small City	BCCP	% Δ BCCP
Data	66.6	-	64.3	68.9	-4.6	-
Baseline	66.6	-	64.0	69.1	-5.1	-
No commuting cost	68.5	2.9	67.7	69.4	-1.7	-66
No higher return on long hours	68.9	3.5	66.5	71.1	-4.6	-9
Childcare price proportional to wages	68.2	2.4	68.6	67.8	+0.8	-115
No taxes	70.0	5.1	69.8	70.3	-0.5	-90
All	73.1	9.8	74.6	70.7	+3.9	-176

First, I proceed to suppress commuting time costs in both cities, which raises participation rates in both cities. In addition, it makes the big city more attractive, especially for couples in which both partners participate in the labour market. This induces them to locate in the big city, which explains why most of the upsurge in participation rates takes place in that type of city, accounting for a 66% decline in the BCCP.

Second, I suppress the convexity in the long-hours occupation, i.e. I set $\theta = 0$. As endless working hours are no longer highly rewarded, intra-household specialization becomes a worse choice. As a result, aggregate participation increases by 3.5%, husband's childcare time increases, and hours worked decrease. Since this occupation clusters in the big city, the effect of this change is larger for that location, and the BCCP falls. However, and in agreement with the findings in the empirical section, this channel has a very small impact on the BCCP relative to the rest of channels. The reason is that big city's comparative advantage in this occupation is somewhat modest.

Third, I make childcare prices proportional to wages while keeping childcare expenditure constant, that is, I set $\beta_2 = 0$ and calibrate β_0 such that childcare expenditure remains constant. Leaving childcare expenditure unchanged allows me to separate the impact of childcare price differentials across cities from the impact of changes in the size of this expenditure. Notice that the resulting childcare price is necessarily in between the original price of both cities. Consequently, the BCCP reverses, that is, the LFP in the big city overtakes that of the small city. Given that I keep childcare expenditure constant, the small response in the aggregate LFP is to be expected.

Fourth, I eliminate all taxes. As I described in the previous section, progressive taxation and joint filing discourage female's LFP because the first dollar she earns is taxed at the marginal rate of the last dollar earnt by her husband. Eliminating taxes increase LFP rates, especially in the big city, where husband's wage is larger. Moreover, and consistent with the findings in Guner et al. (2012), overall female LFP rates increase substantially. The large increase in the aggregate participation rate highlights that this is an important factor preventing LFP in the big city, since it reflects that the larger participation rate in the big city is not the mere result of couples' sorting.

Finally, in the last row of Table 1.9 I show the resulting LFP rates when I perform all previous exercises at the same time. Shutting down all channels results in an almost 10% increase in female LFP. In addition, the participation gap between small and big cities reverses, that is, the resulting participation rate in the big city exceeds that in the small city by nearly as much as the initial BCCP.

It is clear from this exercise, that, quantitatively, the most important channels fueling the BCCP relate to childcare price differentials and to taxation. However, notice that the average commuting time in a city could have a higher impact than what the model captures. While I restrict the impact of the average commuting time in a city to the two-way commute to work, it is likely that long distances and/or high population density, both of which give rise to the long commuting times, affect other activities, constraining individuals' leisure further and thus discouraging LFP.

1.6 Policy Counterfactuals

In this section I use the model to evaluate the impact of two policy reforms: (i) childcare subsidies and (ii) a reduction in commuting times. A key advantage of this framework is that it allows to evaluate how the implementation of any of these policies in one of the two cities spills over the whole economy and to assess the role of sorting on idiosyncratic preferences in carrying over the impact of a policy reform in a city to the other city. Furthermore, this setup is useful to compare how coordinated policies, that can be thought of policies implemented at the federal level, compare to local policies that affect a single city. Therefore, the following exercise aims at exploiting these features by providing comparisons of the economic impact of implementing childcare subsidies in both cities to the case in which only one city subsidizes this expenditure. Next, I use the model to evaluate the impact of an infrastructure investment that lowers the commuting time in the big city but does not affect the travelling time to work in the small city.

1.6.1 Childcare Subsidies

Couples in which both parents work in the labour market usually need to pay for at least some hours of childcare, as working schedules tend to overlap. As explained in Section 1.4, this is modelled as a minimum amount of childcare, k_{min}, that couples need to buy if they both participate. In this subsection, I evaluate the impact of subsidizing this whole cost, which results in a substantial reduction in the cost of participating in the market. In particular, the size of the subsidy is equal to $p_k^B \min(k^*, k_{min})$, where k^* denotes the household optimal choice regarding market-provided childcare. I restrict access to the subsidy to couples in which both parents participate in the labour market, as these are the couples bearing the minimum childcare cost.

I first examine the impact of giving subsidies to childcare in both cities and then proceed to evaluate the effect of subsidizing childcare only in the big city. In order to finance the cost of these subsidies I introduce an additional (proportional) income tax to balance each city budget.[18]

Table 1.10 displays the results regarding participation rates, average welfare, output per capita, and hours worked in the baseline economy (Panel a), when subsidies are introduced everywhere (Panel b), and when the subsidy is limited to the big city (Panel c). Each row reports the aggregate measure in the economy, the percentage change with respect to the baseline economy, as well as the value of each variable in each of the two cities.

The first finding to be noticed is that childcare subsidies are very effective at raising female participation rates. The aggregate participation rate increases by 9 p.p. when the policy is implemented in both cities and by almost 5 p.p. when only the big city subsidizes childcare. Focusing on geographical differences, we can see that the participation rate in the big city surpasses that of the small city in both cases, pointing towards the sorting of two-earner couples

[18] When I introduce the subsidy in both cities, I need to add a 15.2% income tax in the big city and a 11.8% income tax in the small city, while the income tax needed in the big city when the policy is introduced only there amounts to 19%. These additional taxes ensure that a city's tax income is enough to finance the cost of the subsidy on top of the previous (wasteful) government expenditure.

into the big city and of one-earner couples into the small one. The former are attracted by big city's high productivity, which no longer entails a substantial childcare expenditure, while the latter are dissuaded from locating there due to increased taxation, as the additional income tax required to finance the policy is larger in the big city (because childcare is more expensive there).

Table 1.10 – Policy Counterfactual: Childcare

Panel a: Baseline (no subsidy)				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	66.6	-	64.0	69.1
Average welfare	1.07	-	1.17	0.96
Output p.c.	10.4	-	12.1	8.5
Hours worked	8.4	-	8.5	8.3
Panel b: Subsidy in both cities				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	75.5	+13	77.2	73.7
Average Welfare	1.01	-5	1.08	0.95
Output p.c.	10.0	-4	11.8	8.3
Hours worked	8.1	-4	8.1	8.0
Panel c: Subsidy only in the big city				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	71.2	+7	84.2	57.4
Average Welfare	1	-5	1.01	1
Output p.c.	10.0	-4	11.7	8.6
Hours worked	8.2	-2	8.0	8.5

Notes: This table displays the participation rate, average welfare, output per capita, and average hours worked in the whole economy (column *Total*) and in each of the cities (columns *Big City* and *Small City*). Panel a reports the values of these variables in the baseline economy while Panel b and c do so for the cases in which the childcare subsidy is available in both cities and just in the big city, respectively. The column labelled as *% Change* displays the percentage change of a variable with respect to the baseline economy.

It should also be noticed that one-earner couples are faced with higher taxes even though they do not obtain any benefit from this policy. Therefore, geographical sorting on unobserved preferences is exacerbated when the childcare subsidy takes place only in the big city, as it is evident from the large divergence in cities' participation rates. This calls from some caution when evaluating the impact of local policies. Ignoring that substantial sorting on unobserved preferences occurs after the implementation of the policy, for instance by looking at the impact of the policy only in the city that implemented it or, even more common, comparing that city to surrounding ones, would lead to a large overestimation of its impact. Indeed, although big city's participation rate increases by 20 p.p. with respect to the baseline economy, the surge in the aggregate participation rate is limited to about 5 p.p., so that a large part of the cities' differentials is simply due to people moving around. Table A.9 in Appendix A reports the results of performing these policy counterfactuals in the absence of geographical sorting on unobserved preferences, that is, assuming that the underlying distribution of preferences is the same in each city and banning mobility. In that scenario, the rise in participation is entirely due to the effect that childcare subsidies have at improving the options in which the wife participates and at worsening the options in which she does not. Therefore, implementing this policy in just one city does not have any spillover effect in the other location and we do not observe the large divergence in participation rates that is observed here.

Despite the great effectiveness of childcare subsidies in regard to female labour supply, these policies reduce average welfare and output in the economy. The 5% decline in welfare reported in Table 1.10 is driven by the increased taxation that is needed to sustain the policy. Moreover, recall that in this economy long working hours increase substantially average productivity in occupation 1. Therefore, couples in which the husband works in this occupation and the wife stays at home enjoy the highest utility in both cities, since they make the most out of this technological feature. After the enactment of the policy, however, the utility of options in which the wife stays at home is considerably reduced and the increased utility in the rest of options cannot compensate for it.

More surprisingly, childcare subsidies cause a 4% drop in output per capita. The large increase in female LFP comes along with a decline in the intensive margin, as household's market work becomes more evenly shared among couples. Again, given the convex rate in the transformation of working hours to efficient units of labour in occupation 1, this decline in working hours involves a large output cost. This also explains why output per capita raises in the small city when the childcare subsidy is implemented only in the big city, despite the large drop in female LFP. As a result of the childcare subsidy in the big city, households with preferences for a stay-home wife disproportionately locate in the small city, where husbands' long working hours result in increased average productivity and output per capita. Table A.8 in Appendix A displays the result of implementing childcare subsidies in an economy in which productivity does not raise with working hours in any occupation (that, is, setting $\theta = 0$). Although the impact of childcare subsidies on participation rates is similar to the results displayed in this subsection, the output drop is halved and the loss of average welfare is less pronounced.

1.6.2 Commuting Time Reduction

Commuting times are known to discourage female labour force participation substantially. Therefore, in this subsection I analyze the impact that a plausible reduction in commuting times in the big city would have on the economy. In particular, I set the commuting time cost in the big city equal to that of the small city, which can be thought as an improvement of infrastructure investment in the big city. A shortcoming from doing this exercise arises from the fact that commuting times in my model do not react to changes in population. In contrast, in reality, increases in city population are associated to longer average commuting times due to either higher population density or to an extension of the metropolitan area. Therefore, this counterfactual exercise provides an upper bound to the impact that reducing the commuting time in the big city would have. In other words, if the model were to incorporate the response of commuting times to population changes, a fall in one city's commuting time would be attenuated by the resulting increase in population. Consequently, all other responses in the economy, such as changes in LFP rates, would also be more moderate in such a model than in the results included in this section.

Table 1.11 shows the result from implementing the above-mentioned reform. As Table 1.10, Panel a summarizes participation rates, average welfare, output per capita, and hours worked in the baseline economy, while Panel b reports the same variables when the policy takes place.

A decline in the large city's commuting time has a positive impact on this location's female LFP. Given that commuting times act as fixed costs to participation, there is a direct relationship between these two elements, which explains part of the upsurge. In addition, this policy makes the big city more attractive, especially for couples in which both partners participate in the labour market, inducing them to locate there. This in turn explains the drop in the small city's LFP rate, despite the fact that this city does not experience any change. As argued above, ignoring households' location choices would lead to a wrong assessment of the effects of an infrastructure investment in the big city.

Table 1.11 – Policy Counterfactual: Commuting Times

Panel a: Baseline				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	66.6	-	64.0	69.1
Average welfare	1.07	-	1.17	0.96
Output p.c.	10.4	-	12.1	8.5
Hours worked	8.4	-	8.5	8.3
Panel b: Big City's Commuting Time Reduction				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	66.9	0	65.6	68.3
Average Welfare	1.11	+2	1.22	0.97
Output p.c.	10.6	+2	12.4	8.5
Hours worked	8.5	+1	8.7	8.3

Notes: This table displays the participation rate, average welfare, output per capita, and average hours worked in the whole economy (column *Total*) and in each of the cities (columns *Big City* and *Small City*). Panel a reports the values of these variables in the baseline economy while Panel b does so for the case in which big city's commuting time is reduced due to an infrastracture investment in this city. The column labelled as *% Change* displays the percentage change of a variable with respect to the baseline economy.

Overall, the BCCP halves and total female LFP increases slightly. Turning to the effect of this policy in output, the big city experiences an increase 2% in output per capita, due to the increase in labor supply along both the intensive and extensive margins. Notice that this number leaves some room for financing the investment in infrastructure.

1.7 Heterogeneity in skills

In this section, I delve into the results by accounting for skill heterogeneity. In order to establish which explanations are more important for each skill group, I first repeat the empirical analysis for the groups of low- and high-skilled women separately. I then use an extension of the model that is able to capture the different response of low- and high-skilled women to each feature of the model. Therefore, this extension allows to obtain a further validation of the model. For ease of exposition, in this section I focus exclusively on the results concerning women with children.

Table 1.12 – BCCP by skill

Panel A: Low-Skilled						
	(1) LFP Women Children	(2) LFP Women Children	(3) LFP Women Children	(4) LFP Women Children	(5) LFP Women Children	(6) LFP Women Children
Big City	-0.026***	0.003	-0.025***	-0.016***	-0.022***	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Avg Commute		-0.288***				-0.303***
		(0.015)				(0.017)
% Workers LH			-0.030			-0.022
			(0.021)			(0.023)
Avg Wage				-0.066***		-0.005
				(0.011)		(0.014)
(log) Housing					-0.009	0.022***
Price Index					(0.005)	(0.006)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Spouse	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.068	0.069	0.068	0.068	0.068	0.069
Observations	577559	577559	577559	577209	577559	577209
Panel B: High-Skilled						
	(1) LFP Women Children	(2) LFP Women Children	(3) LFP Women Children	(4) LFP Women Children	(5) LFP Women Children	(6) LFP Women Children
Big City	-0.037***	-0.020***	-0.031***	-0.017***	-0.010*	-0.002
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
Avg Commute		-0.169***				-0.080**
		(0.025)				(0.028)
% Workers LH			-0.189***			-0.042
			(0.032)			(0.036)
Avg Wage				-0.147***		-0.039
				(0.017)		(0.022)
(log) Housing					-0.085***	-0.062***
Price Index					(0.008)	(0.010)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Spouse	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.068	0.069	0.069	0.069	0.069	0.069
Observations	230085	230085	230085	229962	230085	229962

Notes: This table shows the impact of several city characteristics on the BCCP of low-skilled (Panel A) and high-skilled (Panel B) women with children. The dependent variable is a dummy equal to one if the woman participates in the labour market. All columns include year and state fixed effects and control for individual's observables and partner's as in Table 1.1. The sample is restricted to married women with children under 12 who live in metropolitan areas. I also exclude women below 25 or over 55 years old. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

1.7.1 Empirical evidence: BCCP by skill

Accounting for skill heterogeneity in the regressions shows that the explanations for the BCCP vary largely with educational attainment, despite the fact that similar differences in the LFP rates of women with children across city size are present for both low-skilled and high-skilled women. In the sequel, I will consider college graduates as high-skilled women while the rest of women will be classified as low-skilled.

Panels A and B in Table 1.12 are the analogues of Table 1.4 in Section 1.3 for low-skilled and high-skilled women, respectively. Thus, they show how the BCCP changes when city characteristics are accounted for. Comparing both sets of results, it can be observed that, while commuting times are an important determinant of the BCCP for both groups, it is much more relevant for low-skilled women. On the contrary, the BCCP of high-skilled women is heavily influenced by the childcare price (proxied by the average housing cost and the average wage in the city). Moreover, the impact of the concentration of the long hours occupation in the big city on the BCCP is more important among high-skilled women than among the low-skilled.

1.7.2 Model extension

In this version of the model, agents also differ in ability, $z \in U[z_{min}, z_{max}]$. To take into account assortative mating, I draw from the data the proportion of households with different education profiles. In particular, I classify couples in four categories according to their education: (i) both partners are low-skilled, (ii) low-skilled husband and high-skilled wife, (iii) high-skilled husband and low-skilled wife, and (iv) both partners are high-skilled. Therefore, I solve the optimization problem of the household for each of the just described education profiles I just described. I assume that idiosyncratic preferences do not vary across education profiles, that is, idiosyncratic preferences for a stay-home wife will be the same regardless of partners' ability. In reality, I would expect preferences for a stay-home parent to be stronger among low-skilled individuals, as educational investments can be seen as a signal for high preferences for work. Therefore, my model is bound to underpredict gaps in LFP rates across skill levels.

Regarding the relationship between productivity and ability, I assume that ability only increases the productivity of workers in the long-hours occupation. More formally, let n_j denote efficient units of labour in occupation j, h, hours worked, and $f^j(z,h) : N^2 \to N$ be the rate at which hours worked are converted to efficient units. As in the baseline model, I assume that labour services are increasing in the amount of hours worked, at a convex rate in occupation 1 and at a linear rate in occupation 2. The new assumption in the extended model is that idiosyncratic productivity only affects efficient units in the long-hours occupation, $f_z^1 > 0$ and $f_z^2 = 0$, which gives rise to a Roy model of occupational choice whereby the most productive individuals sort into the long-hours occupation. Accordingly, I modify the functional form for the relationship between hours worked and efficient units supplied in occupation 1 to:

$$n_1 = z h_1^{1+\theta}, \qquad \theta > 0,$$

while occupation 2's rate remains unchanged.

1.7.3 Quantitative results by skill levels

Parameters

Table 1.13 displays the values of all non-calibrated parameters in the extended model. Since most of them were discussed earlier, I will only comment on two new ones for this extension of the model.

Demographics. I use 2000 US Census to compute the proportion of couples with children and the degree of assortative mating. In particular I compute the proportion of couples with (without) children by the joint education profile of partners (e.g. the proportion of couples with children in which both partners hold a bachelors degree).

Idiosyncratic Preferences. In this version of the model, I let the variance of the idiosyncratic shock, λ_k, to vary with husband's skill level. Therefore, I use both the college and non-college estimates for the elasticity of the labour supply with respect to the wage in Diamond (2016). As discussed earlier, these estimates identify the variance of the idiosyncratic shock in her model.

Table 1.13 – Non Calibrated Parameters

Parameter	Description	Source	Value
Time			
κ	Time cost of children	ATUS	7/14
τ_B	Avg commuting time in B	2000 US Census	1.1/14
τ_S	Avg commuting time in S	2000 US Census	0.8/14
T	Total available time (hours)	Aguiar and Hurst (2007)	14
Occupations			
α	1-Profit Share	Barkai (2016)	0.85
θ	Degree of convexity	Erosa et al. (2017)	0.6
Demographics			
s_κ	% couples with children	2000 US Census	0.7
$s_{LSLS\|\kappa>0}$	% both LS | children	2000 US Census	0.6
$s_{LSHS\|\kappa>0}$	% LS husband-HS wife | children	2000 US Census	0.13
$s_{HSLS\|\kappa>0}$	% HS husband-LS wife | children	2000 US Census	0.14
$s_{HSHS\|\kappa>0}$	% both HS | children	2000 US Census	0.13
$s_{LSLS\|\kappa>0}$	% both LS | no children	2000 US Census	0.64
$s_{LSHS\|\kappa>0}$	% LS husband-HS wife | no children	2000 US Census	0.11
$s_{HSLS\|\kappa>0}$	% HS husband-LS wife | no children	2000 US Census	0.12
$s_{HSHS\|\kappa>0}$	% both HS | no children	2000 US Census	0.13
Preferences			
λ_{LS}	Variance idiosyncratic taste shock, LS	Diamond (2016)	0.25
λ_{HS}	Variance idiosyncratic taste shock, HS	Diamond (2016)	0.47
γ	Intertemporal elasticity labour supply	Domeij and Floden (2006)	0.4
Price of Childcare			
β_1	Price of childcare: Housing	ACS & Childcare Aware	0.25
β_2	Price of childcare: Wage	ACS & Childcare Aware	1.7
Taxes			
t_p	Payroll tax rate	Guner et al. (2012)	0.086
$t_{1,\kappa=0}$	Tax function no children	Guner et al. (2012)	0.113
$t_{2,\kappa=0}$	Tax function no children	Guner et al. (2012)	0.073
$t_{1,\kappa>0}$	Tax function children	Guner et al. (2012)	0.084
$t_{2,\kappa>0}$	Tax function children	Guner et al. (2012)	0.09

Notes: This table summarizes the values and sources for all non calibrated parameters. HS (LS) denotes high-skilled (low-skilled) which are defined as those who hold at least a bachelor degree.

Calibration

Table 1.14 summarizes the calibration exercise when skills are included in the model. Column (2) shows the resulting parameter values while column (3) presents the targeted moment. Next, columns (4) and (5) display the targeted moments in the data and in the model, respectively. Although the displayed set of figures includes all moments targeted in the previous calibration exercise, there are two parameters which are specific to this extension of the model, namely, the productivities of high- and low-skilled individuals (z_{min} and z_{max}). Since in the model productivity is only useful in the long-hours occupation, I calibrate both parameters using moments related to this occupation. Specifically, I match (i) the percentage of low-skilled workers employed in the long hours occupation and (ii) the skill premium within that occupation.

Table 1.14 – Calibrated Parameters

Parameter	Value	Targeted Moment	Data	Model
Preferences				
ψ	0.00013	Average hours worked	8.37	8.35
ν_m	0.08	Male average hours with children	1	1
ν_f	0.11	Female average hours with children	1.6	1.6
Productivity				
A_2^B	1.1	LH occupation wage premium in B	1.30	1.30
A_1^B	1.35	Big city wage premium in LH occupation	1.60	1.59
A_1^S	1.05	LH occupation wage premium in S	1.40	1.40
z_{min}	0.35	% LS workers in LH occupation	44.2	45.5
z_{max}	0.55	Skill premium in LH occupation	1.44	1.44
Childcare				
β_0	-0.4	Average childcare expenditure	0.1	0.1
k_{min}	4	Average hours of childcare	4.5	4.5
Housing				
H_B	3.2	Average housing expenditure	0.24	0.24
H_S	2.2	Ratio of rents B/S	1.46	1.46

Notes: This table summarizes the values and targeted moments for all calibrated parameters. HS (LS) denotes high-skilled (low-skilled) who are defined as those who hold at least a bachelors degree. LH denotes long-hours.

Non-Targeted Moments

The first two rows of each panel in Table 1.15 display the average LFP rate in each city size for each group of women in the data and in the model, respectively. Panel A of the table displays the results for low-skilled women while Panel B focuses on high-skilled women. The model does a good job at predicting the LFP rate of low-skilled women with children and the size of the BCCP for both groups. As expected, the LFP rates predicted for high-skilled women are too low compared to those observed in the data, since I do not allow for correlation between ability and work preferences.

Table 1.15 – Non-Targeted Moments

Participation Rates Women with Children (%)				
	Big	*Small*	*BCCP*	*% BCCP change*
Panel A: Low-skilled women				
Data	62.1	67.4	- 5.3	
Baseline	62.5	68.1	-5.6	
No Sorting	64.8	66.4	-1.6	-71
Panel B: High-skilled women				
Data	72.3	76.0	-3.7	
Baseline	68.4	71.7	-3.3	
No Sorting	70.0	70.3	-0.3	-90

Notes: This table displays the participation rate of women with children in the data and in the model. Panel A focuses on low-skilled women while Panel B focuses on high-skilled women.

The third row in each panel of Table 1.15 shows the participation rates resulting from suppressing across city size differences in the distribution of idiosyncratic preferences. While geographical sorting on unobservables reduces considerably the BCCP of both skill groups, it is more important among the high-skilled women with children, for whom sorting accounts for almost the totality of the gap in participation.

Decomposition

In order to assess skill differences in the relative impact of each channel on the BCCP, I repeat the decomposition exercise summarized in Table 1.9. Panels A and B of Table 1.16 display the results of the decomposition exercise for low- and high-skilled women, respectively. The first and second rows of this table show the LFP rate of women with children in each city size in the data and in the model. The next rows show the result of performing each counterfactual at a time, while the last row presents the results of introducing all changes to the baseline economy at the same time. A careful explanation about how each counterfactual is performed can be found in Section 1.5.6 above. Notice that, in each exercise, the column BCCP shows the resulting BCCP in that counterfactual scenario, while the last column shows the percentage change in the BCCP with respect to the baseline model.[19]

Table 1.16 – Decomposition

	Part. Rate (%)			
	Big City	Small City	BCCP	% BCCP change
Panel A: Low-skilled women				
Data	62.1	67.4	-5.3	-
Baseline	62.5	68.1	-5.6	-
No commuting costs	66.2	68.9	-2.7	-52
No higher return on long hours	65.6	69.9	-4.3	-23
Childcare price proportional to wages	67.0	67.5	+0.5	-109
No taxes	68.6	70.1	-1.5	-73
All	74.6	72.0	+2.6	-146
Panel B: High-skilled women				
Data	72.3	76.0	-3.7	-
Baseline	68.4	71.7	-3.3	-
No commuting costs	71.2	72.0	-0.8	-75
No higher return on long hours	71.6	73.1	-1.5	-45
Childcare price proportional to wages	72.0	70.9	+1.1	-133
No taxes	72.5	72.9	+0.4	-112
All	77.6	74.0	+3.6	-209

Consistent with the empirical findings, the return on working long hours plays a greater role in explaining the BCCP among high-skilled individuals. Moreover, the results of this decomposition exercise show that the LFP of high-skilled women with children is more affected by proportional taxation and the price of childcare than that of low-skilled women. This should come as no surprise since assortative mating implies that husband's income is larger on average among the high-skilled, especially in the large city. As a consequence, the marginal utility of household's consumption is lower and women face a greater marginal tax rate when considering to enter the labour market.

1.8 Conclusions

Despite decades of convergence, gender gaps in LFP remain substantial. In this paper, I exploited variation in female LFP across city size to investigate the role of city characteristics at curtailing female labour supply.

I showed that the BCCP is mostly driven by geographical sorting on unobservables. Higher wages, longer commuting times, and more expensive childcare

[19] Notice that since the BCCP is negative, a positive percentage change corresponds to reduction of the BCCP, while a negative percentage change means that the gap is reversed, that is, that the participation rate in the big city surpasses the participation rate in the small city.

in big cities as compared to small ones set the incentives for intra-household specialization, so that women with low labour attachment disproportionately locate in big cities. Moreover, my framework allows me to draw some conclusions regarding aggregate LFP. While big city's specialization in occupations with a high reward to working long hours has a small impact on the BCCP, technologies in which productivity increases substantially with hours worked are an important deterrent of female labour force participation.

Policy counterfactuals revealed a large effectiveness of childcare subsidies in raising female LFP and some gains from coordinated implementation as opposed to local application. However, they also brought out an interesting trade-off between a more equal division of market work and aggregate output. As high-skilled occupations become increasingly specialized, worker's productivity grows substantially with hours worked. This, in turn, favour a higher degree of intra-household specialization such that long hours rewards materialize and aggregate productivity and welfare raise. Nevertheless, in the presence of gender roles, which partner specializes in market work and which partner in home production is not determined by their relative productivity in each task, giving rise to substantial gender inequality and to labour misallocation. Therefore, a promising avenue for future research relates to the relationship between the long hours occupations, gender roles, and labour misallocation.

I also found that infrastructure investments that shorten commuting times impact positively that location's LFP, welfare, and, output. In order to get more accurate predictions on the impact of the policies considered in this paper, some congestion externalities should be considered. First, commuting times are known to be a function of the population of the city, since more populated cities are necessarily more extensive or denser. Second, it would be interesting to allow the price of housing respond to the cities' demand for housing, especially since it also affects the cost of childcare in the city.

In light of the findings of related literature, I believe future research should also consider the impact of adding agglomeration externalities to this framework, possibly biased towards high-skilled workers. This feature of the model would allow for gains in efficiency from the reallocation of working couples from the small (less efficient) city to the big (more efficient) city.

Chapter 2

Delayed Birth and Gentrification

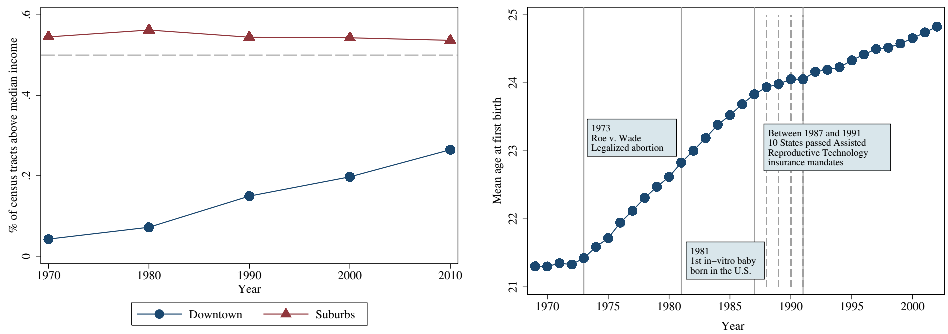
joint work with Clara Santamaria

2.1 Introduction

Over the last decades, the increasing gentrification of downtown neighborhoods in many developed countries has attracted the attention of researchers and policy-makers.[1] While the influx of high-income individuals to low-income central neighborhoods may unfold a range of benefits to incumbent residents, such as amenity improvements or crime reduction, it certainly raises important concerns. Among them, one that stands out is the parallel increase of housing costs in the neighborhood, which may effectively displace low-income residents and thus become an important source of inequality for at least two reasons. First, given that displaced residents face higher traveling time to consumption amenities, displacement affects consumption inequality within the city. Second, and depending on the spatial distribution of jobs, displacement may erode their labour market opportunities by substantially increasing commuting times to high-wage occupations.

[1] In line with most of the economic literature, we define gentrification as the influx of households with higher socioeconomic status to central city neighborhoods, that often results in the displacement of earlier and less affluent residents (David et al., 2017). This definition accounts for several important dimensions, being usually characterized by increased numbers of prosperous and educated individuals living in downtown urban locations, the renovation and refurbishment of houses, and rising rents and house prices.

Although the consequences of gentrification have drawn considerable attention in public policy debates, the identification of the underlying factors behind this growing phenomenon remains controversial, while their understanding is still fairly limited. Recent studies outlined some potential determinants of gentrification, such as the rising valuation of downtown amenities (Couture and Handbury (2017), Couture et al. (2019)), the reduction in downtown crime rates (Ellen et al., 2019), and the increased opportunity cost of long commuting times (Edlund et al. (2015), Su (2018)). Nonetheless, in this paper we take a different perspective of this issue by focusing on the remarkable cultural and socioeconomic transformations that the role of women has undergone in society during recent decades, and which have modified traditional lifestyles and family structures. In particular, we realise that fertility decisions are inextricably linked to residential location choices and propose delayed childbearing as a novel driver of the gentrification process.



(a) Gentrification **(b)** Age at first birth

Figure 2.1 – Delayed fertility and gentrification

We first notice that the decades of rapid gentrification are precisely those in which a steady increase in the age of first-time mothers has taken place. Figure 2.1 shows that from 1970 to 2010 the probability that a census tract in the city center exhibited an average income above the median income in the city went from 5 percent to 25 percent (panel 2.1a). During the same period, the average age of first mother went from just over 21 to close to 25 years old (panel 2.1b). Although, in principle, the causal relationship between these two trends, if any, could go in either direction, our goal in this paper is to identify a plausible causal effect of the delay in fertility on gentrification. We do so by exploiting exogenous policy changes at the state level that decreased the cost of postponing childbearing, and document that the implementation of such policies has important effects on the demographic composition and gentrification of US cities.

Moreover, we show that, over our period of study (1970-2010), families are much more likely to locate in the suburbs of the city than other household types. We argue that families' location patterns are driven by the proximity to better schools and to larger housing size in the suburbs as compared to downtown locations. In contrast, downtowns offer a higher density of consumption amenities, such as restaurants, bars, cinemas and theatres, which are typically more valued among young individuals, especially by those enjoying higher income and living in households without children. Therefore, as technological advances enable women to postpone childbearing, the life period in which downtown amenity consumption is highest is extended, enhancing the incentives to locate in the center among this group of people.

Despite the fact that delayed parenthood is common to all developed countries, postponement can be very costly, since fertility decays sharply with age. In this context, technological advances in Assisted Reproductive Techniques (ART) offer some insurance against the risk of infertility associated to late childbearing. However, given the high cost of these medical treatments, their access remained quite limited. Our empirical strategy exploits state variation in the price of ART induced by infertility insurance mandates. In the late 80's, several US states enacted ART insurance mandates, which in practice implied a substantial reduction of the price of ART treatments that couples faced. This both in a large rise in the access rate to ART and in an increase in the average age at first birth in those states (Hamilton and McManus (2012), Buckles (2005), Schmidt (2007)). Therefore, this policy provides a nice scenario to assess the impact of delayed parenthood in the gentrification of downtown neighborhoods. Admittedly, postponed maternity is a much broader phenomenon which it is certainly not limited to the states that enacted infertility mandates. In this sense, we believe that our results are more general than simply measuring the impact of the mandates on gentrification, since they are also useful to understand the interaction between demographic change and neighborhood development.

Using a difference-in-difference-in-differences (triple difference) approach, we find that the existence of a state mandate to cover ART leads to gentrification. Downtown income relative to the suburbs increases by 5.4% more in treated cities than in cities that belong to the control group, and downtown neighborhoods are 11.3 percentage points (p.p.) more likely to be above the median city income. Moreover, the larger average income of residents in the city center goes in parallel with a demographic change that is consistent with gentrification. Specifically, the share of college graduates in downtown neighborhoods belonging to treated cities increases by around 2 p.p. both relative to the suburbs and the non-treated cities. In addition, the age distribution of women also changed in the expected direction, with an increase of 2 p.p. of women between the ages of 25 and 29 and a subsequent decrease of 2 p.p. among those between 30 and 35 years old. The age distribution of men reacts similarly but lagged by a few years, likely due to male partners being slightly older. We argue that these changes in the age composition downtown are fully consistent with couples postponing childbearing and moving to the suburbs.

The rest of the paper is organized as follows. Section 2.2 reviews the related literature. Section 2.3 provides the background about the implementation of the infertility treatment mandates that we exploit for the causal identification. Section 2.4 presents the data. Section 2.5 introduces the empirical specification and the main results on the causal effect of mandated ART treatments on gentrification while Section 2.6 gathers the results about changes on demographic composition of the city center and the location patterns of families. Section 2.7 provides a quantification of the impact of the overall increase in women's age at first birth on gentrification. Finally, Section 2.8 concludes.

2.2 Related Literature

First and foremost, our paper relates to a growing literature that analyzes the causes of downtown gentrification. Baum-Snow and Hartley (2019) points out that the propensity of young and high-income individuals to live in the city center is largely driven by two factors: (i) divergent preferences towards downtown amenities between different racial groups, and (ii) the rising suburban concentration of labor market opportunities for low-education workers. Likewise, Couture and Handbury (2017) also emphasize the role of amenity valuations, arguing that increases in gentrification in the 2000-2010 period can be explained by a growing taste for downtown amenities among college graduates. On the other hand, Edlund et al. (2015) argue that longer hours worked among high-skilled workers have increased their distaste for commuting, which in turn has pushed up house and rental prices in the city center. Similarly, Su (2018) examines the growing importance of long work hours in well-paid downtown-located jobs as an exogenous factor driving the demand for central locations by high-skilled workers. Couture et al. (2019) evaluate the impact of top-income growth and its associated rise in income inequality on the location choices of rich households. In order to quantify the welfare consequences of urban gentrification, they introduce idiosyncratic preferences shocks and endogenous amenities to a spatial model of urban sorting. Recent work by Almagro and Dominguez-Iino (2019), Curci and Yousaf (2020) and Hoelzlein (2019) also study how endogenous amenities reinforce sorting by income within cities. We contribute to this literature by outlining a novel important channel that leads to gentrification: delayed parenthood.

Moreover, our work speaks to the literature on women's timing of family formation. Goldin and Katz (2002) and Bailey (2006) examine the impact of the availability of the birth control pill on birth, marriage timing, and female labour supply. Goldin and Katz (2002) show that greater access to the pill reduces the likelihood of marrying before age 23 and therefore increases the likelihood of women being employed in professional and high-skilled occupations. In a similar vein, Bailey (2006) pinpoints that reducing the age at which it becomes legal to access to the pill reduces the likelihood of a first birth before age 22 and increases labor supply on both the intensive and extensive margins. Postponing childbearing may benefit women for several reasons. Caucutt et al. (2002) show that fertility delay is related to changes in marriage and labour markets. Thus, high-skilled women delay marriage and fertility in order to obtain a better match, even in the absence of returns to labour market experience. Moreover, when labour market experience is taken into account, fertility is delayed even further. Using biological fertility shocks to instrument for motherhood delay, Miller (2011) finds that postponing motherhood has a statistically significant and positive impact on earnings and career paths, particularly for the highly educated women. Our contribution to this strand of the literature is therefore to highlight a new set of consequences of delayed maternity, those related to neighborhood development and within-city inequality.

Several studies have confirmed the effectiveness of infertility insurance mandates on increasing ART utilization. Hamilton and McManus (2012) find that mandates to cover ART lead to a substantial increase in the usage of these technologies in the market. Moreover, they show that variations in the insurance regulations of states are largely due to different general political preferences rather than to unobserved preferences for ART. Similarly, Jain et al. (2002) find that states with required coverage for In Vitro Fertilization (IVF) - the most effective and most widely used form of ART - have the highest rates of IVF utilization. While these works provide suggestive evidence on how the mandates have increased IVF usage, they do not control for unobservable differences in patients or clinics that may be state-specific. This gap is filled by Bitler and Schmidt (2006), who use a difference-in-differences approach to show that the sizable increase in the use of infertility treatments as a result of the ART mandates is mainly concentrated among highly- educated older women, with no significant impact on other socioeconomic groups.

In addition, there is another stream of literature which has estimated the causal impact of infertility insurance mandates on several outcomes that are relevant to our framework. First, Schmidt (2007) finds a significant increase on first birth rates for women over 35. Consistent with this, Machado and Sanz-de Galdeano (2015) report a positive association between the mandates and women's mean age at first birth. However, they show that fertility rates over women's reproductive lives are unaffected by these mandates. Second, Buckles (2005) encounters that mandates that cover IVF are associated with an increase in labour force participation and earnings for women under 35 and a reduction in participation for older women. Similarly, Abramowitz (2017) points to an increase in women's age at marriage and at first birth after the enactment of these mandates, though only for college graduate women. Third, Kroeger and La Mattina (2017) find that such mandates led to a rise the probability that women hold a professional college degree and work in professional occupations. Our contribution to this literature is to uncover another consequence that had not been previously considered, namely, their effect on the spatial distribution of income within cities.

2.3 Infertility insurance mandates

While several studies have found positive effects of delaying childbearing on women's lifetime earnings (Buckles 2008; Caucutt et al. 2002; Miller 2011; Wilde et al. 2010), it is well known that fertility decays sharply with age (Menken et al. 1986; van Noord-Zaadstra et al. 1991). In particular, the probability of having a successful conception within one year after starting to try to conceive is 75% for women at age 30, while it declines to 66% and 44% at ages 35 and 40 years, respectively (Leridon, 2004). Thus, by enabling women to postpone childbearing, ARTs may relax the career-family trade-off that women often face.

However, ART treatments (specially IVF) are very expensive. According to Hamilton and McManus (2012), one cycle of IVF entails an out-of-pocket cost of \$10,000 to \$15,000 to the patient and it is common to attempt multiple cycles of treatment. Moreover, it is rare that insurers cover these costs unless required by law.

Starting at the end of the 80's, several US states enacted mandates to enhance ART access. In practice, infertility insurance mandates amount to a significant reduction of the price borne by the patient, which is expected to increase utilization by making it affordable to a broader segment of the female population. Although several studies have shown that access to ART remained mostly limited to high-skilled white women (Bitler and Schmidt 2006; Hamilton and McManus 2012), they also notice that these mandates may affect younger women's decisions without necessarily increasing their own utilization of ART afterwards. The reason is that infertility insurance mandates affect the expected value of delaying childbearing: by lowering the cost of ART treatments, they reduce the risk associated with infertility at older ages. On top of that, they may have increased awareness about the availability of IVF and consequently changed women's misconceptions about its effectiveness. Lastly, increased IVF usage may have reduced the stigma associated to marrying and having children at an older age for the whole population of women.

Table 2.1 lists all states that have enacted mandates affecting the insurance of ART procedures over the five decades covering our census samples (1970-2010) and summarises their main features. There are several sources of heterogeneity across state mandates. First, while most states require insurers to cover ARTs treatments in every available insurance policy, mandates in California and Texas only require insurers to offer infertility treatments. In addition, not all mandates include IVF treatments nor affect every type of insurance provider. In particular, some mandates exclude health maintenance organizations (HMOs) while others only target HMOs.

As shown by Hamilton and McManus (2012), this heterogeneity is very relevant. These authors document that "universal mandates" (those requiring all insurers to cover ART) lead to a substantial increase in IVF utilization while other types of insurance mandates have a smaller effect. Consistent with this, studies focusing on the impact of the mandates on different outcomes (see, inter alia, Kroeger and La Mattina (2017), Machado and Sanz-de

Galdeano (2015) or Schmidt (2007)) have found larger effects in states with universal mandates. Therefore, we only include in the treatment group those states that enacted mandates to cover IVF treatment and that applied to all insurers. This means that our group of treated states includes the following ones: Connecticut, Hawaii, Illinois, Maryland, Massachusetts, New Jersey, and Rhode Island.

Table 2.1 – States with mandated infertility insurance

State	**Date enacted**	**Mandate to cover**	**Mandate to offer**	**IVF coverage**	**Type of insurers**	**Treated**
Arkansas	1987	X		X	HMOs excluded	
California	1989		X		All	
Connecticut	1989	X		X	All	X
Hawaii	1987	X		X	All	X
Illinois	1991	X		X	All	
Louisiana	2001	X			All	
Maryland	1985	X		X	All	X
Massachusetts	1987	X		X	All	X
Montana	1987	X			HMOs only	
New Jersey	2001	X		X	All	
New York	1990	X			HMOs excluded	
Ohio	1991	X		X	HMOs only	
Rhode Island	1989	X		X	All	X
Texas	1987		X	X	All	
West Virginia	1977	X			HMOs only	

Notes: This table summarizes the main features of acts mandating infertility insurance in all states that ever passed a mandate of this type. HMOs refers to Health Maintenance Organizations. The column treated displays which states we consider as part of the treatment.

In addition, we eliminate variation in the timing in which the mandates were enacted by pooling together all states that passed reforms between 1980 and 1990, which excludes New Jersey and Illinois from the treated group.[2] As a result of this rule, we are left with five treated states which are listed for convenience in the last column in Table 2.1.

Lastly, there are some US metropolitan areas which belong to several states, such as Boston. In these cases, we consider that a city is treated if at least some part of the metropolitan area belongs to a state in our treated group. The rationale for this choice is that we think it is likely that residents in parts

[2] In our analysis, we use census data because it allows us to identify neighborhoods' location. However, since these data are only available every ten years, we include 1970 and 1980 in the pre-treatment period, and consider 1990, 2000, and 2010, as being part of the post-treatment period.

of the metropolitan area belonging to other states were also affected by the policy, as metropolitan statistical areas have a high degree of economic and social integration. Regarding the control group, it is composed of all states that never enacted any kind of infertility insurance mandates or did it after the 90's.[3]

Finally, our identification strategy requires that assignment to the treatment is exogenous, that is, that the enactment of the mandates in some states did not respond to a greater demand for infertility insurance by the population. As mentioned earlier, these concerns have been addressed in Bitler and Schmidt (2012) and Hamilton and McManus (2012). Both studies show that state differences in the enactment were due to the electorate's view toward mandates in general. More concretely, Hamilton and McManus (2012) show that states that enacted mandates regarding other health issues (such as colorectal cancer screenings, Medicaid funding of abortions, and mental health parity) also adopted regulations for IVF. Thus, it is reasonable to think that state adoption of infertility insurance mandates was due to residents' preferences regarding government intervention in healthcare markets as opposed to a larger demand for infertility insurance on its own. In addition, these authors found no pre-mandates differences across mandate and non-mandate states in ART intensity, measured by the number of clinics in the state or the number of treatments per 10,000 women aged 25-44 years.

2.4 Data

Our analysis is conducted at the census tract level, defined as small geographical units encompassing between 2,500 and 8,000 people, which provides a good approximation for our definition of neighborhoods. We combine decennial Census data and the American Community Survey (ACS) 2008-2012, downloaded from the National Historical Geographic Information System (NHGIS), and construct constant 2010 census tract boundaries using the Longitudinal Tract Data Base (LTDB).

Our definition of city is the Core-Based Statistical Areas (CBSA) constructed by the Census Bureau. Given that gentrification is a big city phenomenon (Hwang and Lin (2016)), we restrict our sample to neighborhoods located in

[3]That is, we drop from our sample states that had a reform but are not in the treated group (all states included in Table 2.1 that are not marked as treated in the last column).

metropolitan areas with more than 1 million inhabitants.[4] The sample size includes 82,129 census track- census year observations (51,469 in the treated group and 30,660 in the control group).

In line with the previous literature, we normalize distance to the city center using the cumulative share of the metropolitan population who lives in the nearest locations. That is, we consider rings of population around the city center such that a given share of population is included; e.g. a distance equal to 0.3 includes the area of the city including the 30 percent of population that is the closest to the city center. In particular, we use data from Lee and Lin (2018) to locate the geographical center of each CBSA and define the city center as the area within 0.1 distance from it. The main advantage of this definition is its flexibility as compared to geographical distances, since it adjusts for the fact that downtowns are generally more extensive in larger metropolitan areas. Similarly, we define the suburbs as the area of the city that contains the 50 percent of population that lives the furthest away from the city center.

Several measures of gentrification are used in this paper. First, we use the probability that the income in a specific census tract is above the median income in the city. This measure provides a good metric to describe gentrification processes, as it captures the income in that particular area relative to the median income in the entire city. However, it potentially misses changes in income at the tails of the distribution. This pitfall is overcome by our second measure, the (log) median income in the census tract. Third, we use the percentage of college graduates in the neighborhood, as gentrification is mostly driven by high-skilled individuals (Couture and Handbury, 2017).

Table 2.2 summarizes the main descriptive statistics of cities in treated and non-treated states before the mandates were introduced. As can be observed, average city size was considerably larger in treated states. Although both groups of states include a similar number of large cities, non-treated states host a greater number of small cities, which results on a large difference in means. It is also noteworthy that, while cities in treated states were richer both in terms of household income and average housing value, their city centers were poorer than those in the control group. Given that differences in city income and population across groups were substantial at the time of the reform, we

[4] In the Appendix, we relax this restriction by replicating our main results for a sample which includes all cities that have more than 100,000 inhabitants, and obtain similar estimates.

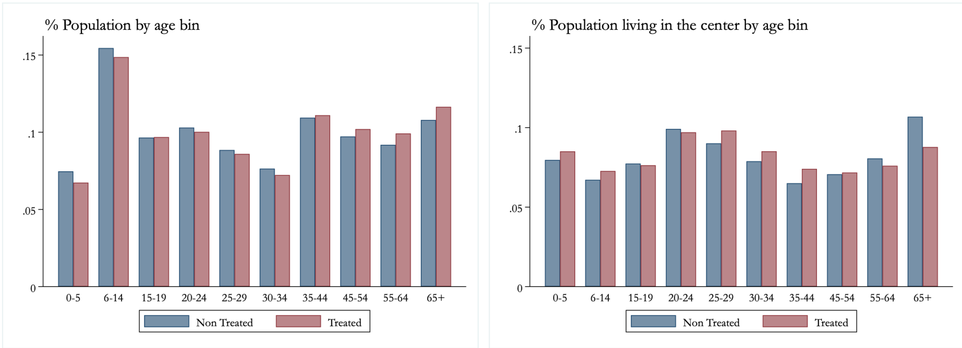
control for these characteristics in all our regression specifications.

Table 2.2 – Summary Statistics before Mandates

	Non Treated	Treated	Std. Diff
Avg City Population	2,766,916	6,798,305	-0.99
	(372,144)	(1,900,509)	
Avg City Household Income	21,431	22,332	-0.35
	(460)	(908)	
Avg City Housing Value	55,047	58,506	-0.27
	(2,325)	(4,335)	
% College Graduate in the City	10	10	-0.11
	(1)	(1)	
Downtown Household Income	15,298	14,658	0.29
	(582)	(753)	
Downtown Housing Value	41,113	40,810	0.02
	(3,872)	(7,272)	
Downtown % College Graduate	9	7	0.38
	(1)	(2)	
Number of observations	51,469	30,660	.
	(.)	(.)	

Notes: This table displays city averages regarding some relevant characteristics in treated and non-treated states in 1980. Standard errors are in parenthesis. The last column shows standardized mean differences for each reported variable.

Lastly, Figure 2.2 shows that the average age distribution before the reform was quite similar across treated and non-treated states. This can be seen in Panel 2.2a, which displays the percentage of population by age bin in treated and non-treated cities prior to the mandates. In addition, both groups of states exhibited a similar spatial distribution of individuals for a given age group. Panel 2.2b in Figure 2.2 documents the absence of statistically significant differences in the percentage of individuals that live downtown within each age group in 1980 between treated and non-treated cities.



(a) City **(b)** Downtown

Figure 2.2 – Age Distribution before Mandates

Notes: This figure displays the age distribution of treated and non-treated cities in 1980, before mandates were enacted. Panel 2.2a displays the percentage of population in treated and non-treated cities by age bin. Panel 2.2b shows the percentage of individuals that locate downtown within each age bin.

2.5 Econometric specification and main results

In this section, we explore the impact of adopting insurance mandates on standard measures of gentrification, i.e., changes in income and the percentage of college graduates in downtown neighborhoods. As already mentioned, we are interested in understanding whether the lower cost of postponing childbirth influences the faster income growth at the city center and on the location patterns of high-skilled individuals. If the cost of living downtown increases with the presence of small kids, delaying childbirth would allow couples to live downtown for a longer period. Therefore, postponing both childbirth and moving to the suburbs would lead to a higher presence in downtown areas of young married couples on a more advanced stage in their professional career paths, therefore implying a higher average income of the residents in the city center. This demographic group not only increases median income at the city center, but their preferences also contribute to endogenously increasing the supply of amenities, making room for further waves of gentrification. Moreover, as the group of people who prefer to live downtown expands, the demand for downtown housing rises, leading to higher sorting of individuals on income, since more wealthy households can afford higher rents.

2.5.1 Triple-Difference Specification

In order to estimate the causal effect of insurance mandates on gentrification, we employ a triple difference specification. The first difference is taken between the pre-treatment and the post-treatment period. As explained in Section 2.3, we consider that the post-treatment period starts in 1990 for all treated states. The second difference is taken between treated and non-treated states. It captures how different was the change in the variable of interest in census tracts that were treated versus those that were not treated between the pre-treatment and the post-treatment periods. The third difference is taken between being part of the city center or of the suburbs. Hence, this triple difference captures how different was the change in the outcome variable of interest: (i) before and after treatment date, (ii) between the city center, and (iii), between the suburbs in the treated states compared to non-treated states. The general form of the regression we run is:

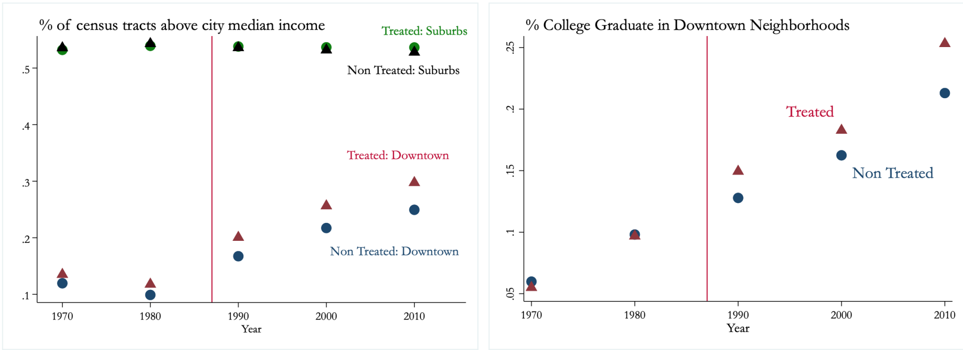
$$\begin{aligned} y_{i,t} = \alpha &+ \beta_1 Treated_i + \beta_2 Post_t + \beta_3 Center_i + \beta_4 Treated_i \times Post_t \\ &+ \beta_5 Treated_i \times Center_i + \beta_6 Post_t \times Center_i \\ &+ \beta_7 Treated_i \times Post_t \times Center_i + \beta_8 X_{i,t} \\ &+ \phi_{State(i)} + \delta_{State(i)} \times t + \psi_{CitySize(i)} + \gamma_{CitySize(i)} \times t + \epsilon_{i,t}, \end{aligned} \tag{2.1}$$

where $y_{i,t}$ is the outcome of interest for a census tract i at time period t. There are three indicator variables: $Treated$, which takes value one for those states which enacted an insurance mandate between 1980 and 1990; $Post$, which takes value one for periods after 1980, both for treated and non-treated states; and $Center$, which takes value one if the census tract is within the radius around the city center which contains 10 percent of the population of the city, $X_{i,t}$ includes controls, which vary at the census tract and year level. Finally, we include state and city size fixed effects (ϕ and ψ), as well as state and city size time trends (δ and γ).

Recall that throughout the analysis we only keep observations in cities with more than one million inhabitants, since gentrification is a large-city phenomenon. Moreover, we control for city size and time trends by city-size category to ensure that the results are not driven by a different presence of large vs small cities in the treated states.

As mentioned earlier, recent studies have pointed out to changes in the spatial distribution of jobs and to growing income as the main drivers of gentrification (Couture and Handbury (2017), Couture et al. (2019), Edlund et al. (2015), Su (2018)). Notice that these trends are widespread across all US cities and, hence, should not affect our estimation. That is, in principle, there are no reasons to think that cities that are located in states that enacted infertility insurance mandates experienced larger changes in the spatial distribution of jobs or greater income growth, except as a consequence of the policy itself. Nevertheless, we include the (log) median income of the city and the share of jobs within 3 miles distance from the census tract as additional controls ($X_{i,t}$) in our regression to ensure that our estimates are not driven by alternative explanations in the literature. Further, we control for the share of college graduates in the city, which is very related to gentrification and could have been increasing faster in treated cities for reasons unrelated to the policy. An important concern when controlling for these variables is that the policy itself may have affected them directly, leading to problems of endogeneity. For instance, postponing the arrival of children is associated with positive effects on female wages, which in turn could affect average income in the city. While this is likely to have occurred, we believe these effects are second order and do not pose a challenge to our estimates. Some support for this conjecture is provided by the results reported in Table B.2 in Appendix B, where it is shown that excluding these controls from our regressions hardly changes our main findings.

Lastly, our identification strategy relies on the existence of parallel trends in the outcomes of interest before mandates were introduced. Figure 2.3 displays the evolution of income by area of the city and the location patterns of college-graduates within the city in treated and non-treated cities over our period of study. Panel 2.3a displays the percentage of census tract whose income is above median income in the city both downtown and in the suburbs. Panel 2.3b shows the percentage of college graduates in downtown neighborhoods. . Both panels support the existence of parallel trends before the dates of the mandates which, together with the exogenous policy enactments, allows us to interpret the estimates in a causal fashion.



(a) Income **(b)** College Graduates

Figure 2.3 – Parallel Trends

Notes: The left panel of this figure displays the percentage of census tract in each area of the city with income above median city income in treated and non-treated cities over time. The right panel of this figure displays the percentage of the city college-graduates that live in the center of the city. The red line signals the time in which infertility insurance mandates were enacted.

2.5.2 Impact of infertility insurance mandates on gentrification

We start by analyzing the effect of the insurance policy mandate on income at the city center. As explained in Section 2.4 above, two complementary outcome variables are being used in this respect: (i)) the probability that a census tract's income is above the median household income in the city and; (ii) the (logged) average income in the census tract. The first two columns in Table 2.3 display the results of running the triple differenced specification in equation 2.1 for each of the two above-mentioned variables. In what follows, we discuss the main findings for each row in Table 2.3 which correspond to the different interactions at play.

The first row shows that census tracts that are located in the city center had lower income levels than the suburbs *before* the treatment took place, both for treated and non-treated cities. Tracts in the city center are 42.6 percentage points (p.p.) less likely to be above the median income and have on average 62.2 percent lower income.

The second row documents that census tracts outside the city center in treated states also became more affluent *after* treatment. Following the policy implementation, median income in these tracts was 6.8 percent higher while the probability of being above the median in the city increased in 5.6 p.p., which implies that the policy also led to favorable effects on income in the suburbs. This effect is in line with the findings in Kroeger and La Mattina (2017), who show that infertility mandates increased the percentage of women that entered professional occupations in treated states as compared to non-treated states, irrespective of whether they lived downtown or in the suburbs. Moreover, it is consistent with the literature that documents positive effects in women's wages from postponing maternity (Caucutt et al. (2002), Goldin and Katz (2002), Miller (2011), among others).

Table 2.3 – The effect of infertility insurance mandates on gentrification

	Prob. above median	Log median income	% College Graduate
	(1)	(2)	(3)
Center	-0.426***	-0.622***	-0.000912
	(0.0132)	(0.0104)	(0.00298)
Treated × Post	0.0557**	0.0677***	0.0278***
	(0.0270)	(0.0211)	(0.00609)
Center × Treated	-0.136***	-0.0376**	-0.00230
	(0.0187)	(0.0146)	(0.00421)
Center × Post	0.0611***	0.0370***	0.0113***
	(0.0161)	(0.0126)	(0.00363)
Treated × Center × Post	0.113***	0.0536***	0.0216***
	(0.0234)	(0.0183)	(0.00526)
Observations	82129	82129	82129
Controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
City Size FE	Yes	Yes	Yes
State Trends	Yes	Yes	Yes
City Size Trends	Yes	Yes	Yes

Notes: This table displays the impact of infertility insurance mandates on several measures of gentrification: (1) the probability that a census tract's income is above median income in the city; (2) the census tract's (log) median income; and (3) the percentage of college graduates in a census tract. Controls include city's population, city's (log) median income, the share of college graduates in the city, and the share of jobs within 3 miles distance from the census tract. This table reports only selected coefficients, the full specification can be found in equation 2.1. Standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$.

The third row reveals that census tracts in the city center (relative to the suburbs) in treated states were poorer than their analogous counterparts in non-treated states. Nevertheless, we do not find that this difference is important for our results.

In the fourth row we see that income in the city center also increased after the treatment in non-treated states. Downtown locations became 6.1 percentage points more likely to be above the median income of the city, and on average this meant a 3.7 percent rise in median income. This is most likely due to a general trend towards gentrification to different degrees everywhere. First, notice that the trend to postpone childbearing is broader than the delay induced by this policy and certainly not restricted to states that enacted infertility insurance mandates (as was illustrated in Figure 2.1). Therefore, the observed widespread growth in downtown income is fully compatible with the mechanism we propose in this paper. In addition, we think that our mechanism is compatible with other drivers of gentrification such as the changing spatial distribution of labor market opportunities or the rising importance of long working hours in high-skilled jobs.

The fifth row implies that center tracts in treated states had an even larger increase in income during the post treatment period compared to the increase in income experienced in non-treated states. In particular, center tracts' income became 11.3 percentage points more likely to be above the median income in the city, beyond the 6.1 percentage point increase of center tracts in non-treated states. This implied an average increase in downtown income of 5.3 percent on top of the 3.7 percent increase in non-treated states.

Next, we switch attention to another commonly used measure of gentrification: the percentage of college graduates that locate downtown. As reported in the third column in Table 2.3, the fraction of college graduates in a neighborhood: (i) was not significantly different in each area of the cities; (ii) increased 2.8 p.p. in treated cities after the mandates were enacted; (iii) was not different in the center of treated cities; and, (iv) had a positive trend downtown everywhere. More crucial to our analysis, college graduates were 2.2 p.p. more likely to locate downtown after the mandates in treated cities as compared to the same difference in non-treated cities. Therefore, the results regarding college graduates' location patterns are fully in line with the observed changes in income downtown, as expected.

To summarize our findings, the effect of the policy on income at the city center and the college graduates' location patterns is statistically significant and sizable in magnitude. As discussed in subsection 2.5.1, treated and non-treated states were on parallel trends leading up to the treatment year. Moreover, it seems reasonable to assume that states differences in the enactment of mandates were politically motivated, instead of being due to some underlying variable that could have increased income at the city center faster than in the suburbs. For this reason, we believe there are sufficiently strong grounds for interpreting the estimated coefficients in a causal fashion.

2.6 Additional Evidence

In this section, we first proceed to analyse the variation in the demographic composition of the city center supports our preferred mechanism whereby the effect of the policy on gentrification could result from women delaying having kids and staying downtown rather than moving to the suburbs. Next, we document the higher propensity of families to locate in the suburbs of the city with respect to non-families and highlight some suburban characteristics that make these areas more suitable for families.

2.6.1 Changes in the demographic composition

We claim that infertility insurance mandates extended the life period in which individuals benefit the most from locating downtown, fueling the process of gentrification. Therefore, we should observe a change in the demographic composition of central neighborhoods towards slightly older couples.

In order to capture this change, we restrict our attention to individuals with ages between 20 and 44 and examine their location choices. We focus on couples in childbearing age because these are the ones for which the timing of family formation influences their residential choices. Therefore, we run again equation 2.1 where this time the dependent variable is the percentage of individuals in each age bin of the census tract population who are in childbearing age (20-44). Figure **??** plots the coefficients of the triple interaction term, which displays the impact of the policy for each of our 4 age bins (20-24, 25-29, 30-34, and 35-39), both for males and females. That is, this figure displays the impact of ART mandates in the age distribution downtown relative to the suburbs in treated cities and compared to the same difference in non-treated cities.

Consistent with the idea that postponing childbearing allows couples to reside in the city center until later stages of their lifetimes, we find that the policy leads to around a 2 percentage points increase in the proportion of adults aged between 20 and 30 living downtown, while in parallel the percentage of older adults goes down. Interestingly, the proportion of women postponing maternity is a bit younger than men. This is consistent with women delaying having kids until the early thirties and moving out of the city center with their partners and kids afterwards. Since, male partners tend to be a little older, the effect is delayed for men. In line with our mechanism, the proportions of men and women aged between 35 and 40 living downtown decrease by around 2 percentage points.

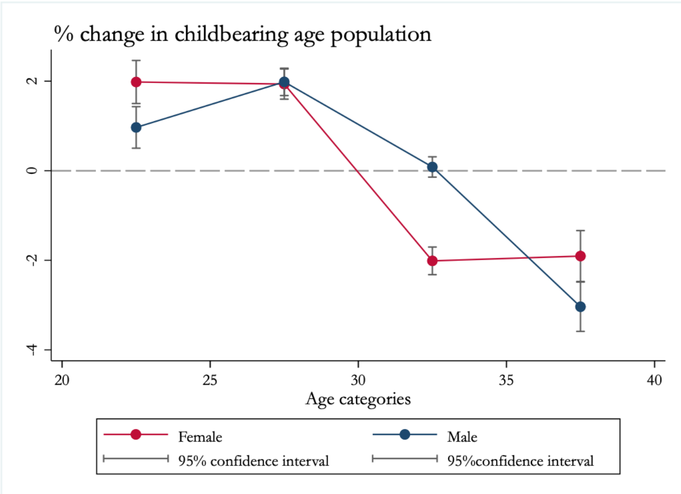


Figure 2.4 – Age composition males vs females

Notes: This figure shows the change in the age distribution of female and male individuals in downtown neighborhoods as compared to the suburbs in treated cities vs non-treated cities. That is, it plots the coefficients of the triple difference of running equation 2.1 for the percentage of females or males in four different age categories.

2.6.2 The location patterns of families

Table 2.4 shows the result of running an analogous triple difference regression to the one considered earlier, this time using the percentage of families in a given neighborhood as the dependent variable. The only change with respect to equation 2.1 is that we now place the focus of the subsequent discussion on the suburbs (rather than the center) as the reference location category. Both specifications are equivalent, we only choose the one presented in Table 2.4 for ease of exposition.

Table 2.4 – Families' location patterns

	% Families in the census tract
	(1)
Suburbs	0.227***
	(0.00321)
Treated	-0.00202
	(0.00653)
Post	-0.212***
	(0.0224)
Treated × Post	-0.00285
	(0.00798)
Suburbs × Treated	-0.00714
	(0.00465)
Suburbs × Post	-0.0485***
	(0.00367)
Treated × Suburbs × Post	0.0321***
	(0.00560)
Observations	73387
Controls	Yes
Year FE	Yes
State FE	Yes
City Size FE	Yes
State Trends	Yes
City Size Trends	Yes

Notes: This table shows the effect of the policy in the location patterns of families. The dependent variable is the percentage of family households in a census tract. Controls include: city's population and city's (log) median income, and the share of jobs within 3 miles distance from the census tract. This table reports only selected coefficients, the full specification can be found in equation 2.1, replacing the dummy for center for the complementary dummy for suburbs. Standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$.

The first row in Table 2.4 shows that the percentage of families in a suburban neighborhood is 22.7 p.p. larger than downtown, while the third row reflects the substantial decline in fertility and family formation after the 80's. Notice that the estimated coefficients on the indicator variable for the treatment and its interaction with the post-treatment period and with the suburbs dummy are all non-significant. This indicates that there were no differences between treated and non-treated cities in the overall percentage of families before and after the mandates nor in their location patterns.

Turning our attention to the last two rows of Table 2.4, we observe that while the percentage of families in suburban neighborhoods decreased by almost 5 p.p. in non-treated cities after the 80's, this decline was limited to about 1 p.p. in treated cities. The decreasing trend in the percentage of families that live in the suburbs could be due to the end of suburbanization, meaning that downtowns are increasingly providing some of the living facilities that families used to find in the suburbs. Moreover, it could be that at least some families are changing their location patterns as downtowns become more attractive. Lastly, the fact that this change is considerably lower in treated cities points towards some difficulties to locate in thriving downtowns among families in these locations, as treated cities experienced gentrification to a larger extent.

Therefore, the evidence reported in Table 2.4 illustrates several patterns that are relevant for our analysis. First, it shows that families are much more likely to locate in the suburbs than non-family households over the last decades. Second, it confirms a severe fall in family formation, which we consider to be a key element in explaining the gentrification process. Lastly, it highlights that the shift in the location patterns of families towards downtown neighborhoods is more contained in cities experiencing a higher degree of gentrification.

One could think of two main reasons explaining why families disproportionately locate in the suburbs. First, the housing stock downtown is not ideal for the children. For instance, houses are too small and usually lack outdoor space. Second, school quality is known to be worse in central neighborhoods. Indeed, in Appendix B we show that houses tend to have a larger number of rooms in the suburbs than downtown. In addition, we include some illustrative evidence about differences in school quality between the suburbs and the center in some cities. Notice that, even if children could attend schools in any location in the city, there are clear advantages of attending nearby schools, as many extracurricular activities are closely linked to the school and other children are also likely to live close by.

2.7 Delayed childbearing and gentrification: an IV approach

So far, we have shown that differences in infertility insurance mandates across US states had a significant effect on the gentrification of downtown neighborhoods. Moreover, we have provided evidence consistent with the postponement of maternity mediating the effect of these policies on the spatial distribution of income. However, the rise in the age at first-time mother is a much more general trend that is not solely related to ART mandates. As illustrated in Figure 2.1 above, starting in the late 70's, the age at first birth has gone from just over 21 to close to 25 years old in 2010. Over the same period, the probability that a census tract in the city center had an average income above the median income in the city went from 5% to 25%. To relate both trends in a causal fashion, in this section we use the mandates as an instrumental variable for the average age at first birth in the city when estimating the impact of the latter variable on gentrification. In addition, we provide some preliminary quantitative assessment of the overall effect that a delayed age at first birth could have on the relative income growth of city centers and suburbs over the sample period under consideration.

The choice of an IV approach in this exercise is dictated by the following reasoning. As pointed out above, downtown neighborhoods tend to be wealthier in cities in which women have their first kid at an older age. However, the direction of causality is unclear. In particular, it could be the case that as gentrification gets stronger (because central areas of the city become more attractive due to shorter commuting times or increased density of amenities), women reacted by postponing having children and moving to the suburbs, leading in this way to reverse causality. Thus, in order to estimate the causal effect of age at first birth on gentrification, we use the ART policy enactment to instrument the average age at first birth in a city, on the basis that the approval of these policies across different states is unrelated to the specific preferences of their populations about delayed fertility treatments (see discussion on this issue in Section 2.3 above). The identifying assumption in this empirical strategy is that the mandates affected gentrification only by affecting the age at first birth, but not directly.

Our specification requires that we run our regressions at the city level instead of using census tract as done in our previous analysis. We obtain the age of first-time mothers at the county level from the National Center for Health Statistics (NCHS) Natality Birth Data. We then construct a measure of gentrification at the city level by dividing the average income in central counties by the average income in suburban counties. Therefore, an increasing income ratio will be indicative of gentrification happening in the city.

$$Gentrification_{i,t} \equiv \frac{\sum_{j \in Downtown} MeanIncome_j}{N_{downtown}} / \frac{\sum_{j \in Suburbs} MeanIncome_j}{N_{suburbs}} \tag{2.2}$$

Since the instrument we employ is a binary instrument, we use the Wald estimator, also known as the grouping estimator. The estimator is implemented through the following three steps. First, we regress our measure of gentrification city i at time t on the instrument and controls:

$$\begin{aligned} Gentrification_{i,t} =& \alpha_0 + \alpha_1 Treated_i + \alpha_2 Treated_i \times Post_t + CitySize_i \\ & + \mu_t + \phi_{State(i)} + \delta_{State(i)} \times t + \epsilon_{i,t}, \end{aligned} \tag{2.3}$$

where μ and ϕ are time and state fixed effects, and δ are state trends and we also control for city size. Next, we run a similar regression for the average age of mothers at their first birth in city i at time t:

$$\begin{aligned} AgeFirstBirth_{i,t} =& \beta_0 + \beta_1 Treated_i + \beta_2 Treated_i \times Post_t + CitySize_i \\ & + \mu_t + \phi_{State(i)} + \delta_{State(i)} \times t + \nu_{i,t}, \end{aligned} \tag{2.4}$$

where the time, state, and state trends are denoted with the same symbols as before for comparability. Finally, we combine both estimates together to obtain the Wald estimator, which captures the effect of age at first birth on gentrification, instrumented with the insurance mandate.

$$\widehat{W} = \frac{\hat{\alpha}_2}{\hat{\beta}_2}, \qquad SE_{\widehat{W}} = \frac{\hat{\alpha}_2}{\hat{\beta}_2} \sqrt{\left(\frac{SE_{\hat{\alpha}_2}}{\hat{\alpha}_2}\right)^2 + \left(\frac{SE_{\hat{\beta}_2}}{\hat{\beta}_2}\right)^2} \tag{2.5}$$

The results of the first and second step are included in Table 2.5. We find that the policy increased the age at first birth by 0.62 years, and the ratio of downtown to suburb income goes up by 1.5 percentage points. For reference, the average ratio of downtown to suburb income in our sample is 57 percent with a standard deviation of 15 percentage points. Moreover, the average mean age at first birth is 24 years with a standard deviation of 2 years. If the exclusion restriction holds, this implies that for each year of increase in the average age at first birth, cities should expect the ratio of income downtown to income in the suburbs to increase by 2.4 percentage points.

Table 2.5 – Causal effect of Age at 1st Birth on Gentrification

	First step	
	Gentrification	Age at 1st birth
	(1)	(2)
Treated × Post	0.015***	0.618***
	(0.0021)	(0.0255)
Observations	72737	72737
Year FE	Yes	Yes
State FE	Yes	Yes
State time trends	Yes	Yes
City Size FE	Yes	Yes
	Second step	
	Gentrification	
	(1)	
Age at 1st birth	0.024***	
(IV: Treated × Post)	(0.0035)	

Notes: This table displays the impact of delayed maternity on gentrification using a Wald estimator. The top panel displays the results of the first step regressions while the bottom panel displays the result of the second step regression. Standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$.

The magnitude of the estimated effect points towards a potentially large economic significance. A back-of-the-envelope calculation tells us that the increase in the age at first birth from 23.27 in 1980 to 24.7 in 2000 could explain an increase in the income ratio of downtown to suburb of 5.8 percentage points. The average increase in the ratio from 1980 to 2000 was 3.7 percentage points. Of course, there are many other mechanisms working at the same time and we do not claim that all of the increase in age at first birth is exogenous, nor that it is the sole driver of gentrification. However, these results are suggestive of the potential central role that a delay in the age at first birth may have played in explaining gentrification.

2.8 Conclusions

In the US, forming a family and having kids is associated with couples moving to the suburbs, where housing is larger and schools are better. However, more and more, young couples are choosing to postpone both fertility and the move to the suburbs. This has been made possible by medical advances in infertility treatments that allow couples to delay childbirth into the 30s without much risk. As couples stay downtown longer, precisely at a time when their incomes are growing fast, they become gentrifiers of their downtown neighborhoods by increasing the demand for amenities such as bars, movie theaters, and restaurants.

This paper provides causal evidence on the importance of delaying fertility on gentrification by exploiting state-level variation in the enactment of policies that essentially decreased the cost of delaying maternity. We find that these policies had a direct and statistically significant effect on the income growth of downtown vs. the suburbs which took place in parallel with a demographic change in the city center consistent with postponing the arrival of children and suburban life.

Recent work on gentrification has highlighted the responsiveness of downtown amenities to changes on the demographic composition of surrounding neighborhoods, which reinforces this process (Almagro and Dominguez-Iino (2019), Curci and Yousaf (2020)). We consider that this is an important avenue to explore in our context, as it could lead to additional incentives to postpone the arrival of children, amplifying the initial effect of the policy. The idea is that, as some couples decide to rely on the eventual utilization of ART treatments and extend the life period in which they live downtown, these areas become more attractive due to the endogenous response of amenities. This, in turn, induce more couples to postpone childbearing and moving to the suburbs, fueling further waves of gentrification. Hence, we believe that a dynamic model of fertility and within-city location would be very useful to account for the general equilibrium effects of delaying maternity, improving our understanding of this matter. Furthermore, such a model would allow us to learn about the effects of delayed parenthood and subsequent gentrification on welfare and inequality, which are central from a policy perspective.

Another promising area for future research that is related to the findings in this paper concerns the spatial distribution of female labour force participation. On the one hand, postponing maternity is associated to increases in wages and to less costly career interruptions, which should raise female labour supply. On the other hand, gentrification affects the location choices of individuals and their commuting times to work, which are known to discourage substantially the labour force participation of women. A general equilibrium model that incorporated both channels would be useful to understand which effect prevails and, more generally, to evaluate the impact on gentrification on labour supply.

Chapter 3

Free Trade and Labour Mobility

joint work with Mathijs Janssen

3.1 Introduction

Technological advances in transportation and communications have led to an increasingly globalised world whose benefits and disadvantages are central to all countries' political agenda. Not only international trade has increased dramatically but also international migration have been on the rise, which raises the question on whether the two phenomena do interact.

Focusing on the European Union, it is noteworthy that labour motivated migration flows from the less advanced periphery to the more advanced core, have experienced a change in their skill composition. While in the sixties migrants were mostly low-skilled, most of the current migration flows between these countries are high-skilled, coinciding with a greater degree of integration due to declines in both trade costs and migration costs. Motivated by these observations we build a theoretical framework that allows to study these issues.

In particular, we develop a two-country, two-sector general equilibrium model and compare equilibrium outcomes for different costs of trade and migration. While the secular reduction of trading costs has received a lot of attention, mainly in the new economic geography literature (Krugman, 1991), to our knowledge the impact of the fall in the cost of migration over time and its interaction with trading costs have not yet been carefully analysed.

In our model, we consider two sectors: *high-tech* and *low-tech*. Each sector employs a single factor of production: *educated* workers in the high-tech sector and *non-educated* workers in the low-tech one. We assume that productivity differences are limited to the high-tech sector and refer to the country with comparative advantage in this sector as the most advanced country. In other words, we are assuming that non-qualified workers, such as waiters, are equally productive in each location while qualified workers, such as engineers, are more productive in the country with a more advanced technology. Workers are identical in each location. We initially take education levels to be fixed, but later relax this assumption and allow workers to choose to become educated by paying a cost that is heterogeneous across agents. Thus, those with an education cost below a given threshold become educated while those above remain non-educated.

Our findings reveal that trade integration leads to a change in the composition of migration flows toward high-skilled labour. In particular, we show that while price equalization, induced by trade integration, reduces incentives to migrate for low-skilled workes, it increases the return to migration for high-skilled workers in technologically disadvantaged countries.

More concretely, we find that trade integration increases emigration of those workers that work in the production of goods that are imported, while it decreases emigration of those workers that work in the production of goods that are exported. Under natural additional assumptions, the technologically least advanced country imports the high-tech good, so that trade integration leads to *brain drain*, the outflow of high-skilled workers from the least advanced country to the most advanced country. In the jargon of the trade and factor mobility literature, trade and educated-migration become complements (as in the Ricardian model), while trade and non-educated migration become substitutes (as in the Hecksher-Ohlin model). To the best of our knowledge, this is novel result, which follows from our realistic assumption that differences in technology only exist in the production of the high-tech good. Free trade equalizes the reward of the labour input in the sector for which there are no technological differences, i.e., the real wage of non-educated workers becomes identical. However, the same argument does not hold in the high-tech sector, as educated individuals are more productive in the most advanced country when producing the high-tech good. Hence, in the sector where technological differences are present, free trade is not enough to equalize the reward of labour, and hence factor mobility is needed to exhaust differences in real wages.

Indeed, this is a good representation of the EU, where the core is more advanced in high-tech sectors while the periphery specializes in low-tech sectors, such as tourism and low-skilled personal services. Moreover, introducing skill heterogeneity among workers, we find that trade integration leads to outflows of the most talented workers among educated workers in the least advanced country, a phenomenon that we label “brain skimming”.

Given the established positive relationship between trade and brain drain, we find interesting to explore the impact of trade and migration costs on the educational attainment of each country’s population. One way to think about these results has to do with time horizons. Since in the short run, a country’s stock of human capital does not respond immediately to changes in trade and migration costs, except for the induced brain drain, the model with exogenous education is a good approximation of the impact of trade liberalization in this time horizon. However, given that individuals do respond to changes in the incentives to get education, a country’s educational attainment does change with declines in trade and migration costs in the long run. Therefore, we endogeneize the worker’s decision to get education and simulate the economy for a given technological gap between the two countries. While these results are far less general than the previous results, they illustrate interesting avenues for future research. According to our simulation, patterns of sectoral specialization are reinforced by the endogenous human capital investments made by workers. That is, incentives to become educated in the most (least) advanced country increase (decrease) when the two countries trade freely, for any cost of migration.

Lastly, we find that, despite gains in efficiency arising from free trade and migration, average welfare of residents in the least advanced country is not always higher under free trade and without migration costs (i.e., free labour mobility). The reason is that, in our model, free trade is the main determinant of brain drain which lowers the average income in that country, since educated workers enjoy higher salaries. In contrast to the sending country, average welfare of residents in the receiving country is highest when both trade and migration are totally free. This should come as no surprise as they experience the flip side of the coin: the arrival of highly productive educated migrants increases the average income in the country.

The rest of the paper is organized as follows: in the next section, we review existing literature on this topic. In Section 3.3, we introduce a general model with exogenous education. In Section 3.4, we discuss the impact of trade on migration according to this model. In Section 3.5, we introduce the endogenous choice of education and provide some insights about the impact of trade on migration by skill. In Section 3.6 we make some parametric assumptions to investigate the relationship between trade, migration, educational attainment, and welfare. Finally, Section 3.7 concludes.

3.2 Related Literature

Our paper is related to the literature on trade and migration. The main contribution in this literature is the generalization of models of factor proportions (i.e., Hecksher-Ohlin models) to account for differences in technology across countries (i.e., Ricardian models). In the former models, in which technology is assumed to be common across countries, differences in factor endowments are the only source of trade. As a consequence, trade and factor mobility become substitutes. In contrast, in Ricardian models, factor endowments are assumed to be identical and it is the different technology across countries that stimulates trade. In this context, Markusen and Svensson (1985) and Wong (1986) among others (see Wong and Grinols (1995) for a comprehensive overview) show that, with differences in technology, trade and factor mobility are complements. Recent availability of data on migration has facilitated testing the complementarity/substitutability hypothesis between trade and migration. However, the evidence is mixed (see Campaniello (2014) for a review of the empirical literature).

Regarding the growing relevance of high skilled migration, which is fully in line with the findings in this paper, Kerr et al. (2016) show that not only the ratio of high- to low-skilled migrants is increasing worldwide but also the number of countries of origin is broadening while the set of destinations is narrowing. Docquier and Rapoport (2012) confirm the increasing importance of the brain drain and point out its higher incidence among middle-income countries in which migrants' financial constraints are less binding than in low-income countries. Even though brain drain is less severe in developed countries, Docquier et al. (2009) report a net outflow of European high-skilled towards other developed destinations, the most important being the US. Tritah (2008) attributes this phenomenom to the weakness of demand for skilled labour in Europe. In addition, there is a high selectivity among emigrants, being those most qualified the most likely to migrate, as Akcigit et al. (2016) recently find for the case of inventors.

Our paper also speaks to the literature that evaluate the impact of brain drain. This literature has identified different channels by which the sending country can lose or gain from skilled emigration. Early literature established the neutrality of brain drain in source countries such as Grubel and Scott (1966) or Berry and Soligo (1969). However, a second wave of literature pointed to losses derived from domestic labour market rigidities and fiscal externalities (Bhagwati and Hamada (1974) and McCulloch and Yellen (1977)). A later trend in the literature highlighted some channels by which the sending country would benefit from skilled emigration such as networks and the diffusion of technology (Kerr (2008) and Agrawal et al. (2011)), global gains in efficiency (McAusland and Kuhn, 2011), return migration (Mayr and Peri, 2009) and the possibility of *brain gain* (Beine et al., 2001), that is, that the prospect of migration fosters the acquisition of skills in the origin country.[1] Our welfare evaluation incorporates the well-known output loss that takes place as the most talented workers leave the country. However, we find an additional benefit from brain drain when it is studied in conjunction with free trade, namely, that a more efficient global labour allocation leads to lower prices that benefit everyone everywhere. We show that for some specific parameters this gain may overweigh the output loss.

Our main contribution is to give a simple yet rich framework that allows us to study the consequences of changes in both trading and migrating costs in the skill composition of migration flows. To the best of our knowledge, we are the first to analyze the connection between free trade and brain drain. In a closely related paper, Iranzo and Peri (2009) study the joint impact of trade costs and migration costs on real wages and specialization. They build a two country, two sector model in which countries differ in TFP and individuals are heterogeneous in skills. Assuming that the differentiated sector employs a more skill-biased technology and that the more advanced country has comparative advantage in this sector, they show that a decrease in trading and/or migration costs intensifies the pattern of specialization and increases the welfare of most individuals, where the last effect is caused by the increase in varieties. They

[1] However, the propect of migration could also lead to the possibility of *brain waste*, that is, that the skills that individuals acquired because of the prospect do not match the needs of the origin country (Di Maria and Stryszowski, 2009) or that even when they succeed in migrating, their skills end up not used in the destination country (Mattoo et al., 2008)

calibrate the model to Western-Eastern European data to study the impact of trade liberalization and the posterior removal of legal barriers to labor mobility. Contrary to our findings, they show that the high-skilled migrate more irrespective of trading costs. The reason is that, in their model, the degree of substitutability between both sectors in consumption is higher. Thus, while in our model as high-skilled emigrate in the absence of trade the skill premium rises, reducing therefore the incentives to emigrate, in their model the skill premium is not affected as consumers simply substitute away from the high-tech sector. In addition, we differ from them in that we model explicitly the choice of education.

3.3 Model with exogenous education

There are two countries indexed by $j \in \{A, B\}$; and two goods: a high-tech good H, and a low-tech good, L. Countries only differ on how efficient they are at producing H, which is measured by a productivity parameter R_j, such that $R_A > R_B$. Thus country A is technologically more advanced than country B. Labour is the only input used in production and could be of two types: *high-skilled* labour h required to produce the H good, and *low-skilled* labour l required to produce the L good. In this section we will take the skill level of workers as given. We will relax this assumption later on. For each country, there is a mass of high-skilled workers M_H and a mass of low-skilled workers M_L.

Countries can trade with each other, but the trade is costly. In particular, we model trade costs as an iceberg cost t, a fraction of goods that is lost in international trade.

Finally, workers can choose to migrate to the other country. If they do, they incur a single utility cost $\kappa \geq 0$, which is specific to each worker. Thus κ has a distribution in the population, whose CDF we denote with F: $F(k) := \Pr(\kappa \leq k)$ if we randomly draw a worker from the population. κ is independent from worker type, i.e. $\Pr(\kappa \leq k) = \Pr(\kappa \leq k|$ worker is high-skilled)

We study a competitive, price taking equilibrium.

3.3.1 Firms

There is a high tech and a low tech firm in each country. Production technology is as follows:

$$Y_A^L = f_L(l_A) \qquad Y_A^H = R_A f_H(h_A)$$
$$Y_B^L = f_L(l_B) \qquad Y_B^H = R_B f_H(h_B)$$

f_L and f_H are given increasing and concave functions. Firms maximize profits, taking wages w_A^L etc. and prices p_A^L etc. as given.

3.3.2 Workers

Workers derive utility from consumption of the two goods in the economy according to $U(x_H, x_L)$, some standard utility function. We will introduce assumption on U as we proceed, to highlight where they are needed. Workers derive income from selling their labour endowment of 1, which they supply inelastically. Thus they have a budget set $B(w, p)$ that depends on the wage they receive and the prices they are facing, $B(w, p) := \{(x_H, x_L) : x_H p_H + x_L p_L \leq w, \}$.

When workers choose whether to migrate, they compare their prospective utility in each country, taking into account their utility cost κ of migrating. With slight abuse of notation, write $U(w, p) = \max_{x \in B(w,p)} U(x)$ for the indirect utility function. Then a worker of type i from j with migration cost κ migrates if $U(w_i^j, p^j) < U(w_i^{\neg j}, p^{\neg j}) - \kappa$. The total number of migrating workers of type i from country j to country $\neg j$ is then $F(U(w_i^{\neg j}, p^{\neg j}) - U(w_i^j, p^j))$.

3.3.3 Trade

As explained earlier, international trade is costly. If a quantity Q is exported, only tQ arrives, $0 \leq t \leq 1$. Thus, t is an inverse index of transportation costs and when we say that trade costs decrease, we mean an increase in t. We define trade-autarky as the situation in which $t = 0$, as trade is impossible, while $t = 1$ corresponds to full trade integration or free trade, in which international trade does not bear any cost. The reason why we distinguish between autarky and trade-autarky is that international migration may still take place even when $t = 1$.

Moreover, since trade involves shipping goods both ways, in trade a quantity t^2 survives. Hence, if a firm has a choice of selling a quantity Q at home for price p or abroad for price p', he will sell abroad if $Qp < t^2Qp'$.

We normalise the prices of the low-tech good in each country to 1 and write p_A, p_B for the prices of the high-tech good.

3.4 The Effect of Trade on Migration

In this section we show that trade integration, modeled as a reduction in trade costs and hence an increase in t, leads to increased migration by those workers who produce goods that are imported and decreased migration by those workers who produce goods that are exported. We then strengthen our assumptions on preferences to argue that country B exports the low-tech good. We can then specify the previous result trade integration leads to a "brain drain" from B, where, in the extreme case where $t = 1$ and there is fully free trade, only high skilled workers emigrate from B.

For the first proposition, we make one strong assumption on preferences and two innocuous ones. The strong assumption is that we take preferences and hence aggregate demand in a country to exhibit the gross substitution property, i.e. an increase in the price of one good leads to an increase in the demand for the other good. While this assumption is restrictive, it makes sense in our context, where the two goods represent very aggregated commodities. A sufficient condition for the gross substitution property, much stronger than needed, is that the utility function is additively separable. While we make the assumption to ensure that price convergence is well behaved as trade costs decrease, it also has the benefit of ensuring uniqueness of equilibrium.

We also make the weaker assumptions that preferences are continuous, so that aggregate demand is continuous in prices, and that they satisfy the Inada condition, so that positive amounts of both goods will be consumed at any price.

Proposition 1 (Trade integration leads to specialised migration.)**.** *Suppose B is a net exporter of good $i \in \{H, L\}$ at a given level of t. Then a marginal increase in t will lead to non-increased emigration of workers from B who* *produce i and non-decreased emigration of workers from B who produce $\neg i$.*

The proof of this proposition can be found in Appendix C, however, we discuss in what follows the intuition for this result. For any positive iceberg cost a country will only export a good if its price is at home is lower than abroad. Decreasing iceberg costs leads to convergence of prices, lowering the price abroad and increasing the price at home. The wage of workers is linearly increasing in the price of the good they produce. Thus the lowering of the price abroad makes migrating less attractive to the workers who produce that good. By the same logic, for goods that are imported, a decrease of iceberg costs will lower the price of the good at home and increase the price of it abroad. This makes migrating more attractive for those workers that produce this good.

A key step in the result is the fact that trade integration leads not just to price convergence, but to convergence to a price that is in between the equilibrium prices in trade-autarky. While the gross substitution assumptions that yield this result are restrictive, the conclusion is rather natural and we therefore believe that the result is quite general.

For the second proposition, we additionally assume that preferences are homothetic, so that the fraction of wealth that is expended on a good depends only on prices, not on wealth levels. It should be noticed that utility functions satisfying homotheticity are widely used in trade models, such as the CES function.

Proposition 2. *B exports L and imports H, if trade takes place.*

Proof. Suppose B exported H. Then it must import L. It must mean that $(1+t)^2 p_B \leq p_A$. Since A produces more of H, it also means that relatively more of H is consumed in A. But then we have the following condition on the marginal rates of substitution: $\frac{U_H^A}{U_L^A} = p_A$, while $\frac{U_H^B}{U_L^B} = p_B$. But we have just argued that the left hand side is smaller in A, while the right hand side is larger. □

Given the comparative advantage of country A in good H, this country produces a larger quantity of the high tech good. The homotheticity assumption ensures that the greater production of good H in country A is not simply consumed by A's residents as a consequence of their higher income but that they also use a fraction of this extra income to buy the low tech good instead.

3.5 Endogenous Education

We now endogenize the education level in the economy by letting workers choose their skill level at a cost. In particular, a worker can obtain a low skill level at no cost, or obtain a high skill level at the cost of getting education. We assume that workers have a cost c of education that is specific to the worker. The distribution of education costs is assumed to be uniform on $[0, 1]$ and is measured in time or labour units. Thus, a worker who obtains education at cost c has $1 - c$ units of labour left to sell in the market. Therefore, c captures the innate talent of a worker, with a worker with low c being more talented.

We now assume that the cost of migration is the same for all workers, κ, again measured as a time cost. Thus a worker who migrates but does not get education has $1 - \kappa$ units of labour left to sell, while a worker with education cost c who migrates and obtains an education has $1 - \kappa - c$ units of labour left to sell. This homogenous cost of migration can be thought as the average of the distribution of migration costs used in the previous section, which allows to compare the results in each of the two sections.[2]

We maintain the assumption that U, the utility function of the worker, is homothetic.

3.5.1 Brain skimming

In this section, we discuss a fairly general result on the joint choice of education and migration by the workers, namely that if there is educated migration, it is by the most talented workers, a phenomenon that we label *brain skimming*. In the next section we make stronger assumptions in order to characterize and analyze equilibrium and comparative statics.

It then satisfies a kind of single crossing property, as the next proposition shows.

Proposition 3 (Only the most talented educate and migrate)**.** *If a worker with education cost c obtains education, then so does any worker with education cost $c' < c$.*

[2] The reason that we assume κ is now the same for all workers is that the distribution of the cost of education already gives us enough worker heterogeneity to have meaningful and non-degenerate comparative statics.

If a worker with education cost c obtains education and migrates, then so does any worker with education cost $c' < c$.

While the proof of this proposition can be found in Appendix C, we provide here some intuition about how this result is obtained. The first part of the proposition is obvious. The second part of the proposition follows from the fact that the cost of migration is a time cost and is therefore relatively less costly when the worker has more time remaining after obtaining education. The homotheticity assumption is used to ensure that the different prices in the two countries do not off-set this comparison.

The proposition therefore establishes that with endogenous education, brain drain, if it occurs, drains the most talented individuals from the country, even among the educated.

3.5.2 Equilibria with endogenous education

In this section, we describe all possible equilibria with endogenous education. In order to make progress, we make much more specific assumptions about preferences and parameters. We hope to generalize our results in further work. We now lay out these specific assumptions.

Firms

There are two types of firms in each country, with production functions given respectively by:

$$\begin{aligned} &Y_A^L(l_A) = l_A \qquad &Y_1^H(h_1) = R_A h_A \qquad &1 < R_B < R_A, \\ &Y_B^L(l_B) = l_B \qquad &Y_2^H(h_2) = R_B h_B, & \end{aligned}$$

where R_j is the productivity of workers employed in the H sector in country j. From the linearity if the production function, it follows immediately that in equilibrium

$$w_A^L \geq 1, \quad w_B^L \geq 1, \quad w_A^H \geq R_A p_A, \quad w_B^H \geq R_B p_B,$$

where w^i_j denotes the wage for labour in good i in country j and p_j the price of the H good in country j. These hold with equality if production is positive in the respective sectors.

Workers

Workers maximize $U(x_L, x_H) = \log(x_L) + \log(x_H)$ subject to $x_L + px_H \leq I$. This is clearly consistent with the earlier, more general assumptions on preferences. Workers supply labour inelastically, once they have made decisions on how much time they spend on education and migration activities. Their income, I, is given by the wage earned during the remaining time in the sector for which they are educated and in the country where they reside. Given I, the demands for each good are given by:

$$x_L = \frac{I}{2} \qquad x_H = \frac{I}{2p_j},$$

Education and migration choice

Taking prices and wages as given, workers make decisions on whether to migrate and become educated as a function of their education-cost type. We compare the utility of the four available possibilities, using the transformation $\tilde{U} = 4\exp U = \frac{I^2}{p_j}$ and writing U instead of $\tilde{U}$. Depending on these choices, the utility of a worker in country B is given by:

$$U_{NE,NM} = \frac{1}{p_B} \qquad U_{NE,M} = \frac{(1-\kappa)^2}{p_A}$$
$$U_{E,NM} = (1-c)^2 R_B^2 p_B \qquad U_{E,M} = (1-\kappa-c)^2 R_A^2 p_A$$

where *NE* stands for no education; *NM*, for no migration; *E* for the worker that gets education, and, lastly, *M* refers to a worker that chooses to migrate.

Classification of Equilibria

First, it should be noticed that, since country A has absolute advantage, real wages are higher in this location and workers from A have no incentive to emigrate to country B. As a result, there is only one threshold in the cost of education in country A, c_A where $0 \leq c_A \leq 1$, that splits its population into educated and non-educated workers.

In light of proposition 3 and the fact that no worker migrates from A, the education and migration choices in both countries can be summarised by 4 numbers: each country's ability threshold below which individuals obtain a higher utility from getting education (c_A and c_B, respectively); country B's ability threshold below which workers prefer to migrate and get education (c_{EM}, where $0 \leq c_{EM} \leq c_B \leq 1$); and the mass of uneducated workers that migrate from country B to country A.

It should be kept in mind that when trade costs made trade unfeasible (trade-autarky), both countries produce both goods, as otherwise the price of the good that is not produced would go to infinity and markets would not clear. Furthermore, country A also produces both goods in the free-trade regime, given that (i) at least the worker with cost of education equal to 1 never chooses to become educated, and (ii) comparative advantage in H ensures the production of that good in country A. This implies that $0 < c_A < 1$.

We now note that only the following constellations of c_B, c_{EM}, M can occur. The proof of this observation is in Appendix C.

1. **No migration** (NM): no worker in B has incentives to migrate. Workers with a relatively low cost of acquiring education do get educated, while the rest do not. Thus, it must be that $0 = c_{EM} < c_B$ and $M = 0$.

2. **Partial brain drain** (PBD): the most able workers from B get education and migrate to A. Workers with higher (but still low on average) costs of migration choose to get education and stay in B while some other workers remain uneducated and do not migrate, that is, $0 < c_{EM} < c_B < 1$ and $M = 0$.

3. **Full brain drain** (FBD): workers in B with a relatively low cost of migration choose to get education and migrate to A, while the rest remain non-educated and stay in B. Hence, in this type of equilibrium $0 < c_{EM} = c_B$ and $M = 0$.

4. **No education** (NE): no worker in B has incentives to get education nor to migrate. Thus, only the L good is produced in B and $c_B = c_{EM} = 0$ and $M = 0$.

5. **Flood** (F): all workers in B prefer to migrate. The most able workers

get education while the rest do not. In this equilibrium $c_{EM} = c_B > 0$ and $M > 0$.

6. **Partial flood** (PF): the most able workers in B choose to get education and migrate to A, while workers with higher (but still low on average) cost of migration choose to get education but stay in B. Among workers with an excessive cost of education that choose to remain as non-educated, some decide to migrate while others decide to stay in B. Hence, in this equilibrium it must be that $0 < c_{EM} < c_B < 1$ and $M > 0$.

We now establish the conditions on parameters that yield the different types of equilibria outlined above. Given a productivity gap ($\frac{R_A}{R_B}$) and a migration cost (κ), we can determine the exact value of all ability thresholds (c_A, c_B, c_{EM}) and the mass of uneducated migrants (M), which in turn determine which equilibrium takes place, as we just described. In order to simplify our analysis, we focus on two extreme cases: trade-autarky ($t = 0$) and free trade ($t = 1$).

Trade-autarky

Because in the absence of trade each country must produce both goods, neither the FBD nor the NE equilibrium can exist without trade.[3]

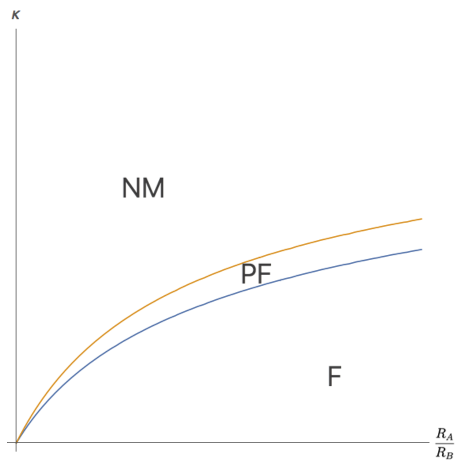


Figure 3.1 – Equilibra in trade-autarky

Figure 3.1 depicts the different equilibria classified in the previous section as a function of the set of parameter values capturing the productivity gap, R_A/R_B, and the cost of migrating, κ. As expected, overall migration is higher

[3] Unless one country loses all its population like in the F equilibrium, in which case nothing will be produced in that country.

the higher the productivity gap between countries and the lower the cost of migration.

First, at a low cost of migration and a productivity gap wide enough to motivate migration, the F equilibrium takes place and all individuals prefer to leave the country. Individuals with high ability in country B get education and move to country A where they are more productive, while individuals with low ability prefer not to get education but also move to A. The reason is that, despite the fact that their productivity remains unchanged, they enjoy lower prices abroad than at home. Second, for given productivities, as the cost of migration increases, incentives to leave B are reduced and the equilibrium shifts to PF, in which some individuals, both educated and non-educated, leave to A while some others, also educated and non-educated, remain in B. Finally, for a relatively high cost of migration given the productivity gap, every worker is better off by staying in B, i.e. the NM equilibrium takes place. Moreover, the PBD does not arise because the case in which uneducated workers are indifferent between migrating or not, so is the worker with no cost of getting education, so the PF takes place.

Free trade

When both countries can trade freely, the price of goods equalizes, wiping out incentives to migrate without acquiring education. Hence, both the F and PF equilibria will not exist. At the same time, complete specialization in good L is possible in country B, as individuals can now import good H from country A, so the FBD and NE equilibria become feasible.

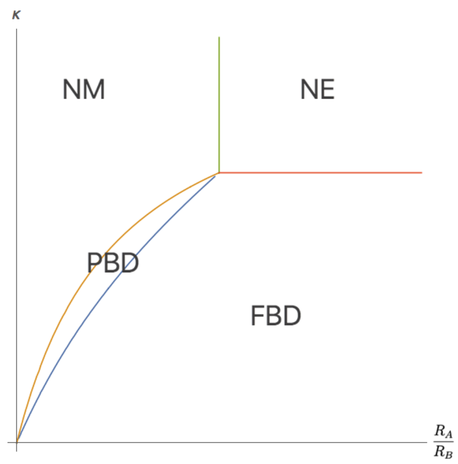


Figure 3.2 – Equilibria with free trade

Figure 3.2 depicts the different equilibria as a function of the set of parameter values for the case of free trade. Again, overall migration is higher, the higher the productivity gap between countries and the lower the cost of migration. We now describe this figure in detail, first for low productivity gaps and then for higher productivity gaps, focusing on how the classification of the equilibrium changes as the cost of migration increases.

For relatively low productivity gaps, all educated individuals leave country B when the cost of migration is relatively small, giving rise to the FBD equilibrium. As the cost of migration increases, the brain drain from country B becomes less severe. In the PBD equilibrium, the most talented workers choose to become educated and migrate up to the threshold worker, c_{EM}, for whom the real wage net of the cost of migration and the cost of education is equalized across countries. Workers with a lower ability than this threshold choose to remain in B. Workers with cost of intermediate cost of education, $c_{EM} < c < c_B$, choose not to migrate, but do obtain education. The threshold worker c_B is indifferent between obtaining an education or not, conditional on not migrating. Finally, for a relatively high cost of migration, no individual chooses to migrate. In that case, high ability individuals get education and produce the H good, while low ability individuals produce the L good.

When the gap in productivity is very high, country B specializes completely in the production of the L good. Hence, for a low cost of migration, the FBD equilibrium holds, that is, all high ability individuals get education and migrate to A. In contrast, when the cost of migration is high, no individual can afford to migrate. Given a high productivity gap, country B is not competitive in producing good H and as a result no one gets education and only the L good is produced in country B.

Migration and trade regimes: a comparison

In this section we compare the regions of parameters for which different equilibria take place in both trade regimes. This comparison allows to draw some general conclusions about the impact of free trade on migration. In particular, we find that the range of parameters that prevent migration in trade-autarky is broader than in free trade, pointing towards a positive relationship between migration and trade.

In Figure 3.3 we depict the loci above which NM equilibria take place as a function of the parameters of the model for both the trade-autarky and free-trade regimes. This is equivalent to superposing the maps reported in the previous section and focusing on the lines that separate the NM equilibria from the rest. As can be observed, the no-migration region is smaller in the free trade regime.

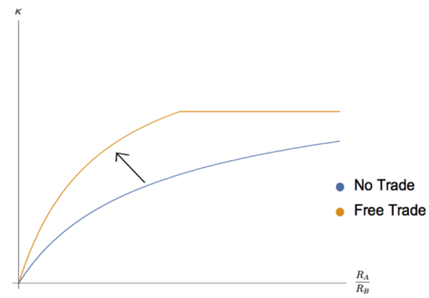


Figure 3.3 – Migration and trade regime

Nevertheless, if parameter values are such that there is migration in trade-autarky (and hence, there is migration also with free trade), the migration intensity may be greater under free trade than in the no-trade regime. Indeed, if the cost of migration is very low, all individuals from country B leave to A (F equilibrium).

3.6 Additional Results: Simulations

In the remainder of the paper, we choose a particular productivity gap and simulate the economy in order to compare the size and educational composition of migration flows in the two trade regimes and to obtain some valuable insights about the effects of trade and migration on each country's welfare. The productivity gap is set at a relatively *low* level in order to prevent the F and NE equilibria, which we find less interesting.[4] However, we do not establish formally the parameter conditions under which these results hold and, hence, they should be taken as merely illustrative.[5]

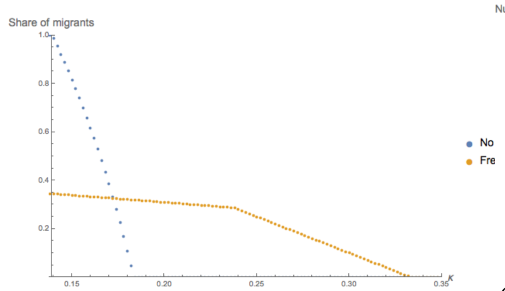
[4]In particular, we set $R_A = 3$ and $R_B = 2$.

[5]We hope to generalize these results in future work. For now, careful checks have been satisfactorily conducted to ensure that these qualitative results are not just a particular case.

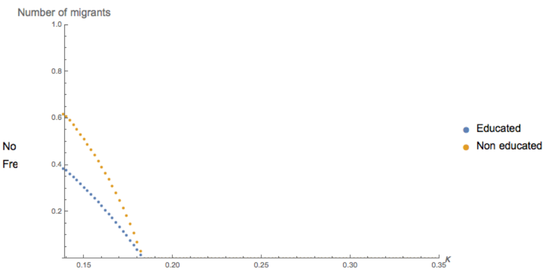
3.6.1 The relationship between migration, trade, and education

Migration flows: size and educational composition

Panel 3.4a in Figure 3.4 compares the share of individuals from country B that decide to migrate in both regimes as a function of the cost of migration, κ, for the chosen productivity gap. Panel 3.4b shows the composition of the migrant workforce as a function of the cost of migration in trade-autarky. For relatively low values of κ, all workers in B leave to country A, since all workers enjoy lower prices and educated workers happen to be more productive than in B. As κ increases, both educated and non-educated workers migrate less, as the reduction of the productive time in country A reduces the gains from migration, up to a point in which the cost of migration is so substantial that no individual chooses to migrate. Observe that in trade-autarky most migrants choose not to get education, more so as the cost of migration increases.



(a) Total migration



(b) Migrants by education (Trade-Autarky)

Figure 3.4 – Migration flows and their educational composition

Recall that free trade wipes out incentives to migrate for low ability individuals, since prices get equalized in both countries. Therefore, while in the trade-autarky, high ability migrants choose to get education and the rest of migrants do not, under free trade only high ability individuals choose to migrate to A. Incentives to migrate and becoming educated for high-ability individuals remain similar under free trade because educated workers are more productive in country A, given differences in the technology employed in both countries, which translates in higher real wages. To see why, notice that, given homothetic preferences, with trade the world relative price of good H falls in between the no-trade relative prices of H in each of the two countries. As a result, the relative price of H in country B falls with free trade with respect to no trade.

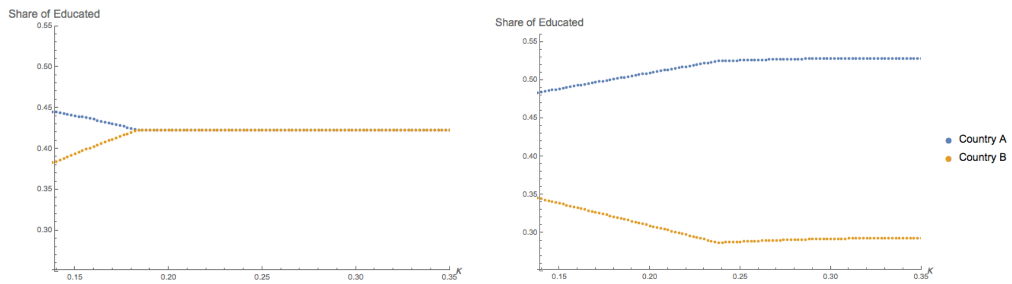
Hence, firms in B cannot offer competitive wages to educated workers, who therefore prefer to pay the cost of emigration and migrate to country A, where wages are higher.

The discussion above implies that, in the jargon of the trade and factor mobility literature, trade and educated-migration become *complements* (as in the Hecksher-Ohlin model), while trade and non-educated migration become *substitutes* (as in the Ricardian model). To the best of our knowledge, this is novel result in the trade-migration literature which follows from our (realistic) assumption that differences in technology only exist in the production of the H good. Free trade equalizes the reward of the labour input in the sector for which there are no technological differences in the production of the L good, i.e., the real wage of non-educated workers becomes identical. However, the same argument does not hold in the H sector, as educated individuals are more productive in country A than in country B when producing the H good. Hence, in the sector where technological differences are present in the production of the H good, free trade is not enough to equalize the reward of labour, and hence factor mobility is needed to exhaust differences in real wages. This means that migration of educated workers is complementary to trade, giving rise to the phenomenon known as *brain drain*.

Educational attainment and trade

Since we have allowed for endogenous education, free trade also affects the incentives to get education in *both* countries. Figure 3.5 illustrates how individuals incentives to become educated, irrespective of the choice on migration, differ across trade regimes, as a function of the cost of migration. Observe that, for any positive cost of migration, the share of workers from country A (B) that gets education is higher (lower) under free trade than under the trade-autarky regime. Therefore, patterns of specialization are reinforced by the endogenous formation of factor inputs.

Let us now consider the effect of an increase in κ on the share of educated individuals absent trade and with free trade. In trade-autarky, the share of individuals that choose to get education in country B is increasing in κ. The reason is that as the cost of migration increases, more individuals stay in country B. As mentioned earlier, without trade, migrants are mostly low-ability individuals that do not get education. Thus, when they stay in country B



(a) No Trade **(b)** Free Trade

Figure 3.5 – Share of Educated Workers by Origin

the real wage of educated workers increase as they become relatively more scarce. To compensate for the rise in the real wage of educated workers, more individuals in B choose to become educated, dragging the skill premium down. Indeed, the maximum level of education in country B is achieved when all workers stay in the country. Individuals from country A experience the other side of the coin.

On the contrary, when both countries trade freely, the share of educated individuals in country A is increasing in κ. The reason is that with free trade, all migrants from B are educated, reducing the skill premium in country A. As the cost of migration increases, less individuals from country B find profitable to enter country A and the skill premium goes up, encouraging individuals from A with a medium level of ability to get education. Another way to put it is that when high ability individuals from B can migrate to country A at a low cost, they effectively displace medium-ability natives from A, who are therefore forced to move to the L sector. Regarding country B, with free trade, the relative price of good H is too low for its low-productivity firms, which cannot compete against A's high-productivity firms. Thus, as migration becomes more costly incentives to get education in country B decrease, since they can no longer migrate and enjoy country A's larger skill premium.

3.6.2 Migration, trade, and welfare

In this section we focus on the impact of trade and migration on each country's welfare. For this purpose, we assume a utilitarian social welfare function. Since the population in each country is changing and individuals are heterogeneous in their ability, we compute welfare per capita of the group of interest, which coincides with the average welfare of the group.

We find that while natives' aggregate welfare is always higher under free trade than in the trade-autarky regime, reflecting gains from trade, natives from A enjoy a greater welfare level as migration cost increase, because they face less competition from top ability immigrants. Moreover, we find that a country's welfare per capita is not always larger under free trade than in trade-autarky nor increases monotonically as migration costs decrease. These results are driven by changes in productive efficiency and by the skill composition of each country's population, something that we explain in detail below.

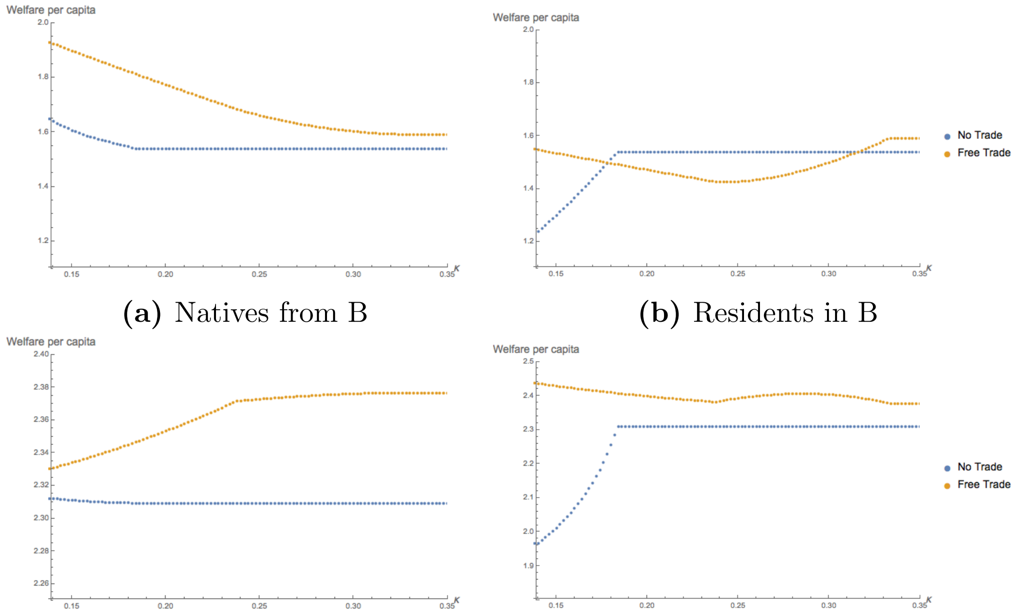
A concern in comparing free trade with trade-autarky is the well-known argument of *winners* and *losers* stemming from trade. That is, when countries open to trade there will be individuals that will benefit from trade and individuals that will be worse off. However, since free trade leads to gains in efficiency in global production, it is possible (under some reasonable assumptions) to find an allocation that makes everybody better off in a Pareto sense (Dixit and Norman, 1980). To deal with this issue, we calculate welfare of the most equal allocation which is the one that maximizes welfare given an utilitarian social welfare function with equal Pareto weights. In other words, we calculate the average allocation of the group, which is the inserted in the utility function.[6] Moreover, an advantage of this measure of welfare with respect to GDP per capita is that the latter may overstate or understate the gains/losses from trade.

Figure 3.6 shows welfare per capita in both countries as a function of the cost of migration, κ, under both regimes. Panels 3.6a and 3.6c display the natives' average welfare from country B and country A, respectively, while Panels 3.6b and 3.6d display the same variable but measured among residents.

Comparing natives' average welfare in trade-autarky to free trade (Panels 3.6a and 3.6c), one can see the gains from trade: welfare p.c. with free trade is above that of no trade for any cost of migration. More interestingly, these two panels show that countries' welfare p.c. behaves in opposite ways as the cost of migration increases. While country B's welfare p.c. decreases monotonically as κ increases, reflecting the gains from migration; country A's welfare p.c. increases monotonically with κ, though in the free trade regime. The reason behind this last relationship is that the share of educated individuals among natives from A is increasing in the cost of migration, as was shown in Figure

[6] For comparison, we have calculated welfare per capita of the equilibrium allocation. While qualitative results remain almost the same, gains from trade are underestimated

3.5. Even though very talented migrants from B lower the price of H, which is produced more efficiently when they work in country A, the displacement into the L sector of medium ability native workers in this country dominates.



(a) Natives from B

(b) Residents in B

(c) Natives from country A

(d) Residents in country A

Figure 3.6 – Welfare per capita

We believe that examining each country's average welfare among residents is informative as well, because, in general, governments can only tax residents in their own country. Thus, while a government could not redistribute income from educated migrants in A to non-educated residents in country B, it could certainly redistribute income from educated residents. Even though we have not featured redistribution in our model, we do focus on the most equal allocation, and hence, we take these results as illustrative of potential gains from redistribution. We are aware that had we included taxes, individuals would have responded to them and some of their choices would have been different; how different is a matter that we leave to future research.

There are two interesting (and related) results regarding the average welfare of residents in B (portrayed in Panel 3.6b): first, welfare p.c. in trade-autarky is above that under free trade for some values of the migration cost; second, welfare p.c. is not monotonically decreasing in the cost of migration.

Given that we focus on the most equal allocation, these results are ultimately driven by the average productivity among residents in B, which increases total output and, hence, our measure of welfare. In the trade-autarky regime, the average productivity of workers increases with the cost of migration, since we showed that most talented individuals have higher incentives to migrate. In addition, since we are looking at country B's residents and there is no migration from country A into this location, a larger cost of migration does not involve a loss of productive time, since no resident in B bears this cost. As a consequence, welfare per capita increases in κ.

In contrast, in the free trade regime, average productivity first decreases with κ and then increases. The reason is that as the cost of migration raises, two opposite forces are at play: on the one hand, there is less emigration from country B, which may increase the average productivity of this country (only if H is produced); on the other hand, the amount of labour that educated emigrants supply in A decreases, which raises the price of the H good, and hence decrease average welfare. Therefore, the average welfare of residents in B will depend on which of the two effects dominate. For low κ, all high-ability individuals choose to get education and migrate, that is, the FBD equilibrium holds. Since no educated workers are left in country B, a rise in κ only has the effect of increasing the prices at which residents in B, all non educated, can purchase goods. Yet, if the cost of migration increases enough as to prevent the migration of medium-ability individuals, country A will not be able to satisfy all the demand for good H. Consequently, some of the good H will be produced in country B where residents will no longer be all non educated, raising average welfare in that country. Lastly, for high κ welfare per capita is larger with free trade than in trade-autarky, which indicates that the price effect dominates.

Finally let us consider the impact of free trade and migration in welfare p.c. of all residents in A, which includes migrants and the native population (depicted in panel 3.6d). First, welfare p.c. of residents in country A is increasing in κ in trade-autarky, because, with this regime, the amount of uneducated migrants decreases with migration costs. Second, and unlike in country B, average welfare per capita among A's residents is always larger in free trade than in trade-autarky. This reflects gains from trade together with the fact that with free trade this country experiences only educated (top ability) migration. Finally, in the free trade regime, the welfare p.c. of residents in country A as a function of κ is non monotonic, as was the case in country B.

3.7 Conclusions

In this paper we have shown the importance of studying trade and migration jointly. Using a simple model that features technological differences only in the sector that employs high-skilled workers, we found that trade integration, modelled as a reduction of trade costs, leads to the outflow of educated workers from the least to the most advanced country. More generally, this result highlights the impact of trade on the skill composition of migration flows.

We have also provided a compelling example suggesting that trade integration can also affect the educational attainment of countries on the long run. In particular, the simulated results pointed towards a divergence in the educational attainment of each country's population, which reinforced sectoral patterns of trade. We believe this issue deserves further attention, since it has important implications on global inequality. Thus, it would be interesting to establish formally which exact conditions give rise to a divergence in the human capital stock of countries as they trade. Furthermore, these considerations suggest that valuable insights could be obtained from examining these issues in a dynamic framework and possibly relate them to the literature on poverty traps.

Appendix A

Differences in LFP over time

The lower LFP of women with children in big cities relative to small cities has been a persistent feature of the US labour markets over the past decades. In Figure A.1, I plot the participation rates of women with children in small MSAs (dashed line) and big MSAs (continuous line) for all years between 1990 and 2010 for which city information is available. Though the overall trend is increasing, that is, the LFP of women with children is rising over time, spatial differences have been stable at around 5%.

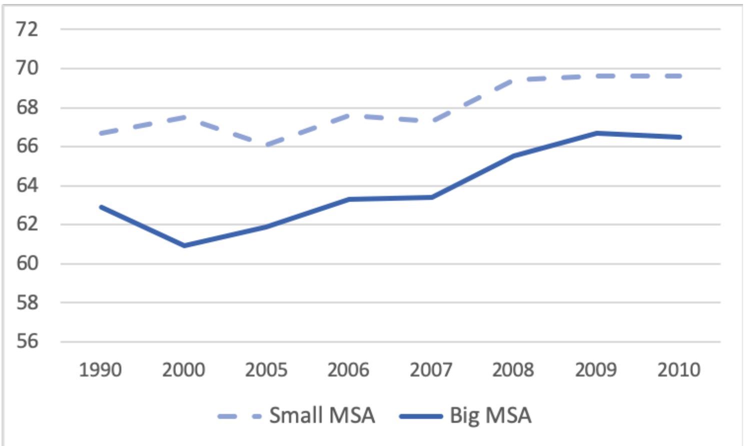


Figure A.1 – LFP rates of women with children over time

Notes: This figure shows the LFP rates of women with children over time for big (solid line) and small MSAs (dashed line) for recent years in which city size can be constructed. Source: 1990 and 2000 Census, and ACS 2005-2010, available in IPUMS, own elaboration.

LFP and City Size

Table A.1 displays the coefficients of some of the control variables included in the regressions reported in Table 1.1. All control variables have the expected sign.

Table A.1 – BCCP and City Size: Detailed

	(1) LFP Women Children	(2) LFP Women Children	(3) LFP Women Children	(4) LFP Women Children
Big City	-0.057***	-0.038***	-0.038***	-0.032***
	(0.001)	(0.002)	(0.002)	(0.002)
Age			0.016***	0.024***
			(0.001)	(0.001)
$Age^2/100$			-0.000***	-0.000***
			(0.000)	(0.000)
Edu: group 2			0.097***	0.101***
			(0.002)	(0.002)
Edu: group 3			0.166***	0.187***
			(0.002)	(0.002)
Edu: group 4			0.173***	0.240***
			(0.002)	(0.003)
Edu: group 5			0.312***	0.439***
			(0.003)	(0.004)
Foreign-Born			-0.105***	-0.086***
			(0.002)	(0.002)
Two children			-0.069***	-0.069***
			(0.001)	(0.001)
Three children			-0.130***	-0.130***
			(0.002)	(0.002)
Spouse's edu 2				0.008***
				(0.002)
Spouse's edu 3				-0.010***
				(0.002)
Spouse's edu 4				-0.112***
				(0.002)
Spouse's edu 5				-0.216***
				(0.003)
Year FE	Yes	Yes	Yes	Yes
State FE	No	Yes	Yes	Yes
Controls	No	No	Yes	Yes
Spouse	No	No	No	Yes
R-squared	0.004	0.012	0.073	0.086
Observations	745138	745138	745138	745138

Notes: This table displays differences in the likelihood to participate among married women with children across city size. The dependent variable is a dummy equal to one if the woman participates in the labour market. Additional controls include: age and age squared, dummies for races and a dummy for foreign-born individuals for the wife and for her husband. The sample is restricted to married women with children under 12 in metropolitan areas. I also exclude women below 25 or over 55 years old. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

Intensive margin

In Table A.2 I show that city size differences in the probability of working full-time are similar (but smaller) to the extensive margin.

Table A.2 – Intensive Margin and City Size

	(1) Full-time Single Men	(2) Full-time Married Men No Children	(3) Full-time Married Men Children	(4) Full-time Single Women	(5) Full-time Married Women No Children	(6) Full-time Married Women Children
Big City	0.008***	-0.001	-0.000	0.017***	0.004*	-0.000
	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)	(0.001)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.016	0.010	0.013	0.024	0.019	0.013
Observations	696182	550141	937777	489768	522560	937777

Notes: This table displays differences in the probability of working full-time across city size. The dependent variable is a dummy equal to 1 if the person works full-time. In all regressions I control for age, age squared, dummies for race, and a dummy equal 1 if the person completed a bachelor. I also include the same spouse characteristics when the sample includes married women. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05. The sample is restricted to workers with ages between 25 and 55 in metropolitan areas.

Fertility Rates

Fertility rates are fairly equal across city size even after controlling for age, education, race, and nativity. As can be seen in Table A.3, the average number of children in a city that is twice as large is 0.006 p.p. lower.

Table A.3 – Fertility Rates and City Size

	(1) Number of Children	(2) Number of Children	(3) Number of Children	(4) Number of Children
Big City	0.029***	-0.012***	0.002	-0.030***
	(0.002)	(0.002)	(0.002)	(0.002)
Age			-0.006***	-0.006***
			(0.000)	(0.000)
High Skill			-0.090***	-0.089***
			(0.002)	(0.002)
Black/African American/Negro			0.028***	0.055***
			(0.002)	(0.002)
American Indian or Alaska Native			0.122***	0.109***
			(0.011)	(0.011)
Chinese			-0.290***	-0.301***
			(0.005)	(0.005)
Japanese			-0.204***	-0.240***
			(0.010)	(0.010)
Other Asian or Pacific Islander			-0.084***	-0.097***
			(0.004)	(0.004)
Other race, nec			0.148***	0.133***
			(0.003)	(0.004)
Two major races			0.008	0.003
			(0.006)	(0.006)
Three or more major races			0.050*	0.032
			(0.023)	(0.023)
Foreign-Born			0.161***	0.163***
			(0.002)	(0.002)
Year FE	Yes	Yes	Yes	Yes
State FE	No	Yes	No	Yes
R-squared	0.000	0.005	0.015	0.019
Observations	3525273	3525273	3525273	3525273

Notes: This table displays differences in the number of children across city size. The sample is restricted to women with children in metropolitan areas. I also exclude women below 20 or over 55 years old. An individual is considered high-skilled if she has completed at least one year of college. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

Similarly, the probability of having children does not vary too much across city size. The probability of having children in a big city is 2.5 p.p. lower than in a small one. Notice that differences in the probability of having children could point towards selection.

Table A.4 – Probability of Having Children and City Size

	(1) Probability of Having Children	(2) Probability of Having Children	(3) Probability of Having Children	(4) Probability of Having Children
Big City	-0.006***	-0.017***	-0.015***	-0.025***
	(0.001)	(0.001)	(0.001)	(0.001)
Age			0.003***	0.003***
			(0.000)	(0.000)
High Skill			-0.076***	-0.075***
			(0.001)	(0.001)
Black/African American/Negro			0.134***	0.136***
			(0.001)	(0.001)
American Indian or Alaska Native			0.049***	0.051***
			(0.004)	(0.004)
Chinese			-0.067***	-0.066***
			(0.002)	(0.002)
Japanese			-0.102***	-0.102***
			(0.004)	(0.004)
Other Asian or Pacific Islander			-0.019***	-0.020***
			(0.001)	(0.001)
Other race, nec			0.104***	0.103***
			(0.001)	(0.001)
Two major races			0.032***	0.033***
			(0.002)	(0.002)
Three or more major races			0.062***	0.064***
			(0.007)	(0.007)
Foreign-Born			0.081***	0.084***
			(0.001)	(0.001)
Couple			0.339***	0.338***
			(0.001)	(0.001)
Year FE	Yes	Yes	Yes	Yes
State FE	No	Yes	No	Yes
R-squared	0.000	0.002	0.138	0.139
Observations	6263940	6263940	6263940	6263940

Notes: This table displays differences in the probability of children across city size. The dependent variables is a dummy equal to one if the woman has at least one child. The sample is restricted to women with children in metropolitan areas. I also exclude women below 20 or over 55 years old. An individual is considered high-skilled if she has completed at least one year of college. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

Cities' Summary Statistics

Table A.5 displays some summary statistics regarding differences in the big and the small city. The first row of each variable displays the mean of that variable in each city size while the second row displays the standard errors of the mean.

Table A.5 – Summary Statistics

	Small City	Big City
Commuting time in min	22.4	31.6
	(0.000)	(0.000)
% Workers in LH occ	0.43	0.47
	(0.000)	(0.000)
Hours per day in LH occ	8.35	8.47
	(0.000)	(0.000)
Hours per day in linear occ	7.04	7.06
	(0.000)	(0.000)
(log) wage in LH occ	2.73	2.96
	(0.000)	(0.000)
(log) wage in linear occ	2.51	2.62
	(0.000)	(0.000)
Years of schooling in LH occ	9.89	11.3
	(0.000)	(0.000)
Years of schooling in linear occ	8.72	9.17
	(0.000)	(0.000)

Notes: This table reports summary statistics for small and big cities. Commuting times refer to one-way travelling time to work in minutes. LH stands for the long hours occupation. Standard errors of the mean are reported in parentheses.

BCCP: interactions

In this section I show that differences in LFP across city size do not change when I consider a fully saturated model with interactions for sex, marital status, and parenthood instead of running separated regressions for each demographic groups (as in Table 1.2).

Table A.6 – BCCP interactions

	(1)	(2)	(3)
	LFP	LFP	LFP
Big City	0.022***	0.015***	0.050***
	(0.001)	(0.001)	(0.001)
Female × Big City			-0.019***
			(0.001)
Married × Big City			-0.050***
			(0.001)
Children × Big City			-0.040***
			(0.003)
Married × Children × Big City			0.050***
			(0.003)
Female × Married × Big City			0.019***
			(0.002)
Female × Children × Big City			0.022***
			(0.004)
Female × Married × Children × Big City			-0.052***
			(0.004)
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
Controls	No	Yes	Yes
Spouse	0.029	0.086	0.087
R-squared	3915346	3915346	3915346

Notes: This table displays the results of regressing a dummy equal to one if the person participates in the labour market on the (log) city population. All regressions include year and state fixed effects and control for observable characteristics: age, age squared, race, education, and a dummy equal to 1 if the person is foreign-born. The sample is restricted to individuals who live in metropolitan areas and have ages between 25 or 55. Individuals with children over 12 years old have been excluded from the sample. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

Unemployment rate

In this section I show that differences in female unemployment rates across city size are not an important explanation of the BCCP.

Table A.7 – BCCP and average unemployment rate

	(1) LFP Women Children	(2) LFP Women Children	(3) LFP Women Children	(4) LFP Women Children
Big City	-0.032***	-0.032***	-0.002	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Female Unemp. Rate		-0.014		0.008
		(0.025)		(0.026)
Avg Commute			-0.255***	-0.256***
			(0.015)	(0.015)
% Workers LH			-0.037	-0.037
			(0.020)	(0.020)
Avg Wage			0.002	0.002
			(0.012)	(0.012)
(log) Housing			-0.008	-0.008
Price Index			(0.005)	(0.006)
Year FE	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
Spouse	Yes	Yes	Yes	Yes
R-squared	0.086	0.086	0.087	0.087
Observations	745138	745138	744685	744685

Notes: This table shows the impact of several city characteristics in the BCCP. The dependent variable is a dummy equal to one if the woman participates in the labour market. All columns include year and state fixed effects and control for individual's observables and partner's as in Table 1.1. The sample is restricted to married women with children under 12 who live in metropolitan areas. I also exclude women below 25 or over 55 years old. Robust standard errors in parenthesis, significance levels: *** p< .001, ** p<.01, * p<.05.

Policy Counterfactuals: additional results

Table A.8 displays the result of implementing childcare subsidies as described in Section 1.6 but eliminating the convexity, that is, setting $\theta = 0$. Therefore, the first panel describes the baseline economy when occupation 1 does not highly reward long working hours while the other two panels report the results of each policy counterfactual while holding $\theta = 0$. Although the impact of childcare subsidies on participation rates is similar to the baseline results, the output drop is halved and the loss of average welfare is less pronounced.

Table A.8 – Policy Counterfactual (without Convexity): Childcare

Panel a: Baseline (no subsidy)				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	68.9	-	66.5	71.1
Average welfare	1.12	-	1.23	0.99
Output p.c.	8.7	-	10.0	7.4
Hours worked	7.8	-	7.9	7.8
Panel b: Subsidy in both cities				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	78.2	+13	80.8	75.7
Average Welfare	1.08	-4	1.16	1
Output p.c.	8.5	-2	9.9	7.2
Hours worked	7.4	-5	7.4	7.4
Panel c: Subsidy only in the big city				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	73.8	+7	87.3	58.7
Average Welfare	1.08	-4	1.12	1.04
Output p.c.	8.5	-2	9.8	7.3
Hours worked	7.6	-3	7.3	7.9

Notes: This table displays the participation rate, average welfare, output per capita, and average hours worked in the whole economy (column *Total*) and in each of the cities (columns *Big City* and *Small City*) when θ is set to 0. Panel a reports the values of these variables in the baseline economy while Panel b and c do so for the cases in which the childcare subsidy is available in both cities and just in the big city, respectively. The column labelled as *% Change* displays the percentage change of a variable with respect to the baseline economy with $\theta = 0$.

Table A.9 displays the result of implementing childcare subsidies as described in Section 1.6 in the absence of sorting on unobserved preferences, that is, assuming that the underlying distribution of preferences is the same in both cities and banning mobility. Therefore, the first panel describes the baseline economy without sorting while the other two panels report the results of each policy counterfactual and the changes with respect to the baseline economy in the absence of sorting. When mobility is not available to couples, the rise in participation is entirely due to the effect that childcare subsidies have at improving the options in which the wife participates and at worsening the options in which she does not. Therefore, implementing this policy in just one city does not have any effect in the other location.

Table A.9 – Policy Counterfactual (without sorting): Childcare

Panel a: Baseline (no subsidy)				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	66.7	-	66.3	67.2
Average welfare	1.03	-	1.07	0.99
Output p.c.	10.3	-	11.8	8.8
Hours worked	8.4	-	8.5	8.4
Panel b: Subsidy in both cities				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	75.7	+13	76.3	75.0
Average Welfare	0.99	-4	1.01	0.96
Output p.c.	10.0	-3	11.4	8.8
Hours worked	8.1	-4	8.1	8.0
Panel c: Subsidy only in the big city				
	Total	*% Change*	*Big City*	*Small City*
Participation rate	71.8	+8	76.3	67.2
Average Welfare	1.0	-3	1.01	0.99
Output p.c.	10.1	-2	11.4	8.8
Hours worked	8.2	-2	8.1	8.4

Notes: This table displays the participation rate, average welfare, output per capita, and average hours worked in the whole economy (column *Total*) and in each of the cities (columns *Big City* and *Small City*) in the absence of geographical sorting on unobserved preferences. Panel a reports the values of these variables in the baseline economy while Panel b and c do so for the cases in which the childcare subsidy is available in both cities and just in the big city, respectively. The column labelled as *% Change* displays the percentage change of a variable with respect to the baseline economy without sorting.

Appendix B

Impact of infertility insurance mandates on gentrification

Table B.1 – The effect of infertility insurance mandates on gentrification

	Prob. above median	Log median income	% College Graduate
	(1)	(2)	(3)
Center	-0.418***	-0.620***	-0.00413*
	(0.0102)	(0.00773)	(0.00217)
Treated × Post	0.0326*	0.00210	0.0233***
	(0.0198)	(0.0151)	(0.00424)
Center × Treated	-0.142***	-0.0322***	-0.000950
	(0.0164)	(0.0125)	(0.00351)
Center × Post	0.0449***	0.0357***	0.0116***
	(0.0131)	(0.00994)	(0.00279)
Treated × Center × Post	0.123***	0.0433***	0.0175***
	(0.0211)	(0.0160)	(0.00451)
Observations	104608	104608	104608
Controls	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
City Size FE	Yes	Yes	Yes
State Trends	Yes	Yes	Yes
City Size Trends	Yes	Yes	Yes

Notes: This table displays the impact of infertility insurance mandates on several measures of gentrification: (1) the probability that a census tract's income is above median income in the city; (2) the census tract's (log) median income; and (3) the percentage of college graduates in a census tract. Controls include: city's population, city's (log) median income, the share of college graduates in the city, and the share of jobs within 3 miles distance from the census tract. This table reports only selected coefficients, the full specification can be found in equation 2.1. Standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$.

The following evidence shows that restricting our attention to cities with at least one million inhabitants barely changes our results. Table B.1 displays the results of running the exact same specification than that in Table 2.3 but including all cities with at least 100,000 inhabitants instead. The results are very similar.

Table B.2 shows that our results are robust to the exclusion of the control variables list in the main specification: (log) income of the city, the share of jobs within 3 miles distance from the neighborhood, (log) population of the city, and the share of college graduates in the city.

Table B.2 – The effect of infertility insurance mandates on gentrification

	Prob. above median	Log median income	% College Graduate
	(1)	(2)	(3)
Center	-0.427***	-0.622***	-0.00126
	(0.0133)	(0.0104)	(0.00301)
Treated × Post	-0.0118	0.151***	0.0120**
	(0.0222)	(0.0175)	(0.00504)
Center × Treated	-0.135***	-0.0371**	-0.00187
	(0.0187)	(0.0147)	(0.00424)
Center × Post	0.0577***	0.0317**	0.00803**
	(0.0161)	(0.0127)	(0.00365)
Treated × Center × Post	0.116***	0.0621***	0.0254***
	(0.0234)	(0.0184)	(0.00530)
Observations	82129	82129	82129
Controls	No	No	No
Year FE	Yes	Yes	Yes
State FE	Yes	Yes	Yes
City Size FE	Yes	Yes	Yes
State Trends	Yes	Yes	Yes
City Size Trends	Yes	Yes	Yes

Notes: This table displays the impact of infertility insurance mandates on several measures of gentrification: (1) the probability that a census tract's income is above median income in the city; (2) the census tract's (log) median income; and (3) the percentage of college graduates in a census tract. This table reports only selected coefficients, the full specification can be found in equation 2.1. Standard errors in parentheses. $*p < 0.10, **p < 0.05, ***p < 0.01$.

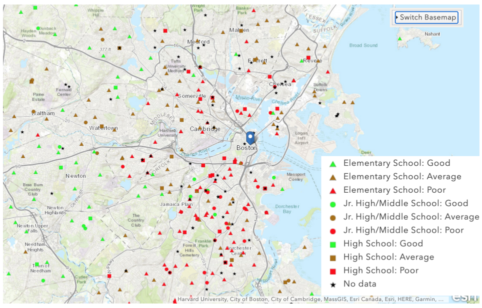
The location patterns of families

As we argued in the main text, one reason why couples may decide to move out of the city center whenever they have children is that the characteristics of the housing stock may not be ideal for children. For instance, houses downtown may be smaller and lack outdoor space. To provide some evidence of this channel, Figure B.1 shows the distribution of houses by number of rooms in the suburbs and downtown. As can be inspected, houses are larger in the suburbs than downtown and thus more suitable for family life.

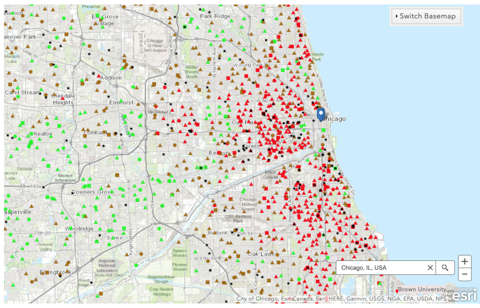
Figure B.1 – Housing Size

Notes: This figure displays the percentage of houses downtown/in the suburbs by the number of rooms in 2000.

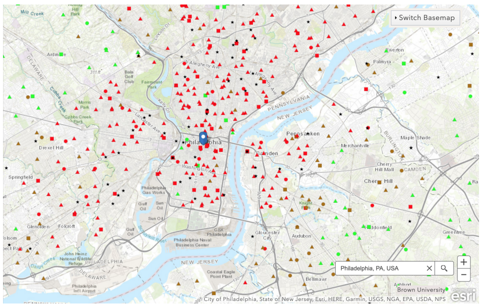
Another reason why families may prefer to relocate to the suburbs is their proximity to children specific amenities that families without children do not value. An important one is the quality of surrounding schools. To illustrate this point, we provide some examples on schools location by quality in Figure B.2. We have accessed maps by “Map US Schools”, which is part of the American Communities Project at Brown University, led by John Logan. These maps have been constructed using data from the National Center for Education Statistics and contain detailed information on the location and quality of most US schools. All cities accessed show a similar pattern, low school quality for all school education levels in the centre of the city and much high school quality for all education levels in the suburbs. We have selected some well-known cities that are located in treated and non treated states to simply illustrate the pattern.



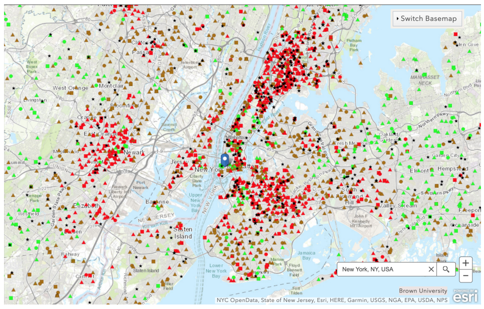
(a) Boston



(b) Chicago



(c) Philadelphia



(d) New York

Figure B.2 – Schools Quality within the City

Notes: These figures displays the location and quality of schools in different cities in 2000. Source: National Center for Education Statistics, accessed via https://s4.ad.brown.edu/Projects/usschools/index.html

Appendix C

Proposition 1 [Trade integration leads to specialised migration]

Suppose B is a net exporter of good $i \in \{H, L\}$ at a given level of t. Then a marginal increase in t will lead to non-increased emigration of workers from B who produce i and non-decreased emigration of workers from B who produce $\neg i$.

Proof. A high-tech firm in B is willing to export if $p_B \leq t^2 p_A$. A low-tech firm in B is willing to export if $p_A \leq t^2 p_B$. Therefore for any $t > 0$ not both firms in B will export.

A high-tech firm in A is willing to export if $p_A \leq t^2 p_B$. A low-tech firm in A is willing to export if $p_B \leq t^2 p_A$. Therefore both countries cannot export the same good. This implies that if a country is exporting all it produces of a good, there is no supply of it in that country. But this contradicts the Inada condition. Therefore if a country is exporting, it does not export all it produces and hence the firm is indifferent between exporting and selling at home: $p_B = t^2 p_A$ or $p_A = t^2 p_B$.

By continuity of aggregate demand, if a country is exporting a positive quantity at a given level of t, then the country will still export a positive quantity if t is marginally increased to $t + \epsilon$. Thus prices will still satisfy $p_B = (t+\epsilon)^2 p_A$ or $p_A = (t+\epsilon)^2 p_B$. Thus, if B is an exporter of the high-tech good, a marginal increase in t will decrease p_A/p_B and if B is an exporter of the low-tech good, a marginal increase in t will increase p_A/p_B.

Now suppose that the marginal change in t moves p_A and p_B in the same direction. By the gross substititution assumption, this would then move demand in both countries in the same direction. But then markets do not clear in both countries. Hence one price must increase and the other decrease.

From profit maximizing of the firm we obtain that the wage w_i^j in country $i \in \{A; B\}$ for worker type $j \in \{H, L\}$ is given by $w_i^j = f_j'(M_i^j(p))p_i^j$, where $M_i^j(p)$ is the number of workers of that type in the country, taking into account migration, which is a function of the prices in both countries.

Assume now that at t, B is exporting L, so that $p_A = t^2 p_B$. The argument for when it is exporting H is identical, but with the roles of p_A and p_B inverted. The claim is then that $M_H^B(p)$ is non-increasing in t and $M_L^B(p)$ is non-decreasing. We will prove each claim in turn by contradiction.

Observe that, if $w = Ap$ and $w' = Ap'$, with $p < p'$, then $B(w, p) \subseteq B(w', p')$. Indeed, if $(x, y) \in B(w, p)$, then $x + py \leq Ap$, so $y \leq A$, so $x + p'y - Ap' = x + (p' - p)(y - A) + p(y - A) \leq x + p(y - A) \leq 0$, so $(x, y) \in B(w', p')$. Clearly, $B(Ap, p) \subseteq B(A'p, p)$ iff $A < A'$ and $B(w, p) \subseteq B(w, p')$ iff $p > p'$.

Suppose that $M_H^B(p(t))$ is increasing in t. Since B is an exporter of the low-tech good, a marginal increase in t will increase p_A/p_B. Since prices have to move in opposite directions, p_A is increasing, while p_B is decreasing in t. Denote $K := R_B f_H'(M_H^B(p(t)))$ Then $w_H^B(t + \epsilon) = R_B f_H'(M_H^B(p(t + \epsilon)))p_B(t + \epsilon) < R_B f_H'(M_H^B(p(t)))p_B(t+\epsilon) = Kp_B(t+\epsilon)$, because $M_H^B(p(t))$ is increasing in t by assumption and f is concave. So $B(w_H^B(t+\epsilon), p_B(t+\epsilon)) \subseteq B(Kp_B(t+\epsilon), p_B(t+\epsilon)) \subseteq B(Kp_B(t), p_B(t)) = B(w_H^B(t), p_B(t))$, where we have used the rules on budget sets derived above. Hence $U(w_H^B(t + \epsilon), p_B(t + \epsilon)) \leq U(w_H^B(t), p_B(t))$.

Similarly, denote $K' := R_A f_H'(M_H^A(p(t)))$ Then $w_H^A(t + \epsilon) = R_A f_H'(M_H^A(p(t + \epsilon)))p_A(t + \epsilon) > R_A f_H'(M_H^A(p(t)))p_A(t + \epsilon) = K'p_A(t + \epsilon)$, because $M_H^A(p(t))$ is decreasing by assumption and f is concave. So $B(w_H^A(t + \epsilon), p_A(t + \epsilon)) \supseteq B(K'p_A(t+\epsilon), p_A(t+\epsilon)) \supseteq B(K'p_A(t), p_A(t)) = B(w_H^A(t), p_A(t))$, where we have used the rules on budget sets derived above. Hence $U(w_H^A(t + \epsilon), p_A(t + \epsilon)) \geq U(w_H^A(t), p_A(t))$.

But now for any κ such that $U(w_H^A(t), p_A(t)) - \kappa \geq U(w_H^B(t), p_B(t))$ we also have $U(w_H^A(t + \epsilon), p_A(t + \epsilon)) - \kappa \geq U(w_H^B(t + \epsilon), p_B(t + \epsilon))$. Thus any agent that would migrate at t migrates at t'. Thus contradicts the assumption that $M_H^B(p(t))$ is increasing.

Suppose $M_L^B(p(t))$ is decreasing in t. Denote $K := f_L'(M_L^B(p(t)))$. Then clearly $w_L^B(t + \epsilon) = f_L'(M_L^B(p(t + \epsilon))) > f_L'(M_L^B(p(t))) = K$, so that $B(w_L^B(t + \epsilon), p_B(t+\epsilon)) \supseteq B(K, p_B(t+\epsilon)) \supseteq B(K, p_B(t))$. Hence $U(w_L^B(t+\epsilon), p_B(t+\epsilon)) \geq U(w_L^B(t), p_B(t))$.

Similarly, denote $K' := f'_L(M_L^A(p(t)))$. Then $w_L^A(t+\epsilon) = f'_L(M_L^A(p(t+\epsilon))) < f'_L(M_L^A(p(t))) = K'$, so that $B(w_L^A(t+\epsilon), p_A(t+\epsilon)) \subseteq B(K', p_A(t+\epsilon)) \subseteq B(K', p_A(t))$. Hence $U(w_L^A(t+\epsilon), p_A(t+\epsilon)) \leq U(w_L^A(t), p_A(t))$.

But now for any κ such that $U(w_L^A(t), p_A(t)) - \kappa \leq U(w_L^B(t), p_B(t))$ we also have $U(w_L^A(t+\epsilon), p_A(t+\epsilon)) - \kappa \leq U(w_L^B(t+\epsilon), p_B(t+\epsilon))$. Thus any agent that would not migrate at t also does not migrate at t'. Thus contradicts the assumption that $M_H^L(p(t))$ is decreasing. □

Proposition 3 [Only the most talented educate and migrate] If a worker with education cost c obtains education, then so does any worker with education cost $c' < c$.

If a worker with education cost c obtains education and migrates, then so does any worker with education cost $c' < c$.

Proof. In light of the proposition, there exist two thresholds in the cost of education, c_{EM} and c_B where $0 \leq c_{EM} \leq c_B \leq 1$, such that a worker migrates and educates if his cost of education c is below c_{EM}, educates without migrating if $c_{EM} < c < c_B$ and does not educate if $c > c_B$.

Denote by $V(I,p)$ the indirect utility at income I and price p. Denote by $V_{x,y}(c), x \in \{M, NM\}, y \in \{E, NE\}$ the indirect utility of becoming educated or not and of migrating or not for a worker in B. Since the indirect utility of not becoming educated is independent of the idiosyncratic cost, we drop the the argument in that case: $V_{NE,y}(c) =: V_{NE,y}$. The proposition is then equivalent to the following.

If $V_{E,M}(c_{EM}) = V_{E,NM}(c_{EM})$, then $V_{E,M}(c) > V_{E,NM}(c)$ iff $c < c_{EM}$. Furthermore, if $V_{E,NM}(c_B) = V_{NE,NM}(c_B)$, then $V_{E,NM}(c) > V_{NE,NM}(c)$ iff $c < c_B$.

To see the first part, observe that

$V_{E,M}(c_{EM}) = V((1-c_{EM}-\kappa)w_H^A, p_A) = V((1-c_{EM})w_H^B, p_B) = V_{E,NM}(c_{EM})$. By homotheticity then: $(1-c_{EM}-\kappa)V(w_H^A, p_A) = (1-c_{EM})V(w_H^B, p_B)$. Let $c < c_{EM}$. Then $V_{E,M}(c) = (1-c-\kappa)V(w_H^A, p_A) = (1-c-\kappa)\frac{1-c_{EM}}{1-c_{EM}-\kappa}V(w_H^B, p_B) = \frac{(1-c_{EM})(1-c)-(1-c_{EM})\kappa}{1-c_{EM}-\kappa}V(w_H^B, p_B) > \frac{(1-c_{EM})(1-c)-(1-c)\kappa}{1-c_{EM}-\kappa}V(w_H^B, p_B) = (1-c)V(w_H^B, p_B) = V_{E,NM}(c)$.

To see the second part, note that:

Non-educated \ Educated	Full Migration	Partial Migration	No Migration	Empty
Full Migration	Flood	X	X	X
Partial Migration	X	Partial Flood	X	X
No Migration	Full Brain Drain	Partial Brain Drain	No Migration	No Education
Empty	X	X	X	X

$V_{E,NM}(c_B) = V((1-c_B)w_H^B, p_B) = V(w_L^B, p_B) = V_{NE,NM}$. Since the price is always the same in the preceding equation, it follows that $(1-c_B)w_H^B = w_L^B$. □

Classification of equilibria

We show that classification of equilibria in section 3.5.2 is indeed exhaustive. Consider the table of possible equilibrium constellations.

The work consists in showing that the constellations marked with an X are indeed impossible. Consider first the constellations that are marked in green. These all have in common that no worker chooses to remain uneducated, which is impossible, given that there are worker with $c = 1$. Consider next the constellations marked in blue. These have in common that both the educated and the non-educated workers emigrate from B, while one group partially stays in B. With free trade, non-educated workers never migrate, so this can only happen in autarky. But then, there would remain workers in B that only have access to one good, which is impossible.

It remains to show that the constellations marked in red are impossible. This is the content of the next lemma.

Lemma 1. *There exist no equilibrium in which all migrants choose not to get education.*

Proof. Observe first that when there is free trade incentives to emigrate as a non-educated worker disappear, since prices are equalized across countries. Thus, we only need to show that the Lemma 1 also holds in autarky.

Clearly, it cannot be that $M = 1 - c_B$, that is, it cannot be that all workers that do not get education migrate, as the supply of L would be 0. Therefore, we focus on the case where the uneducated workers are indifferent between migrating and not, so some of them do while others remain in the country.

This implies:

$$\frac{1}{p_B} = \frac{(1-\kappa)^2}{p_A}$$

The threshold worker, c_B, must be indifferent between getting education and not migrating and not get education (whether he migrates or not). This implies:

$$(1-c_B)^2 R_B^2 p_B = \frac{1}{p_B} \Rightarrow p_B = \frac{1}{R_B(1-c_B)}$$

where we compared getting education and not while not migrating.

Finally in country A, the threshold worker that is indifferent between getting education or not is defined by:

$$(1-c_A)R_A p_A = 1$$

The educated workers, those whose cost of education is below the threshold $\tilde{c} < c_B$,must prefer to not migrate, which implies:

$$(1-\tilde{c})^2 R_B p_B \geq (1-\kappa-\tilde{c})^2 R_A p_A \quad \forall \tilde{c} \leq c_B$$

Evaluated at $\tilde{c} = 0$ and using the indifference condition, this implies $\frac{R_B}{R_A} \geq (1-\kappa)^2$.

Combining the three indifference conditions and this inequality, we find

$$c_B = 1 - \frac{R_A}{R_B}(1-\kappa)^2(1-c_A) \geq 1-(1-c_A) = c_A$$

Market clearing in the two countries implies:

$$p_B = \frac{2(1-c_B-M)}{R_B(1-(1-c_B)^2)}$$
$$p_A = \frac{2(1-c_A+M)}{R_A(1-(1-c_A)^2)}$$

Together with the indifference conditions and $\frac{R_B}{R_A} \geq (1-\kappa)^2$, this implies:

$$\frac{1 - c_A + M}{1 - (1 - c_A)^2} \leq \frac{1 - c_B - M}{1 - (1 - c_B)^2}$$

Now, observing that $\frac{1 - x}{1 - (1 - x)^2}$ is decreasing on $[0, 1]$ and $c_A < c_B$, this implies $M < 0$, a contradiction. Thus there can never be an equilibrium in which only uneducated workers migrate. □

Bibliography

Joelle Abramowitz. Assisted reproductive technology and womens timing of marriage and childbearing. *Journal of Family and Economic Issues*, 38(1): 100–117, 2017.

Ajay Agrawal, Devesh Kapur, John McHale, and Alexander Oettl. Brain drain or brain bank? the impact of skilled emigration on poor-country innovation. *Journal of Urban Economics*, 69(1):43–55, 2011.

Mark Aguiar and Erik Hurst. Measuring trends in leisure: the allocation of time over five decades. *The Quarterly Journal of Economics*, 122(3):969–1006, 2007.

Ufuk Akcigit, Salome Baslandze, and Stefanie Stantcheva. Taxation and the international mobility of inventors. *American Economic Review*, 106(10): 2930–81, 2016.

Christoph Albert and Joan Monras. Immigrants' residential choices and their consequences. 2018.

Alberto Alesina, Andrea Ichino, and Loukas Karabarbounis. Gender-based taxation and the division of family chores. *American Economic Journal: Economic Policy*, 3(2):1–40, 2011.

Milena Almagro and Tomas Dominguez-Iino. Location sorting and endogenous amenities: Evidence from amsterdam. 2019.

Martha J Bailey. More power to the pill: The impact of contraceptive freedom on women's life cycle labor supply. *The quarterly journal of economics*, 121 (1):289–320, 2006.

Simcha Barkai. Declining labor and capital shares. *Stigler Center for the Study of the Economy and the State New Working Paper Series*, 2, 2016.

Nathaniel Baum-Snow and Daniel Hartley. Accounting for central neighborhood change, 1980-2010. *Journal of Urban Economics*, page 103228, 2019.

Michel Beine, Frédéric Docquier, and Hillel Rapoport. Brain drain and economic growth: theory and evidence. *Journal of development economics*, 64 (1):275–289, 2001.

R Albert Berry and Ronald Soligo. Some welfare aspects of international migration. *Journal of political economy*, 77(5):778–794, 1969.

Marianne Bertrand, Claudia Goldin, and Lawrence F Katz. Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, 2(3):228–55, 2010.

Jagdish Bhagwati and Koichi Hamada. The brain drain, international integration of markets for professionals and unemployment: a theoretical analysis. *Journal of Development Economics*, 1(1):19–42, 1974.

Marianne Bitler and Lucie Schmidt. Health disparities and infertility: impacts of state-level insurance mandates. *Fertility and sterility*, 85(4):858–865, 2006.

Marianne P Bitler and Lucie Schmidt. Utilization of infertility treatments: the effects of insurance mandates. *Demography*, 49(1):125–149, 2012.

Dan A Black, Natalia Kolesnikova, and Lowell J Taylor. Why do so few women work in new york (and so many in minneapolis)? labor supply of married women across us cities. *Journal of Urban Economics*, 79:59–71, 2014.

Richard Blundell, Thomas MaCurdy, and Costas Meghir. Labor supply models: Unobserved heterogeneity, nonparticipation and dynamics. *Handbook of econometrics*, 6:4667–4775, 2007.

Kasey Buckles. Stopping the biological clock: Infertility treatments and the career% family tradeoff, obu dissertation. 2005.

Kasey Buckles. Understanding the returns to delayed childbearing for working women. *American Economic Review*, 98(2):403–07, 2008.

Nadia Campaniello. The causal effect of trade on migration: Evidence from countries of the euro-mediterranean partnership. *Labour Economics*, 30: 223–233, 2014.

Jeffrey Capizzano. The hours that children under five spend in child care: Variation across states. 2000.

Francesca Carta and Marta De Philippis. You've come a long way, baby. husbands' commuting time and family labour supply. *Regional Science and Urban Economics*, 69:25–37, 2018.

Elizabeth M Caucutt, Nezih Guner, and John Knowles. Why do women wait? matching, wage inequality, and the incentives for fertility delay. *Review of Economic Dynamics*, 5(4):815–855, 2002.

Janice Compton and Robert Pollack. Why are power couples increasingly located in large cities? *NBER Working Paper*, 10918, 2004.

Dora L Costa and Matthew E Kahn. Power couples: changes in the locational choice of the college educated, 1940–1990. *The Quarterly Journal of Economics*, 115(4):1287–1315, 2000.

Victor Couture and Jessie Handbury. Urban revival in america, 2000 to 2010. Technical report, National Bureau of Economic Research, 2017.

Victor Couture, Cecile Gaubert, Jessie Handbury, and Erik Hurst. Income growth and the distributional effects of urban spatial sorting. Technical report, National Bureau of Economic Research, 2019.

Federico Curci and Hasin Yousaf. Gentrifying cities, amenities and income segregation: evidence from san francisco. 2020.

H David, Christopher J Palmer, and Parag A Pathak. Gentrification and the amenity value of crime reductions: Evidence from rent deregulation. Technical report, National Bureau of Economic Research, 2017.

Morris A Davis and François Ortalo-Magné. Household expenditures, wages, rents. *Review of Economic Dynamics*, 14(2):248–261, 2011.

Corrado Di Maria and Piotr Stryszowski. Migration, human capital accumulation and economic development. *Journal of Development Economics*, 90 (2):306–313, 2009.

Rebecca Diamond. The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000. *American Economic Review*, 106(3):479–524, 2016.

Avinash Dixit and Victor Norman. *Theory of international trade: A dual, general equilibrium approach*. Cambridge University Press, 1980.

Frédéric Docquier and Hillel Rapoport. Globalization, brain drain, and development. *Journal of Economic Literature*, 50(3):681–730, 2012.

Frédéric Docquier, B Lindsay Lowell, and Abdeslam Marfouk. A gendered assessment of highly skilled emigration. *Population and Development Review*, 35(2):297–321, 2009.

David Domeij and Martin Floden. The labor-supply elasticity and borrowing constraints: Why estimates are biased. *Review of Economic dynamics*, 9(2): 242–262, 2006.

Lena Edlund, Cecilia Machado, and Maria Micaela Sviatschi. Gentrification and the rising returns to skill. Technical report, National Bureau of Economic Research, 2015.

Jan Eeckhout, Roberto Pinheiro, and Kurt Schmidheiny. Spatial sorting. *Journal of Political Economy*, 122(3):554–620, 2014.

Ingrid Gould Ellen, Keren Mertens Horn, and Davin Reed. Has falling crime invited gentrification? *Journal of Housing Economics*, 46:101636, 2019.

Andrés Erosa, Luisa Fuster, Gueorgui Kambourov, and Richard Rogerson. Hours, occupations, and gender differences in labor market outcomes. 2017.

Lidia Farré, Jordi Jofre-Monseny, and Juan Torrecillas. Commuting time and female labor force participation. 2019.

Claudia Goldin. A grand gender convergence: Its last chapter. *American Economic Review*, 104(4):1091–1119, 2014.

Claudia Goldin and Lawrence F Katz. The power of the pill: Oral contraceptives and womens career and marriage decisions. *Journal of political Economy*, 110(4):730–770, 2002.

Herbert B Grubel and Anthony D Scott. The international flow of human capital. *The American Economic Review*, 56(1/2):268–274, 1966.

Nezih Guner, Remzi Kaygusuz, and Gustavo Ventura. Taxation and household labour supply. *The Review of economic studies*, 79(3):1113–1149, 2012.

Barton H Hamilton and Brian McManus. The effects of insurance mandates on choices and outcomes in infertility treatment markets. *Health economics*, 21(8):994–1016, 2012.

Mattias Hoelzlein. Two-sided sorting and spatial inequality in cities. 2019.

Jackelyn Hwang and Jeffrey Lin. What have we learned about the causes of recent gentrification? *Cityscape*, 18(3):9–26, 2016.

Susana Iranzo and Giovanni Peri. Migration and trade: Theory with an application to the eastern–western european integration. *Journal of International Economics*, 79(1):1–19, 2009.

Tarun Jain, Bernard L Harlow, and Mark D Hornstein. Insurance coverage and outcomes of in vitro fertilization. *New England Journal of Medicine*, 347(9):661–666, 2002.

Sari Pekkala Kerr, William Kerr, Çağlar Özden, and Christopher Parsons. Global talent flows. *Journal of Economic Perspectives*, 30(4):83–106, 2016.

William R Kerr. Ethnic scientific communities and international technology diffusion. *The Review of Economics and Statistics*, 90(3):518–537, 2008.

Mark R Killingsworth and James J Heckman. Female labor supply: A survey. *Handbook of labor economics*, 1:103–204, 1986.

Sarah Kroeger and Giulia La Mattina. Assisted reproductive technology and womens choice to pursue professional careers. *Journal of Population Economics*, 30(3):723–769, 2017.

Paul Krugman. Increasing returns and economic geography. *Journal of political economy*, 99(3):483–499, 1991.

Ilyana Kuziemko, Jessica Pan, Jenny Shen, and Ebonya Washington. The mommy effect: Do women anticipate the employment effects of motherhood? Technical report, National Bureau of Economic Research, 2018.

Sanghoon Lee and Jeffrey Lin. Natural amenities, neighbourhood dynamics, and persistence in the spatial distribution of income. *The Review of Economic Studies*, 85(1):663–694, 2018.

Henri Leridon. Can assisted reproduction technology compensate for the natural decline in fertility with age? a model assessment. *Human reproduction*, 19(7):1548–1553, 2004.

Matilde P Machado and Anna Sanz-de Galdeano. Coverage of infertility treatment and fertility outcomes. *SERIEs*, 6(4):407–439, 2015.

James R Markusen and Lars EO Svensson. Trade in goods and factors with international differences in technology. *International Economic Review*, pages 175–192, 1985.

Aaditya Mattoo, Ileana Cristina Neagu, and Çağlar Özden. Brain waste? educated immigrants in the us labor market. *Journal of Development Economics*, 87(2):255–269, 2008.

Karin Mayr and Giovanni Peri. Brain drain and brain return: theory and application to eastern-western europe. *The BE Journal of Economic Analysis & Policy*, 9(1), 2009.

Carol McAusland and Peter Kuhn. Bidding for brains: Intellectual property rights and the international migration of knowledge workers. *Journal of Development Economics*, 95(1):77–87, 2011.

Rachel McCulloch and Janet L Yellen. Factor mobility, regional development, and the distribution of income. *Journal of Political Economy*, 85(1):79–96, 1977.

Jane Menken, James Trussell, and Ulla Larsen. Age and infertility. *Science*, 233(4771):1389–1394, 1986.

Amalia R Miller. The effects of motherhood timing on career path. *Journal of population economics*, 24(3):1071–1100, 2011.

John Odland and Mark Ellis. Variations in the labour force experience of women across large metropolitan areas in the united states. *Regional Studies*, 32(4):333–347, 1998.

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

Stuart S Rosenthal and William C Strange. Agglomeration and hours worked. *The review of economics and statistics*, 90(1):105–118, 2008.

Esteban Rossi-Hansberg, Pierre-Daniel Sarte, and Felipe Schwartzman. Cognitive hubs and spatial redistribution. 2019.

Steven Ruggles, Sarah Flood, Ronald Goeken, Josiah Grover, Erin Meyer, Jose Pacas, and Matthew Sobek. Ipums usa: Version 8.0 [dataset]. *Minneapolis, MN: IPUMS*, 2018.

Clara Santamaria. Small teams in big cities: Inequality, city size, and the organization of production. 2019.

Lucie Schmidt. Effects of infertility insurance mandates on fertility. *Journal of health economics*, 26(3):431–446, 2007.

Yichen Su. The rising value of time and the origin of urban gentrification. *Available at SSRN 3216013*, 2018.

Ahmed Tritah. The brain drain between knowledge-based economies: the european human capital outflow to the us. *Économie internationale*, (3): 65–107, 2008.

Boukje M van Noord-Zaadstra, Caspar W Looman, Hans Alsbach, J Dik Habbema, Egbert R te Velde, and Jan Karbaat. Delaying childbearing: effect of age on fecundity and outcome of pregnancy. *Bmj*, 302(6789):1361–1365, 1991.

Clare Ward and Angela Dale. Geographical variation in female labour force participation: an application of multilevel modelling. *Regional Studies*, 26 (3):243–255, 1992.

Elizabeth Ty Wilde, Lily Batchelder, and David T Ellwood. The mommy track divides: The impact of childbearing on wages of women of differing skill levels. Technical report, National Bureau of Economic Research, 2010.

Kar-yiu Wong. Are international trade and factor mobility substitutes? *Journal of International Economics*, 21(1-2):25–43, 1986.

Kar-yiu Wong and Earl L Grinols. *International trade in goods and factor mobility*. MIT press Cambridge, MA, 1995.