THÈSE de DOCTORAT

de l'UNIVERSITÉ TOULOUSE CAPITOLE

Présentée et soutenue par

Monsieur Sébastien MONTPETIT

Le 5 juillet 2024

Essais sur l'offre de travail des femmes dans les pays développés

École doctorale : **TSE Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche : **TSE-R - Toulouse School of Economics - Recherche**

Thèse dirigée par *Monsieur Philippe BONTEMS et Monsieur François POINAS*

Composition du jury

Rapporteure : Mme Marie CONNOLLY
Rapporteur : M. Jean-William LALIBERTé
Examinatrice : Mme Marie-Louise LEROUX
Directeur de thèse : M. Philippe BONTEMS
Co-directeur de thèse : M. François POINAS

Essays on Female Labor Force Participation in Developed Countries

Sébastien MONTPETIT

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics
Toulouse School of Economics

June 15, 2024

Summary

Over the past century, the rise in women's participation in the economy has been one of the most significant transformations in the labor market. This thesis presents three essays studying the trade-offs women face in their participation choices and their consequences for social welfare. The first chapter focuses on universal childcare provision and its positive impacts on mothers' labor supply while the second and third chapters investigate the joint decision of integrating economically and preserving their culture for Muslim women.

In the first chapter, co-authored with Pierre-Loup Beauregard and Luisa Carrer, we evaluate the welfare effect of universal childcare provision. Leveraging the introduction of universal low-fee daycare in Québec in 1997 and novel data on daycare coverage rates within Québec, we show that the positive impacts on maternal labor supply and childcare use are larger in areas where daycare expanded more. Thus, childcare availability, rather than just the price decrease, is also responsible for the observed behavioral responses. In the second part of the chapter, we estimate the benefit-to-net-cost ratio of the policy while notably taking into account its non-marginal nature. We estimate mothers' utility gains using a model of maternal labor supply and childcare choices, incorporating non-pecuniary benefits for mothers, such as non-monetary costs of childcare use, and childcare availability. Structural estimates indicate that mothers' benefits are more than 3.5 dollars per dollar of net government spending – more than twice that obtained when solely focusing on earnings gains. As such, our findings suggest that non-pecuniary benefits for mothers are a key component of welfare gains of universal policies. Counterfactual simulations suggest that channelling more resources towards opening spots, rather than lowering prices, could have led to even larger social returns.

In the second chapter, Antoine Jacquet and I study the relationship between veiling behavior and economic participation using the largest sample of Muslim women in France. We demonstrate a significant negative relationship between veiling and economic participation, which contrasts with the existing economic theory of veiling in Muslim-majority countries. We show that a model which also accounts for reduced economic opportunities for veiled women is consistent with this finding. We then develop and estimate a discrete-choice model of veiling and economic participation to disentangle the various motivations behind the joint decision. Our results indicate that veiled women are less active not due to religious preferences, but rather because their benefits of economic participation are lower. Additionally, our results emphasize the significance of personal religious motives in the decision to veil, rather than community-based religious pressure. This calls into question the rhetoric used to justify policies that restrict the wearing of religious symbols in France.

In the third chapter, I estimate the effect of prohibiting the wearing of the Islamic veil for pupils on educational attainment of Muslim women. In a difference-in-difference analysis, I find that the directive to school principals to ban the veil in French schools in 1994 induced a large decline in high-school completion rates of Muslim women. There is further evidence that the effect on the intensive margin of education lasts in the medium-run. The data suggests that the ban operates through increased experiences of discrimination against Muslims and mistrust of the French school rather than through a change in Muslim parents' investments into their daughters' education. I show how using an inappropriate measure of the treatment group as in previous work substantially alters conclusions on the impacts of the policy. In the long-run, cohorts affected by the ban display lower levels of religiosity.

Résumé

Dans le dernier siècle, la hausse du taux de participation des femmes dans l'économie est l'une des transformations les plus importantes du marché du travail. Cette thèse présente trois essais étudiant les compromis auxquels les femmes doivent faire face dans leur choix de participer ainsi que leurs conséquences sur le bien-être social. Le premier chapitre se penche sur les défis pour les mères de jeunes enfants à trouver un équilibre entre le travail et la garde d'enfant alors que les second et troisième chapitres examinent la décision conjointe de l'intégration économique et de la préservation de leur culture pour les femmes musulmanes.

Dans le premier chapitre, corédigé avec Pierre-Loup Beauregard et Luisa Carrer, nous évaluons l'effet de la provision universelle de services de garde sur le bien-être social. Tirant avantage de l'introduction universelle de services de garde à bas prix au Québec en 1997 et de nouvelles données sur les taux de couverture en garderies au niveau local, nous démontrons que les impacts sur l'offre de travail des mères et l'utilisation des services sont plus grands dans les régions où l'offre de garde a augmenté davantage. Ainsi, la disponibilité de places est en partie responsable des réponses observées, et non seulement la baisse des prix. Dans la seconde partie du chapitre, nous estimons le ratio bénéfices-coûts nets de la politique, en tenant notamment compte de sa nature non-marginale. Nous estimons les gains d'utilité des mères à l'aide d'un modèle d'offre de travail et de choix de garde, incorporant les bénéfices non-pécuniers pour les mères, tels que les coûts non-monétaires de l'usage de services de garde et la disponibilité des places. Les estimés structurels indiquent que les bénéfices pour les mères sont plus de 3,5 dollars par dollar net de dépense publique – plus du double de ceux obtenus en se limitant aux gains de revenus. Ainsi, nos résultats suggèrent que les bénéfices non-pécuniers pour les mères sont une composante clé des gains de bien-être des politiques universelles. Des simulations contrefactuelles suggèrent qu'investir davantage dans l'ouverture de places, plutôt que de baisser les prix, aurait généré des retours sur investissement encore plus grands.

Dans le second chapitre, Antoine Jacquet et moi étudions la relation entre le port du voile et la participation économique à l'aide du plus grand échantillon de femmes musulmanes en France. Nous démontrons une relation négative forte entre le port du voile et la participation économique, contraire à la théorie économique existante dans les pays à majorité musulmane. Nous montrons qu'un modèle qui tient également compte des opportunités économiques réduites pour les femmes voilées est cohérent avec ce résultat. Ensuite, nous développons et estimons un modèle à choix discrets de port du voile et de participation pour distinguer les motifs derrière cette décision conjointe. Nos résultats indiquent que les femmes voilées sont moins actives, non pas en raison de préférences religieuses, mais plutôt parce que leurs bénéfices à participer sont inférieurs. De plus, nos résultats suggèrent l'importance des motifs religieux personnels, plutôt que celle des pressions communautaires. Cela remet en question la rhétorique utilisée pour justifier les

politiques restreignant le port de symboles religieux en France.

Dans le troisième chapitre, j'estime l'effet de l'interdiction du port du voile pour les élèves sur l'éducation des musulmanes. Dans une analyse de différence-en-différences, je trouve que la directive aux responsables d'écoles de bannir le voile dans les écoles françaises en 1994 a induit un grand déclin dans le taux de complétude du secondaire des femmes musulmanes. L'effet sur la marge intensive de l'éducation persiste à moyen terme. Les données suggèrent que l'interdiction opère à travers une augmentation des expériences de discrimination envers les musulmanes et d'une perte de confiance envers l'école française, et non via des changements dans les investissements des parents dans l'éducation de leurs filles. Je montre que l'utilisation d'un proxy inapproprié du traitement dans des études précédentes a une grande influence sur les conclusions obtenues. À long terme, les cohortes affectées affichent une plus faible religiosité.

Acknowledgements

This thesis would not have come to life without the precious help, encouragements, and support from countless individuals.

First and foremost, I am indebted to many professors at the Toulouse School of Economics who have taken some of their precious time to chat and help me progress in my research. I am extremely grateful for the continuous support and availability of my dissertation advisors, Mohamed Saleh, François Poinas, and Philippe Bontems. I thank Professor Saleh for his useful guidance before he left for another institution. His advice on my projects on veiling among Muslim women has been excellent. Mohamed, you are such a kind and nice person. I have very happy that our paths have crossed. After Professor Saleh left, and given that my thesis was taking a different focus, I decided to work with other Faculty. Professor Philippe Bontems, with whom I have been collaborating for organizing the ENTER exchange network, quickly showed great interest in the paper that would become the first chapter of this thesis. I rapidly knew that things would work well with you, Professor Bontems. I am especially grateful for your incredible availability, which proved extremely important during the Job Market period. Your reading (in great detail) of my first chapter substantially helped improving it along the way.

François Poinas has co-advised my work since the beginning of my PhD journey. François, I cannot thank you enough for continuously finding the good words to encourage me and challenge me when necessary. I will especially always remember the Summer of 2023, when I was struggling with the estimation in the Canadian Research Data Centers and the Job Market approaching fast. You offered me a lot of your time for Zoom calls to give me weekly advice on implementing the model estimation. Without your amazing advice, I would not have learned as fast and the first chapter would not be what it is today. Your great qualities as a communicator, teacher, and advisor are certainly among the reasons behind your promotion as Dean of Education, which is absolutely deserved.

Several other TSE and IAST professors have contributed to my successful PhD journey. I thank, in particular, in no particular order, Angelique Acquatella, Matteo Bobba, Margot Dazey, Olivier De Groote, Victor Gay, Christian Hellwig, Paul Seabright, and Karine Van Der Straeten for insightful discussions on my work. I am even more grateful for the precious advice I received from the Faculty managing the Job Market period at TSE, namely Fabrice Collard, Ulrich Hege, and Anouch Missirian. Ulrich, your availability to talk about my ups, downs, anxieties, and many other emotions was priceless. You helped me stay rational and keep my hopes up in that particularly difficult period. I thank you so much.

I also benefited from the great advice of many other scholars from other institutions. For their useful comments at various stages, I thank Jean-Paul Carvalho, Gordon Cleveland, Pierre Lefebvre, Thomas Lemieux, Philip Merrigan, Kevin Milligan, Timea Laura Molnár, and Alain Paquet. I am also particularly indebted to Nicole Fortin, Marie Connolly, and Marie-Louise Leroux, who generously helped me obtain invitations to work with Canadian data.

I also thank Laurence Delorme, Ludmila Namolovan, Louise Stachan, and Corinne Vella of the administrative staff at TSE, for their precious help to navigate the bureaucracy and all those tasks that students do not typically enjoy. Mrs. Strachan, the Job Market period would have been so much more complex without your incredible efficacy. Thank you so much! I also recognize the generous help I have received from Statistics Canada staff in the Canadian Research Data Centres, namely Wendy Kei, Chunling Fu, Pascal Doray-Demers, Loïc Courtemanche, and Anne-Charlotte Latour.

Such a tedious and often suffering adventure that is the PhD would be simply impossible to accomplish without the continuous support from and precious moments spent with friends and family. To my precious friends of the Fabulous France (FF) Club, Oscar "Wey" Fentanes Tellez and José Ignacio "Weón" Hernández, for all the great laughter, the incredible FF moments, the (too) many lunches at the RU, and the great inside jokes, I love you guys. Luisa Carrer, also a great friend and an amazing colleague, I thank you for your kind encouragements during the difficult Job Market period and for your great ideas which became the seeds of the first chapter of this thesis. Jean-François Fournel, an ex-colleague at McGill who I was very lucky to have around again in Toulouse also played an important role in my PhD journey. For the many random discussions, afternoon drinks, French poutines, and advice for the Job Market, thanks J-F! I also thank the many other TSE and IAST students with whom I've had some quality time and discussions: Basile Dubois (and Margot), Antoine Jacquet, Marine Coinon, Jonas Gathen, Moritz Loewenfeld, Esteban Muñoz Sobrado, José Alfonso Muñoz Alvarado, Anaïs Fabre, Federica Lo Polito, Peter Neis, Yifan Zhang, Max Lesellier, Oscar Vilargunter, Léa Bignon, Lisa Botbol, and Guillem Foucault Llopart. The TSE community was simply the best I could hope for! The same thoughts go to some PhD candidates at other institutions, in particular to Sarah Vincent, Pierre-Loup Beauregard, and Katherina Thomas. In addition, I thank the "Team Airbnb", my flatmates in Vancouver when I visited the Department of Economics at UBC, Daniel, Ricardo ("Weón"), Josua, Marco, Ryu, and Kanako for helping me making the most out of this experience. Another friend I must mention is Clémentine Gand. She has been an incredible friend for my partner Camille and I. Having the opportunity to talk about topics of naturalists was very helpful to laugh and think about something else than economics.

Despite the distance separating us, I always felt supported by my family and friends oversees. To my father, Martin, I am deeply touched by how proud you are of my accomplishments. I thank you very much for every little thing you did that contributed to this thesis: the multiple rides to/from the airport, the vacations when I needed to relax, and your generosity. To my mother, Isabelle, I am especially grateful for your kind words in the most difficult moments. You helped me quickly get back on my feet when I needed it most. To my parents-in-law, Martin, Linda, and Sylvie, you have been a second family to me. Thank you for all the amazing moments we have spent together and for your unequalled hospitality. I also thank Guillaume, Myriam, Chantal, Samuel, Myliann, Brigitte, Sylvain, and Les Boys from high-school for the precious time spent together in Québec, which helped changing my mind when needed most.

Last, to my life partner, my love, Camille. For the both of us, this experience had its ups and downs. Nevertheless, your presence with me and your love made this whole PhD journey so much easier. I am so lucky to have you in my life.

Contents

List of Figures

List of Tables

Chapter 1

General introduction

This thesis presents three essays examining economic issues related to the labor supply of women in developed countries. Over the past century, the rise in women's participation in the economy has been one of the most significant transformations in the labor market. This "quiet revolution" (Goldin, 2006) has brought new opportunities and challenges for households and policymakers. In this vein, the three essays which compose this thesis investigate some key trade-offs faced by women in their participation choices and their consequences for social welfare, namely childcare duties and conflicts with traditional culture. The thesis thus mostly contributes to the fields of gender and applied microeconomics. In terms of methods, Chapter 2 further contributes to the literature in public economics developing empirical approaches for welfare analysis. Chapters 3 and 4 also provide insights relevant to the economics of culture.

One key societal issue impacted by the evolution in gender norms is childcare since childrearing responsibilities have traditionally fallen upon women (Gauthier et al., 2004). Public investments in childcare have been at the heart of discussions on family policy in North America over the past three decades. Until recently, their economic benefit for society was disputed, mainly because of their large costs. Recent research in economics, however, has provided a strong momentum in favor of public intervention in childcare markets. In particular, recent evidence shows that the three main programs implemented in the United States generated large returns on initial public investment. The HighScope Perry Preschool and the Carolina Abecedarian projects – often referred to as "model" preschool interventions – were intensive and high-quality small-scale programs aimed at enhancing life outcomes of participants using established scientific protocols. Notable evaluations of these programs by Heckman et al. (2010) and García et al. (2020, 2023) suggest very high rates of return for society stemming from both increased maternal employment and various positive effects on children. Similarly, Head Start, a large-scale program offering subsidized childcare all over the country provided sizable societal benefits exceeding its cost (Kline and Walters, 2016; Bailey et al., 2021). Hendren and Sprung-Keyser (2020), in a comparative welfare analysis across policy domains, substantiate that these programs are among the most welfare-improving policies on record. Those programs also have another important feature in common in that they target disadvantaged households, in those cases disadvantaged Black American families.

Proposals to expand childcare reforms are often motivated by encouraging results from such targeted interventions. However, these might not necessarily extend to programs open to all, in large part because the composition of the treated population is substantially different (Baker, 2011; List et al., 2021; List, 2022; Duncan et al., 2023). By definition, universal reforms grant access to both the disadvantaged and better-off families. Kottelenberg and Lehrer (2017) and Cornelissen et al. (2018), among others, show that while the disadvantaged are likely to benefit from childcare policies, children in better-off families might not. As such, the current evidence base on universal reforms is mixed, with policy impacts ranging from positive to negative.

In Chapter 2, we rather focus on another overlooked issue in welfare evaluations of universal childcare reforms. That is, we take into account the fact that the reform is a *non-marginal* change in the economic environment. Our approach contrasts with standard benefit-cost analysis, which typically relies on sufficient-statistic approaches to empirical welfare analysis. These methods have the advantage of allowing for welfare evaluations solely using causal impacts of an implemented policy on key outcomes, especially beneficiaries' earnings. Given the recent developments in causal estimation in econometrics, this framework has rapidly gained in popularity among empirical economists. However, the sufficient-statistic method hinges on the assumption that the policy change should be small from the beneficiaries' point-of-view. In other words, beneficiaries should only react to the policy at the margin and do not re-optimize behavior as a response to the reform. While it is well-established in the public finance literature that this approach is biased for large policy changes, we document that applying this method to non-marginal policies as if they were infinitesimal is common practice. The key challenge is that, for large reforms, one cannot simply express the welfare effect of a policy as a fiscal externality, i.e. the impact of a policy on the government's budget.

In a separate literature, scholars have long studied welfare impacts of hypothetical changes in policy using estimated structural models, often referred to as *ex ante* program evaluation. More recently, a growing literature has shown how this approach can be fruitfully combined with *ex post* program evaluation. Todd and Wolpin (2023) and Buera et al. (2023) discuss how combining these methods allows to both more credibly identify structural models and enhance scholars' understanding of policy impacts beyond the limitations of reduced-form causal analysis. In Chapter 2, we leverage these insights to provide the first empirical evaluation of the extent of bias in sufficient-statistic estimators when applied to non-marginal changes in policy. To do so, we compare two estimators of welfare: a sufficient-statistic estimator following the approach described above and a structural estimator that allows for non-marginal responses by recipients. For the latter, we specify and estimate an economic model of maternal labor supply and childcare choices to infer mothers' willingness-to-pay for the policy change. Our model integrates the presence of supply shortages in a Chaparro et al. (2020) model of a mother who has to meet childcare and time constraints and cares about the human capital accumulation of the child. This strategy also allows us to incorporate non-pecuniary benefits for mothers. In particular, increased availability reduces non-monetary costs associated with childcare use, such as time spent commuting to the caregiver and search effort to find a place when supply is limited.

Our empirical setting is a childcare reform implemented in 1997 in the Canadian province of Québec. The program introduced universal subsidized preschool daycare in a new set of regulated settings and increased daycare supply in the province. Prior to the reform, Québec was lagging behind the other provinces in terms of maternal labor-force participation, which was one of the main motives for implementing the policy at the time. Impacts of this policy on a variety of outcomes have been extensively studied. Two main patterns emerge: positive and large effects on childcare utilization and on maternal labor supply and adverse consequences on child development on average (see Baker et al., 2008; Lefebvre and Merrigan, 2008, among others). As of now, the literature on this policy, while recognizing the increase in daycare supply, has put much emphasis on the substantial price reduction induced by the generous subsidies. This might be in part due to the fact that it was not possible to disentangle between the supply increase and the price channels in previous causal analyses.

In a first step, we provide new insights on the impacts of the policy change on economic behavior and how they interact with the daycare supply increase. We digitize new data on daycare coverage rates at the administrative-region level within Québec from a series of annual management reports of the Ministry of the Family. Using this data and a difference-in-differences design, we show that the positive impacts on maternal labor supply and childcare use about 40% larger in areas where daycare expanded more. Thus, childcare availability, rather than just the price decrease, is also responsible for the observed behavioral responses. In a heterogeneity analysis, we find that the supply increase can help explain increased labor supply of low-educated mothers despite the financial incentives being lower for them. Indeed, to finance the policy, the Québec government abolished a refundable childcare credit that was rapidly decreasing with household income. However, even if the financial incentives were low, our results suggest that, for mothers without a post-secondary qualification, access to a space was the main incentive to take up employment. In a separate causal analysis, we assess whether the negative short-term impacts of children's health and behavior documented by Baker et al. (2008) extend to educational outcomes in the long-run. We find no evidence of such negative consequences in the long-run, which mirrors the surprisingly parallel trends in educational attainment across Canadian provinces over decades.

In the second part of the paper, we calculate the policy's benefit-net cost ratio, also called the marginal value of public funds (MVPF). We compare the welfare estimates obtained from the reduced-form analysis following the sufficient-statistic approach to our structural estimator accounting for non-pecuniary benefits and the non-marginal nature of the policy. We first estimate mothers' earnings gains to obtain a benchmark value of the policy's benefits. We use a non-linear difference-in-differences model to estimate earnings gains across mothers' income distribution. Taking into account the effect of this heterogeneity on fiscal returns to the government, we estimate the implied fiscal externality. We find that mothers' earnings gains amount to 1.42 dollars per dollar of net government spending, a relatively modest MVPF compared to targeted preschool interventions studied in Hendren and Sprung-Keyser (2020).

In a second step, we move beyond earnings gains and estimate mothers' willingness-to-pay structurally. We find that accounting for non-monetary benefits and for the non-marginal nature of the policy yields a WTP more than twice as large as when considering earnings gains alone. Our estimates further reveal that only about 37% of utility gains are attributed to labor-market choices, implying that most of the increase in welfare stems from non-pecuniary benefits. Our structural estimator yields an MVPF estimate of 3.56. Therefore, our findings suggest that universal preschool policies do have the potential to generate substantial social returns. Focusing on earnings gains only would substantially underestimate the policy's benefits.

Finally, counterfactual analyses further confirm the crucial role of availability of daycare spaces for welfare. Our results suggest that much of the welfare gains are due to increased coverage so that, in the Québec context, higher welfare gains could be achieved by channeling more resources towards opening spots rather than lowering childcare fees as much.

In the second and third essays, we turn to a different trade-off that women face in their employment decisions, which is the potential conflict between the evolving gender norms and the desire to preserve their culture. We focus on a specific group for which this trade-off is particularly relevant, namely Muslim women in France. We contribute to a vast literature in the social sciences by studying a particularly salient decision in Muslim women's lives: whether or not to wear the Islamic veil. We approach this issue through two different angles. In Chapter 3, we focus on evaluating the economic cost associated with veiling in France as well as understanding the main motives behind the joint choice of veiling and economic participation. In Chapter 4, we analyze the economic consequences of veiling restrictions in France by exploiting a reform that gradually prohibited the wearing of conspicuous religious symbols in French schools. Taken together, these two essays aim at deepening our understanding of the obstacles to economic integration faced by Muslim women in secular countries.

The question of the Islamic veil, a burning issue in France since at least three decades, has been extensively studied in sociology and anthropology. This literature has uncovered a large number of potential reasons as to why women wear the Islamic veil despite expecting important costs to their integration into French society. When asked why they wear the veil, Muslim women mostly invoke religious duties (76%) and issues of safety (35%) (Institut Montaigne, 2016). Young women in particular mention "the difficulty to reconcile their families' demands with those of the society" (Khosrokhavar, 2004, p. 90). Other motives go from signaling piety to potential husbands, or even fashion (Patel, 2012), as well as affirming their distinction with the rest of society and to feel closer to their community of origin (Silhouette-Dercourt et al., 2019). Overall, the literature established that veiling is a heterogeneous and complex practice. However, in the public debate, the issue is often reduced to a "one-sided debate" with proponents of secular policies defending the idea that a "silent majority" of Muslim women are forced to wear the veil by their families or communities.

In Chapter 3, we test the validity of this assertion. While the current evidence from interviews and small-scale surveys contradicts this argument (IFOP 2019, Institut Montaigne 2016), these methodologies – on which most of the current evidence is based – often suffer from small sampling and representativeness issues. In this essay, we perform an in-depth descriptive analysis of the relationship between veiling and economic participation, using rich survey data over more than 3,000 Muslim women in France. This sample constitutes the largest source of data on Muslim women and their veiling practices in France that we are aware of. The richness of the survey not only allows us to account for an unusually large number of confounding factors, but also to measure individual religiosity, women's religious environment, and parents' investments in transmitting their religion. This strategy improves upon interview evidence because in the latter, respondents are typically aware that the topic of the interview is veiling behavior and are thus more susceptible to social desirability bias.

We first document important costs in the labor market for veiled women. Our results suggest that, even after accounting for differences in human capital, family structure, and religiosity, reporting to always wear a conspicuous religious symbol in public spaces is associated with a 23 percentage points decline in economic participation. This correlation is large and economically significant: in our preferred specification, it is equivalent to the effect of having an additional 1.4 children aged less than 4 years old. This result further survives an extensive set of robustness checks.

Such negative correlation contrasts with the existing economic theory of veiling in Muslim-majority countries. The seminal economic theory of veiling of Carvalho (2013) considers veiling as a technology available to Muslim women in order to alleviate the intrinsic and social costs of their integration. By providing a practical protection against opportunities to engage in religiously prohibited behaviors, veiling acts both as a commitment to oneself and as a signal of this commitment to others. According to this theory, therefore, in Muslim-majority contexts, the veil is a tool for Muslim women to participate in the economy while preserving their reputation. Shofia (2020) provides empirical evidence for this mechanism in a study of veiling among Indonesian schoolgirls. Her study supports a *positive* association between veiling and economic participation in Muslim-majority contexts. We show that a model which also accounts for reduced economic opportunities for veiled women is consistent with our findings in the Muslim-minority context and the sociological evidence in France.

Translating this extended model into a discrete-choice model of veiling and labor force participation, we disentangle the various motivations behind the joint decision to veil and to be economically active. Our results indicate that veiled women are less economically active not due to religious preferences, but rather because the benefits of economic participation are lower for women who veil compared to those who do not. This finding echoes previous findings in the literature regarding labor-market discrimination against individuals who signal their religious affiliation. Additionally, our results emphasize that personal religious motives are a stronger predictor than community-based or parental pressure in the decision to veil.

While we cannot identify a causal relationship in the second essay, in Chapter 4 I exploit an exogenous policy change to investigate the consequences of veiling restrictions on economic outcomes of Muslim women. This essay contributes to the literature on the economic integration of minority populations, and more specifically to the question of whether "integration" or "assimilation" policies yield the most desirable societal outcomes. Previous literature studying *assimilationist* regulations shows that these can unintentionally backfire. Fouka (2020) shows that German language prohibitions in U.S. schools after WWI made German-Americans less likely to volunteer in World War II and increased their cultural distance with the majority. On the contrary, *integration* policies, such as easier access to citizenship for immigrants in their host country were found to improve labor-market attachment and social integration of immigrants (Gathmann and Garbers, 2023). A notable exception is Dahl et al. (2022) who show that the introduction of automatic birthright citizenship in Germany had negative impacts on Muslim girls. They find that Muslim girls born soon after the policy have lower life satisfaction and self-esteem and are less socially integrated into German society. These results might suggest that Muslim girls, whom perhaps face stronger pressures from their peers to follow religious norms, might benefit from an assimilationist policy which would free them from these pressures. To fill this literature gap, in Chapter 4 I analyze whether a veil ban in public schools boosts or impedes economic integration of Muslim women in France using a difference-in-difference design.

In 1994, following the election of a right-wing government, the French minister of education, François Bayrou issued a circular asking school principals to prohibit conspicuous religious symbols worn by students. The directive was enshrined in law ten years later, but most schools had already adopted the ban after the directive was issued. The headscarf ban targets pupils at the primary and secondary levels, but does not apply to students attending college.

Two recent papers reach opposite conclusions on the effects of the French headscarf ban on educational attainment of girls of African origin. On the one hand, Abdelgadir and Fouka (2020) find that the 2004 ban depressed schooling outcomes of French girls of North-African origin. Their results suggest that the negative impact of the ban operates through increased perceptions of discrimination at school. On the other hand, Maurin and Navarrete-H. (2023) find that the Bayrou circular had a positive impact on girls of African origin. They find that the issuance of the circular is positively associated with other measures of social integration such as mixed marriages. One potential reason behind the fragility of these results is the fact that they do not use the same measures of treatment in their analysis, sometimes even within the same paper.

Incorporating the second wave of the rich survey used in Chapter 3 in which religion is observed, my results suggest that properly identifying the treatment group is crucial in this context. I show that the documented positive impact of the Bayrou circular holds for individuals of African origin, but that the impact on the actual treatment group (Muslim women) is of opposite sign. I find a very large short-term negative impact of the circular issuance on Muslim women's probability to have completed high school. The main point estimate suggests a decline in the high-school completion of about 25% of the pre-treatment mean.

The results suggest that measurement error is driving results in the two previous studies. In a heterogeneity analysis, I show that using the father's nationality at birth as a proxy likely captures positive impacts on other religious groups than Muslims. In addition, the data suggests that parental religious influence, the mechanism suggested in Maurin and Navarrete-H. (2023), is not the main channel at play. I rather find evidence consistent with Abdelgadir and Fouka (2020) in that affected cohorts are more likely to report experiences of discrimination due to their religion. Last, I find that in the long-run, treated Muslim women display lower levels of religiosity later in life rather than a strengthening of religious identities.

In the last chapter of the thesis, I conclude by summarizing the main lessons learned from the three essays and formulate recommendations for policymakers.

Chapter 2

A Welfare Analysis of Universal Childcare: Lessons From a Canadian Reform

Sébastien Montpetit   Pierre-Loup Beauregard*   Luisa Carrer†

Abstract

Recent research shows that early-childhood interventions targeted at disadvantaged households can yield large returns on initial public investment. However, the extent to which such benefits extend to universal programs remains an open question. Leveraging the introduction of universal low-fee daycare in Québec in 1997, we evaluate the welfare effect of universal childcare provision. First, using novel data on daycare coverage rates within Québec and a difference-in-differences design, we show that the positive impacts on maternal labor supply and childcare use are larger in areas where daycare expanded more. Thus, childcare availability, rather than just the price decrease, is also responsible for the observed behavioral responses. In the second part of the paper, we estimate the benefit-to-net-cost ratio of the policy while notably taking into account its non-marginal nature. We estimate mothers' utility gains using a model of maternal labor supply and childcare choices, incorporating non-pecuniary benefits for mothers, such as non-monetary costs of childcare use, and childcare availability. Structural estimates indicate that mothers' benefits are more than 3.5 dollars per dollar of net government spending – more than twice that obtained when solely focusing on earnings gains. As such, our findings suggest that non-pecuniary benefits for mothers are a key component of welfare gains of universal policies. Counterfactual simulations suggest that channelling more resources towards opening spots, rather than lowering prices, could have led to even larger social returns.

Keywords: universal childcare, daycare coverage, social welfare, sufficient statistics

JEL Codes: H43, J13, J22

*University of British Columbia. Email: pierreloup.beauregard@gmail.com.
†Toulouse School of Economics. Email: luisa.carrer@tse-fr.eu.

2.1 Introduction

Over the past century, the rise in female labor-force participation has been one of the most significant transformations in the labor market. This "quiet revolution" (Goldin, 2006) has presented challenges for parents with young children in balancing employment and family duties since childrearing responsibilities have traditionally fallen upon women (Gauthier et al., 2004). Public investments in childcare can help mothers face this trade-off, as improved access to out-of-home care alternatives may reduce their opportunity cost of employment.

Proposed childcare reforms are often motivated by existing evidence of large potential benefits of early-childhood programs targeted at disadvantaged families. Findings in this literature suggest substantial returns on initial public investments.[1] However, these results might not necessarily extend to programs open to all (Baker, 2011; List et al., 2021; List, 2022; Duncan et al., 2023). Evidence on universal reforms is mixed, with policy impacts ranging from positive to negative.[2] Moreover, universal programs require substantial public expenditures.[3] Because of their high costs and the considerable variability in estimated benefits across contexts, little is known about the social desirability of these policies. Nevertheless, as policymakers are considering expanding current programs, "decisions about government funding of preschool enrolment should be based on evidence that benefits exceed costs" (Duncan et al., 2023, p. 19).

In this paper, we assess the welfare impact of universal childcare provision by exploiting a major policy change in Québec in the late 1990s, which introduced universal subsidized preschool daycare and increased daycare supply in the province. This task is particularly challenging for two main reasons. On the one hand, improved access to childcare alternatives may impact mothers' welfare through various channels. Firstly,

[1] See, among others, Heckman et al. (2010); Kline and Walters (2016); García et al. (2020); Bailey et al. (2021); Algan et al. (2022); Barr and Gibbs (2022); García et al. (2023). A notable exception is Cascio (2023), who using variation in program features across states, finds no average impact of targeted preschool enrolment on children's test scores in the United States.

[2] Policy impacts vary substantially depending on the specific context studied. For example, in the case of maternal labor supply, effects depend to a large extent on the counterfactual childcare market (Kline and Walters, 2016; Karademir et al., 2023). In contexts with a high initial prevalence of informal care (by grandparents, siblings, etc.), studies find little impact on parents' labor supply (Cascio, 2009; Fitzpatrick, 2010; Havnes and Mogstad, 2011a,b; Chaparro et al., 2020; Kleven et al., 2021). On the contrary, policies that crowd-out parental care hours typically have positive impacts on maternal employment (Gelbach, 2002; Baker et al., 2008; Nollenberger and Rodriguez-Planas, 2015; Herbst, 2017; Carta and Rizzica, 2018; Hojman and Lopez Boo, 2022). Regarding child development, the estimated direction of impact appears to primarily depend on both program care quality and children's socio-economic status (Baker et al., 2008, 2019; Havnes and Mogstad, 2015; Kline and Walters, 2016; Kottelenberg and Lehrer, 2017; Felfe and Lalive, 2018; Cornelissen et al., 2018; Fort et al., 2020; Chaparro et al., 2020; Hojman and Lopez Boo, 2022; Akee and Clark, 2023). See Morrissey (2017), Hotz and Wiswall (2019), Evans et al. (2021), and Blau (2021) for reviews of the literature on subsidized childcare provision.

[3] For example, in his Covid Recovery Plan, US president Joe Biden announced $39 billion in investments specifically dedicated to childcare (The White House, 2021). Similarly, for the current Canadian expansion, which should bring down daily daycare fees to $10 per day by 2026, the Trudeau government already committed $30 billion in federal investments (Seward et al., 2023).

opting for non-maternal care frees up time for market work or leisure but also implies that the child is cared for in a different environment, affecting the child's human capital development. In addition, impacts on early child development might have long-term consequences on economic outcomes as children age.[4] Secondly, although mothers might enjoy time spent with their children, providing development-enhancing care can be exhausting (Chaparro et al., 2020). Thirdly, increased availability reduces non-monetary costs associated with childcare use, such as time spent commuting to the caregiver (Bravo et al., 2022; De Groote and Rho, 2023) and search effort to find a place when supply is limited. On the other hand, the policy under consideration is a *non-marginal* change in the economic environment. This implies that standard sufficient statistics for welfare benefits, which assume an infinitesimal change in policy, would be biased (Kleven, 2021). We overcome these challenges by estimating a tractable structural model of maternal time-allocation and childcare choices and using it to infer mothers' willingness-to-pay for the reform. This paper is the first to incorporate the various sources of benefits mentioned above and to take into account the non-marginal nature of the policy into a unified welfare analysis of an implemented reform.

In a first step, we provide new evidence on the policy's impacts. Using novel data on regional daycare coverage rates within Québec we manually digitized, we revisit established findings regarding the short-term effects of the reform. We provide several pieces of evidence suggesting that the increase in local daycare coverage has the characteristics of a supply shock and is not driven by mothers' demand. Leveraging this variation, we estimate the policy's heterogeneous effects by local daycare supply, employing an intent-to-treat (ITT) difference-in-differences approach that compares mothers of young children in Québec to their counterparts in the rest of Canada. Our findings indicate that the local expansion of daycare supply, not just the decrease in prices, is an important channel of impact on childcare use and maternal labor supply. In regions where daycare supply increased the most (defined as the top two terciles of daycare coverage expansion), maternal employment (childcare use) increased 67% more (38% more), even after controlling for regional-level covariates. A heterogeneity analysis further reveals that low-educated mothers, who had smaller financial incentives due to the abolition of a refundable childcare credit,[5] also respond in high-expansion regions. These results suggest that increasing local daycare supply is key for the effectiveness of preschool policies.

[4] See Almond et al. (2018) for a review of the literature on long-term impacts of childhood circumstances and Duncan et al. (2023) for a review of impacts of public investments in early childhood.

[5] To finance some of the policy's costs, the Québec government also implemented some changes to other family benefits. The most notable fiscal change was the abolition of a refundable childcare credit, which reimbursed a share of childcare costs to claimant families, for families using a low-fee space. The credit rate declined rapidly with family income (see Appendix Figure A.1). Thus, the net price of childcare declined substantially more for high-educated families. See Section 2.2.1 and Baker et al. (2005) for additional details on these changes.

Then, we estimate mothers' earnings gains to obtain a benchmark value of the policy's benefits. We use a non-linear difference-in-differences model to estimate earnings gains across mothers' income distribution. Taking into account the effect of this heterogeneity on fiscal returns to the government, we estimate the implied fiscal externality (i.e. the return to the government from additional earnings) and calculate the policy's marginal value of public funds (MVPF). The MVPF is the ratio of the beneficiaries' willingness-to-pay (WTP) for a policy over its net cost (of fiscal externalities) to the government. We estimate that mothers' earnings gains amount to 1.42 dollars per dollar of net government spending, a relatively modest MVPF compared to targeted preschool interventions studied in Hendren and Sprung-Keyser (2020).

Our reduced-form analysis extends to examining the impact of the policy on eligible children's educational attainment later in life. We find that the negative impacts on child behavior documented by Baker et al. (2008, 2019) do not translate into worse economic outcomes later in life. Therefore, this evidence suggests the absence of negative fiscal impacts stemming from eligible children's economic outcomes in the long run.[6]

In a second step, we move beyond earnings gains and incorporate non-pecuniary benefits for mothers. To do so, we specify and estimate a model of maternal labor supply and childcare choices to infer mothers' WTP for the policy change. Our model integrates the presence of supply shortages in a Chaparro et al. (2020) model of a mother who has to meet childcare and time constraints and cares about the human capital accumulation of the child. The model captures key trade-offs families face such as to balance employment and care as well as to decide how much parenting effort to exert when at home.

To assess the validity of the model, we verify that it accurately replicates the ITT impacts on maternal labor supply and childcare use. We simulate the main features of the policy in our model, which is estimated on pre-reform data, and compare the simulated changes in mothers' choices to the reduced-form estimates. The model closely matches the key behavioral responses to the policy, supporting our structural assumptions on behavior. Additionally, our estimation algorithm leverages causal estimates from the first step to directly identify some key parameters of the model. In particular, we show how the Québec natural experiment can serve as an instrument for maternal-care hours to identify a (potentially) non-linear cost of parenting effort by adapting the identification approach of Chaparro et al. (2020) to this context.

We find that accounting for non-monetary benefits and for the non-marginal nature of the policy yields a WTP more than twice as large as when considering earnings gains alone. Our simulations reveal that only about 37% of utility gains are attributed

[6]Nevertheless, behavioral problems could impact the government budget through other channels. In robustness exercises, we notably calculate the potential costs associated with increased youth criminal activity found in Baker et al. (2019), using recent estimates of costs of crime that consider victimization costs and productivity losses. Given the relatively "benign" nature of typical juvenile crimes, this additional societal cost turns out to be relatively small compared to mothers' gains in this context.

to labor-market choices, implying that most of the increase in welfare stems from non-pecuniary benefits. Our structural estimator yields an MVPF estimate of 3.56. Therefore, our findings suggest that universal preschool policies do have the potential to generate substantial social returns. Focusing on earnings gains only would substantially underestimate the policy's benefits.

This result has implications beyond childcare policy. A recent body of literature shows that, under standard assumptions, monetary gains are a sufficient statistic for beneficiaries' WTP in the case of sufficiently small fiscal policy changes.[7] However, our results indicate that applying this method to non-marginal reforms, such as the Québec childcare policy, which we document to be common practice,[8] might significantly compromise welfare conclusions. While it is well-established in the public finance literature that this sufficient-statistic approach is biased for large policy changes, our results provide insight on the magnitude of this bias empirically. In cases of policies that entail large costs and substantial non-pecuniary benefits such as the Québec childcare reform, such an estimator can substantially underestimate welfare gains.

Finally, we use our structural model of behavior to perform counterfactual analyses, providing insights into which feature of the reform drives most of the welfare gains and on the optimal policy scheme. By removing each feature of the reform one-by-one, we evaluate which policy feature yields the largest increase in mothers' WTP. That is, we compare, in turn, the WTP for simulated counterfactual reforms in which *(i)* childcare prices remain unchanged, *(ii)* there is no increase in daycare coverage, and *(iii)* the refundable childcare credit is maintained. We find that not reducing the childcare price reduces the WTP for the reform by only 16%. However, reducing the daycare price without increasing daycare coverage entails very modest utility gains. Similarly, maintaining the refundable credit has little impact, given the small net price reduction it represents under the 5$/day regime. Therefore, this suggests that most of the welfare gains are due to increased coverage.

Last, we ask whether the government could have achieved a higher "bang for the buck" under different policy schemes. Specifically, we compare the Québec childcare reform to alternative price reduction and daycare expansion levels. Consistent with our previous results, given that parents are willing to pay substantially more for an increase in coverage than a reduction in price, we find substantially higher MVPFs for reforms that further increase coverage and reduce prices to a smaller extent. Our results thus suggest that, in the Québec context, higher welfare gains could be achieved by channelling more resources

[7]Hendren (2016) recognized that this result applies to small changes in *fiscal* policy but that pecuniary benefits are no longer sufficient if the policy in question changes the state of public good provision.

[8]Our own survey of a sample of MVPF estimates recorded in Hendren et al.'s (2023) Policy Impacts Library reveals that computing the MVPF of large policy changes as if they were infinitesimal is indeed common practice. We stress, however, that this exercise requires several judgement calls and should only be seen as being at best suggestive that policies considered in this literature are often non-marginal. See Appendix A.4 for the detailed results of our survey.

towards opening spots rather than lowering childcare fees.

This paper contributes to three strands of literature. First, we provide new evidence on the impact of universal childcare provision. In this rapidly growing literature, most studies have focused on measuring the causal impact of preschool enrolment on child development or maternal labor supply, yet yielding mixed evidence (see footnote 2). However, despite considerable efforts invested in this line of research, still little is known about the overall societal implications of such policy changes. Notable exceptions, such as Guner et al. (2020) and Daruich (2022), estimate general-equilibrium models of the family to study impacts of childcare programs. Compared to these studies, our work evaluates the welfare effect of an implemented reform rather than a hypothetical policy scheme. Additionally, we consider a broader range of potential non-pecuniary benefits for mothers, including the fact that parents might enjoy time spent with their children.[9] In the context of the Québec policy, we show that public provision of low-fee preschool can generate positive returns to society, even with a more diverse pool of beneficiaries compared to targeted programs. Moreover, we show that negative short-run impacts on non-cognitive outcomes (Baker et al., 2008, 2019; Haeck et al., 2015) do not necessarily translate into depressed economic outcomes later in life. This paper also highlights the role of local daycare supply in shaping impacts of universal programs, consistent with Yamaguchi et al. (2018a,b) for Japan and Cornelissen et al. (2018) for Germany. Lastly, we estimate mothers' WTP for the policy, including non-pecuniary gains, a dimension not addressed in previous cost-benefit analyses of implemented universal childcare programs (e.g. Fortin et al. (2013); Haeck et al. (2018) for the same policy or Andresen and Havnes (2019) for Norway).

Second, we link sufficient statistics and structural approaches to provide an empirical evaluation of the extent of bias in sufficient-statistic estimators. A few recent papers theoretically identified conditions under which transparent sufficient statistics for non-infinitesimal policy changes can be derived (Hendren, 2016; Finkelstein and Hendren, 2020; Kleven, 2021). However, as Kleven (2021) argues, those are in most cases beyond empirical reach. The key challenge is that, for large reforms, one cannot simply express the welfare effect of a policy as a fiscal externality. Kang and Vasserman (2022) propose welfare bounds for non-marginal reforms, but these would be difficult to apply to a policy changing many parameters in the economic environment such as the Québec program. We take an alternative approach and show how using a tractable structural approach can inform practitioners on the extent of bias in sufficient-statistics methods applied to non-marginal reforms empirically. In cases of policies involving substantial costs and significant non-pecuniary benefits, such as the Québec childcare reform, such an estimator

[9] In another related paper, Bravo et al. (2022) estimate the welfare effect of reduced distance to childcare centres induced by a national expansion in Chile, focusing on a marginal policy change. In contrast, we are interested in the overall effect of a reform at scale.

can substantially underestimate welfare gains.

Third, we contribute to a growing literature that combines reduced-form ex post estimation of policy impacts and structural estimation.[10] While some recent studies, such as Griffen (2019) and Chaparro et al. (2020), specify structural models of the family to interpret experimental impacts of targeted childcare programs (Head Start and IHDP, respectively), we do so in the context of universal childcare provision. This paper also relates to Chan and Liu (2018) who study a different policy scheme, which provided cash transfers to stay-at-home mothers in Norway.[11] We contribute to this literature by showing that a tractable behavioral model, which incorporates non-pecuniary considerations of childcare decisions, can replicate the reduced-form impacts of the Québec program. Furthermore, we leverage the natural experiment to build a transparent identification argument for some key model parameters.

The rest of the paper is structured as follows. Section 2.2 describes the institutional background and the data we use in our empirical analysis. In Section 2.3, we present our reduced-form analysis, namely our results on heterogeneity by local daycare availability, on long-run effects on children, and on the policy's impact on mothers' earnings. Section 2.4 presents our structural model of labor supply and childcare decisions and its estimation. Section 2.5 presents our estimates of the policy's MVPF as well as counterfactual simulations. Last, Section 2.6 concludes.

2.2 Background and Data

2.2.1 The Québec Childcare Reform

On September 1, 1997, a large-scale reform of preschool daycare was initiated by the provincial government of Québec, the second most populous province in Canada. At the time, the province was lagging behind the other Canadian provinces in terms of female labor-force participation. The major reform was thus designed to address this issue as well as to fight poverty and promote equality of opportunity for children (Japel et al., 2005). The centrepiece of the policy was the introduction of reduced-fare spaces in regulated childcare facilities at an out-of-pocket price of $5 per day per child (which increased to $7 in 2004).[12] Those low-fee spaces were allocated through the creation of a network

[10] Scholars such as Heckman (2010), Todd and Wolpin (2023), and Buera et al. (2023) discuss the merits of this approach, which combines "the best of both worlds" in empirical research. On the one hand, structural models can help interpret the mechanisms through which a given policy change impacts relevant outcomes and allow for counterfactual policy experiments that deviate from the implemented policy. On the other hand, ex post policy evaluation can be fruitfully used to identify and discipline behavioral models.

[11] Another relevant literature uses experimental results to estimate production functions of child cognition (e.g. Attanasio et al., 2020a).

[12] The average daily gross price in March 1997 was approximately $21 (Office des services de garde à l'enfance, 997a). A reduced-fee space thus represents annual gross savings of more than $4,100 for a

of new regulated facilities named *Centres de la petite enfance* (CPEs). The reform was phased in by age of the child over a period of 4 years. Initially, only children aged 4 (as of September 30$^{\text{th}}$) were eligible. In the following years, the age requirement was gradually lowered: one year later, 3-year-olds became eligible, followed by 2-year-olds in September 1999. In September 2000, subsidized spaces became available to all children aged less than 59 months. Access to the program was universal so that there were no entry requirements such as labor-force participation. In other words, the only condition for eligibility was the age of the child.[13] Importantly, (gross) prices remained constant for parents until 2014 – with the exception of an increase to $7 per day per child from 2004 – with the provincial government subsidizing the remaining fees.

Eligibility for subsidized spaces, however, did not imply that parents would actually find a spot for their child. Indeed, there were important shortages of spaces, especially in the first years of implementation. There were long waiting lists at each regulated childcare facility. Figure 2.1, which shows the evolution of the daycare coverage rate in the province by administrative region, illustrates this low supply. In 2000, only 35% of children aged 1-4 had access to low-fee childcare services. The slow growth in the number of spaces at the beginning is in large part due to the government's decision to freeze the number of spaces in unregulated daycare.[14] To remain as for-profit entities, daycare providers could only sign an agreement with the government and open additional spaces at a reduced fee. The moratorium on the creation of for-profit daycares was lifted in June 2002, after which the for-profit market expanded.[15] New spaces kept being created at a fast pace over the following years, raising the share of children with access to subsidized spots to 65% in 2008 (Lefebvre et al., 2009). Despite successive governments' efforts to increase supply, shortages remain a reality nowadays.

The reform also included the abolition, for households with a subsidized space, of some universal family allowances as well as of a refundable childcare credit prior to the adoption of the policy.[16] The credit rate, shown in Appendix Figure A.1, was decreasing

family signing a contract for the maximum number of days.

[13] To be eligible to the universal subsidy, families were required to enrol their child full-time for a maximum of 260 days per year. Families would typically sign yearly contracts to keep their space. The fees were billed monthly and had to be paid even if the child was absent from daycare. Note, nevertheless, that the for-profit market was not much more flexible, providers preferring to offer only full-time spaces because they were easier to manage (Haeck et al., 2018).

[14] Note that, due to long-run fertility trends, the daycare coverage rate (defined as the number of spaces per preschooler) did increase in the late 1990s even if the number of spaces stagnated somewhat.

[15] Until 2009, growth in the regulated network was still superior to that in the for-profit network. Data assembled by Haeck et al. (2016) which we complemented with recent years using ministerial reports for 2017 to 2019, however, shows that, from 2010, the for-profit network rapidly expanded as the regulated network stabilized. In the regulated network, the average annual number of newly created daycare spaces in the province from 2002 to 2009 was as high as 8,600 but dropped to 2,909 over the following decade. In contrast, between 2002 and 2009, 854 spaces were created annually in the for-profit network on average, but this figure dramatically increased to an annual growth of 6,322 spaces between 2010 and 2019.

[16] See Baker et al. (2005) for a detailed description of the changes to family allowances and other subsidies.

with household income. As a consequence, the reform was most beneficial for middle- and high-income households and changed financial incentives mostly for those families, raising concerns about equity (Baril et al., 2000).

Quality of care in childcare facilities

In addition to equity concerns, quality of care after the reform is one of its most controversial aspects. Under strong public pressure to open more spots at a reduced fee, the Québec government maintained minimal educational standards for daycare workers to facilitate entry into the profession at the implementation of the reform. A report on childcare quality in regulated settings from the *Institut de la Statistique du Québec* (Québec Statistical Institute) in 2004 emphasized the need for improving quality of care in those institutions (see Drouin et al., 2004). Moreover, an audit study conducted by Japel et al. (2005) between 2000 and 2003 revealed that the majority of childcare settings (61%) only met the basic criteria (ensuring the children's health and safety), and that their educational component was minimal. Almost one-eighth of them failed to meet the minimum standards.

However, Japel et al. (2005) also found that CPEs, on average, outperformed all other childcare settings on the vast majority of the criteria they considered for quality. For example, 26.5% of unregulated daycares (home-based or for-profit) were rated as inadequate in terms of quality, but only 6% of CPEs were rated as such. In the same vein, only 12.5% of unregulated daycares provided more than the "minimal" quality, while 33% of CPEs were deemed to provide a good service. Therefore, this evidence suggests that quality issues were actually more important in the private childcare market.[17] One part of the solution to improve average quality, some observers argued, was thus to increase the number of reduced-fee regulated spaces in CPEs.

In response to these quality issues, the provincial government, in addition to increasing the quantity of subsidized spaces, also gradually implemented some quality changes. In 2000, as documented by Molnár (2023), the educational requirements and wages of staff in regulated facilities were substantially increased over a four-year period. The average wage of child care workers was raised by 38 to 40 percent over this time span. The staff-to-child ratios remained unchanged (except for four and five year olds whose ratio increased by 25%) despite the increase in maximum capacity (Baker et al., 2005). Qualification requirements for the staff in centre-based CPEs were raised, and they were then extended to all centre-based care in 2006. In addition to political will, the increase in parents' involvement (in the board of directors, for example) is also an important factor that led to these quality changes.

[17] One potential reason for those quality differences is the greater generosity of infrastructure subsidies given by the government to daycare providers of subsidized spaces than to those operating privately.

2.2.2 Existing evidence on the Québec childcare reform

A substantial amount of work has been dedicated to the study of the impacts of the Québec reform on economic outcomes. First, two patterns are clear, namely, the overall positive and large effects on childcare utilization and on maternal labor supply. Baker et al. (2008) find that the reform induced an average increase in child care usage of 14.6 percentage points and in labor-force participation of mothers (of two-parent families) of 7.7 percentage points (14.5% of baseline participation) in Québec compared to the rest of the country. Subsequent studies (Lefebvre and Merrigan, 2008; Haeck et al., 2015; Kottelenberg and Lehrer, 2017) report similar estimates confirming this general pattern. On care use, Haeck et al. (2015) additionally found positive effects at the intensive margin. Indeed, the policy also increased the number of hours families sent their children to daycare conditional on already using non-maternal care. However, there is substantially less consensus about which mothers increase their labor supply the most. Lefebvre and Merrigan (2008) and Haeck et al. (2015) find similar responses for the low-educated and the high-educated in absolute terms. Lefebvre et al. (2009) found that the effects on maternal labor supply were long-lasting and that they were mainly driven by low-educated mothers. Molnár (2023), on the contrary, finds that eligible high-educated mothers are substantially more likely to increase labor-market participation thanks to the reform, both in absolute and relative terms on both margins. Thus, the estimated relative response on this education margin varies across studies, perhaps because the authors used different datasets and considered different specifications.[18]

Second, a set of papers documented the impacts of the reform on child development. Baker et al. (2008) found striking evidence that, on average, eligible children in two-parent families experienced worse development outcomes and were exposed to worse parenting practices, such as worse health outcomes and lower consistency in parenting. Moreover, the overall negative impacts of the universal childcare program on child outcomes and parenting were confirmed in many other studies (e.g. Kottelenberg and Lehrer, 2013; Haeck et al., 2015, 2018; Baker et al., 2019; Haeck et al., 2022). In particular, Baker et al. (2019) find that the initial negative impacts on children had long-lasting consequences for exposed cohorts later in life such as higher crime rates and lower life satisfaction. However, in a series of studies, Kottelenberg and Lehrer (2013, 2014, 2017, 2018) provided ample evidence of substantial heterogeneity in those impacts by the age of the child, the initial development score distribution, and the child's gender.[19] In the same vein, Molnár (2023) finds that while eligible parents are less likely to read daily to their child, there is an increase in reading time at the lower end of the reading distribution: the propensity

[18] In a recent study, Karademir et al. (2023) find that the reform also had a small positive impact on employment of grandmothers.

[19] Moreover, the results of Baker et al. (2019) were challenged by Haeck et al. (2018), who found that after accounting for variation in treatment dosage, the long-term negative effects were substantially less severe than what Baker et al. (2019) estimated.

to never read decreased, and the likelihood of reading once or several times per week increased.

In a recent contribution, Molnár (2023) also shows that the increase in labor supply driven by increased childcare use was accompanied by compensating behavior of parents in their children's education. For example, she finds that eligible parents also increase the focused time spent with their children and rather reduce home production time and leisure time.[20] Overall, while the positive and large effects on childcare take-up and maternal labor supply are clear, the negative effects on child development documented by Baker et al. (2008) appear to be only true on average, but the subsequent literature uncovered substantial heterogeneity.

Despite all the efforts invested in estimating the short- and medium-run impacts of the Québec reform on a variety of economic outcomes, we still know very little about the overall implications of the policy change. In the few attempts at measuring the cost-benefit ratio (see Fortin et al., 2013; Haeck et al., 2015), authors typically only focus on the direct cost of subsidies and the labor-supply response of parents. In particular, Haeck et al. (2015) find that the fiscal gains for the government stemming from the increased labor supply of mothers do not cover the upfront cost of the daycare subsidies. In such situations in which the policy does not pay for itself, accurately measuring the WTP for the reform can substantially affect the welfare conclusions.

We thus contribute to this literature by bringing together *(i)* previous reduced-form results with additional regional heterogeneity and *(ii)* non-pecuniary gains for mothers to measure welfare. Local capacity constraints might not only help explain the mixed evidence on policy impacts by education group but are also important for welfare assessments given that having spots available matters for families. Moreover, in the case of childcare subsidies, which are likely to be valued for non-financial reasons, the benefits of the policy are most likely underestimated by previous papers. Since the policy is large and the transfer is in-kind, we account for such benefits by estimating an economic model of the family.

2.2.3 Data sources

For our empirical analysis, which includes both the reduced-form ex post evaluation of the Québec reforms and the estimation of the structural model, we utilize several sources of Canadian micro-data. The main source is the National Longitudinal Survey of Canadian Youth (NLSCY), a common dataset in the literature. These data contain

[20]She finds that this average reallocation is driven by high-educated mothers. Low-educated mothers rather increase leisure time and, if anything, decrease the time spent with the child. This latter effect is, however, statistically insignificant. Moreover, exposed families increased the share of household spending dedicated to food, games and toys, and domestic help, again suggesting that parents tried to compensate for the increase in work hours.

rich information on a representative sample of Canadian children and their parents over the period of the reform. We notably observe measures of care quality, labor-market participation of parents, and daycare expenditures, all of which are crucial to estimate parents' WTP for the policy. These repeated cross-sectional surveys covering the period 1994-1995 to 2008-2009 also contain a longitudinal component allowing to follow a subset of children over several survey cycles. Table A.2 reports summary statistics comparing Québec and the rest of Canada, our control group. To estimate the long-run effects of the childcare reform on eligible children, we use the Canadian Censuses of population of 2016 and 2021. These recent datasets allow us to compare individuals who are old enough to have completed their education. We relegate further details of these more standard datasets to Appendix A.1.1 and rather dedicate more space to describe our novel data source on daycare supply within Québec.

Daycare supply in Québec

While previous studies of the Québec childcare reform estimate ITT effects of the policy at the provincial level, we investigate treatment effect heterogeneity at a more granular level *within* Québec. To this end, we assemble a novel dataset of the daycare coverage rate at the administrative-region level in Québec from a series of annual management reports – as well as some reports on childcare demand for pre-reform years – of the Ministry of the Family (*Ministère de la Famille*, which also held various other names). We use this information to allow treatment dosage to vary by region of residence depending on the extent to which coverage increased over the period of analysis. Specifically, the daycare coverage rate is defined as the ratio between the number of childcare slots and the total number of preschool-age children (0-4 year olds) in the region. These reports include information on the number of daycare spaces by administrative region by type of facility (centre-based, CPE, for-profit, etc.) from 1994 to 2019. Unfortunately, prior to 2004, we cannot distinguish between regulated and for-profit spaces. This is not a major issue for our empirical analysis for two reasons. First, the share of for-profit spaces in Québec was less than 7% of total spots until 2010, thus making the for-profit market rather marginal. Second, we restrict our empirical analysis to until 2002-2003, the period over which the government froze the number of for-profit spaces.

There are 17 administrative regions in Québec, which makes them a relatively granular level given that the total provincial population was approximately 7 million inhabitants in the late 1990s. Moreover, in the Québec context, using the coverage rate at the municipal level might not be an ideal strategy since many families send their children to daycare in other cities.[21] Therefore, it might be problematic to assume that children

[21] From other ministerial reports, we can confirm that this phenomenon is rare at the administrative-region level but is actually common at the municipal level. The share of children in daycare coming from other administrative regions is low (7.5% for Montréal and Laval and only 1.8% on average in other

attend facilities in their city of residence, but it appears as a reasonable assumption at the administrative-region level.

Appendix Figure A.2 provides another illustration of the differential expansion across regions in Québec. Before the reform, childcare coverage was very low across the entire province, but substantial heterogeneity across regions already existed then. Coverage increased considerably from the late 1990s at different rates, with some low-coverage regions eventually catching up with high-coverage regions, until childcare availability exceeded 0.35 space per preschooler in all regions in 2011.

Determinants of local childcare expansions

The substantial spatial variation in childcare expansion across regions within the province of Québec is clearly appealing from an identification standpoint. It notably allows us to assess the extent of heterogeneity in our intention-to-treat estimates. However, to interpret this evidence as causal, local childcare expansions need be plausibly exogenous to the evolution of parents' labor-market outcomes and childcare arrangements. In other words, expansions need to be a daycare supply change and not be driven by an increase in demand. In what follows, we present some evidence to support the reliability of this assumption.

First, to assess the extent to which the local daycare supply increase can be regarded as quasi-random, we explore the potential determinants of changes in regional coverage rates (see, e.g. Cornelissen et al., 2018; Yamaguchi et al., 2018a,b). We obtain region-level information from public datasets of the Québec Statistical Institute, which we complement with other indicators from the Canadian Census of 1996. In Figure 2.2, we plot the correlation between the daycare expansion level and each characteristic in turn. We define the expansion level as the change in the daycare coverage rate between 1997 and 2003 – namely before the raise of the daily fee to 7$ and when the CPE network stabilized. The Figure reveals that the share of educated inhabitants and the initial coverage rate in the region are positively and negatively correlated with the expansion level, respectively, but that all the other characteristics are uncorrelated with expansion levels. While the negative correlation with initial coverage is mostly mechanical, we could be worried that local daycare expansions might capture differences in mean educational attainment across regions. Therefore, we control for these education shares in our regressions in the following Section along with region fixed effects, which capture variation in time-invariant regional characteristics.[22] In Appendix Table A.4, we regress the expansion level on all

regions in 2001), but it is large at the municipal level. Indeed, in 2001, 25.1% and 15.8% of children in daycare in the Montréal and Laval regions were coming from another CLSC territory (a geographical unit grouping a few municipalities), respectively. A similar figure is observed in other regions: 23.6% of children in CPEs and 19% of those in other daycare came from other municipalities in 2001 (Ministère de la Famille et de l'Enfance, 2001b).

[22]Another plausible demand-side channel could be that mothers take-up employment in public

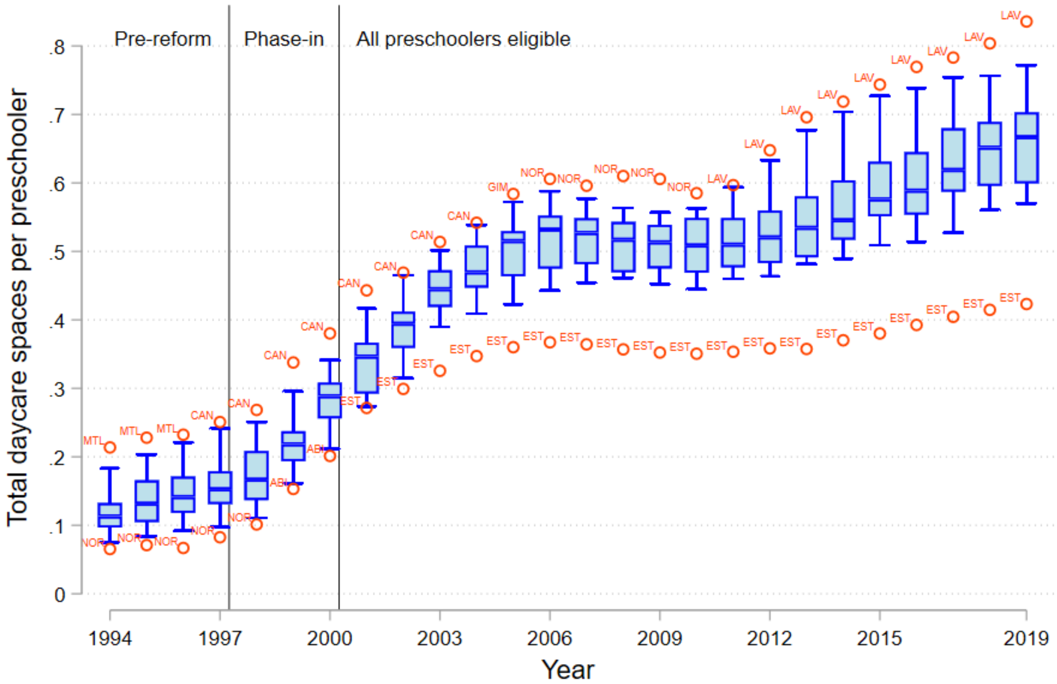


Figure 2.1: Evolution of the daycare coverage rate by administrative region, Québec

Note: This Figure displays the median (middle blue line), the 25th and 75th percentiles (light blue boxes), and the 10th and 90th percentiles (blue brackets), and outliers (orange circles) of the daycare coverage rate distribution at the administrative-region level in a given year. It is constructed using data from a series of ministerial reports for daycare spaces (*Ministère de la Famille*, which held various other names), from the Québec Statistical Institute for the children population, and from authors' calculations. The dependent variable is the ratio between the total number of spaces (the sum of subsidized and non-subsidized spaces) for preschool children (as of March 31st) over the number of children aged 0-4 years (as of July 1st) in a given region. Before the reform, low-income families had access to childcare subsidies and subsidies became universal in 1997. The reduced-fee program ($5/day/child) began in September 1997 only for children aged 4. All preschool children (0-4 years old) became eligible only in September 2000. In January 2004, the daily fee was raised to $7. ABI = Abitibi-Témiscamingue; CAN = Capitale-Nationale; EST = Estrie; GIM = Gaspésie-Îles-de-la-Madeleine; LAV = Laval; MTL = Montréal; NOR = Nord-du-Québec;

the characteristics and reassuringly, we find that none of these variables can predict local childcare expansions. Indeed, we cannot reject the null hypothesis that the estimated coefficients are jointly zero ($p = 1.0$).

A second potential threat to identification could arise from households endogenously sorting into different regions according to childcare availability. To the extent that such residential choice is correlated with unobservable characteristics affecting our outcomes of interest, we would be erroneously attributing the observed changes in outcomes to the increase in daycare availability. For instance, it could be that mothers from low-coverage regions chose to move to other areas where offers in daycare were more abundant precisely because they wish to keep working following childbirth. Such a situation would generate non-random selection into treatment and bias our estimates. To get a sense of whether

childcare services. However, data from the 1996 and 2001 Canadian Census suggests that the share of mothers' employment in our sample of interest (mothers of preschoolers in two-parent families) is very small and in fact does not increase from 1996 to 2001. Indeed, this share decreases from 3.69% in 1996 to 1.5% in 2001.

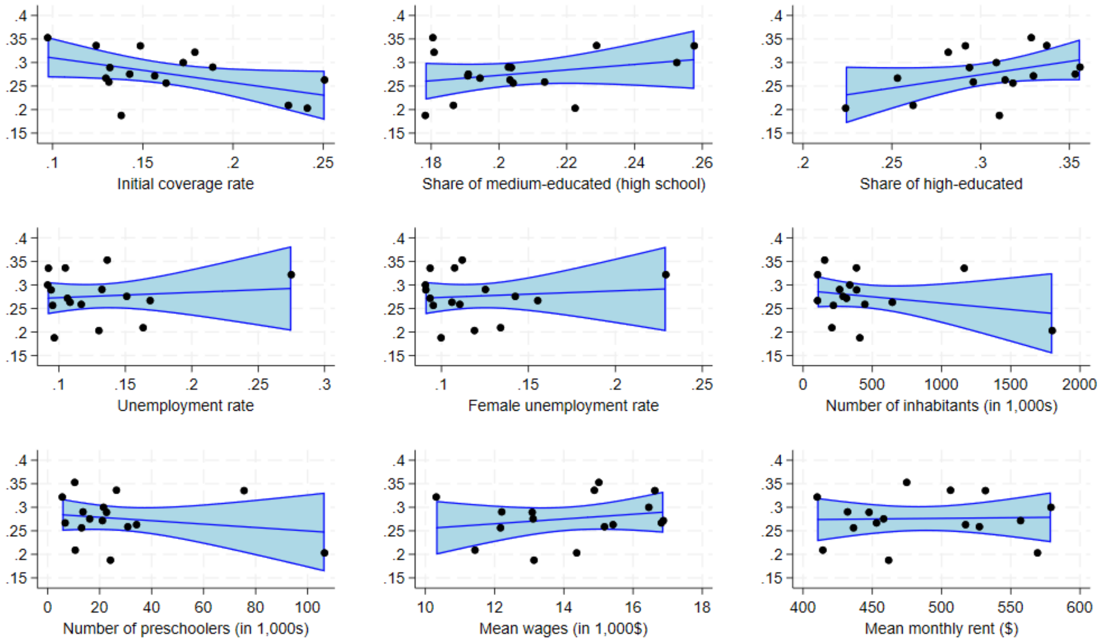


Figure 2.2: Regional daycare expansion and region-level characteristics

Note: This Figure illustrates the relationship between the daycare expansion level and region-level characteristics potentially associated with childcare demand. The daycare expansion level is defined as the increase in the daycare coverage rate from 1997 to 2003. Region-level variables are calculated from 1997 data of the Québec Statistical Institute and the 1996 Canadian Census of population. Shaded areas are 95% confidence intervals.

this phenomenon is relevant in our setting, we study the trends in interregional migration of families with a preschool-age child in Appendix Figure A.3. We find that, while there is an increase in migration flows of approximately 10% in regions which experienced the largest childcare expansion in the following years, we observe a similar trend in those where childcare provision did not expand as much. Therefore, it is not the case that families systematically moved to areas that experienced greater increases in daycare supply. This additional evidence suggests that this type of self-selection into treatment is not a major concern in our context.

Last, in Appendix Table A.3, we show that pre-reform characteristics in low- versus high-expansion regions are very similar. Not only are families in the two groups of regions similar in terms of demographic characteristics (such as parents' education and the number of children in the household), but they are also comparable along our outcomes of interest. With this evidence in hand, we now evaluate how the impacts of the price decrease interacts with changes in local supply.

2.3 Impacts of the reform

In this section, we begin our empirical analysis by estimating the impact of the Québec childcare reform on parents' behavior and exposed children's educational outcomes as they age. Using our new data on regional daycare coverage rates within Québec, we estimate heterogeneous effects of the policy on maternal employment, childcare use, and parenting practices. This constitutes our first contribution. We notably find that this heterogeneity can help explain some puzzling evidence of positive impact of the reform on low-educated mothers (Lefebvre and Merrigan, 2008; Molnár, 2023) despite their financial incentives being small.[23] Second, we investigate whether the negative impacts on child health and non-cognitive outcomes found by Baker et al. (2008, 2019) extend to economic outcomes later in life. Our empirical strategy exploits the most recent Canadian Censuses, which allow us to observe exposed children with completed education.

2.3.1 Impact on parents' time allocation

In this section, we estimate the impact of the Québec reform on parents' labor-market outcomes that constitute the first source of fiscal externalities. We focus on short-term impacts for three reasons. First, a major change in Québec's parental-leave policy occurred in 2006 and had a substantial impact on mothers' employment and earnings (see Patnaik, 2019; Karademir et al., 2023). We are therefore reticent to use all cycles of the NLSCY, especially for the evaluation of the earnings impacts, as this other policy change might introduce bias in estimates of long-term effects. Second, some features of the reform evolved a few years after its implementation. As mentioned in Section 2.2.1, some quality changes were implemented starting in the 2000s. Additionally, daycare providers could open spaces in the unsubsidized network from June 2002 onward. To avoid capturing these confounding effects, we focus our analysis on the reform as it was originally conceived. Third, for comparability reasons of the heterogeneous impacts by the local level of childcare availability, we follow the original empirical approach established in the literature on this reform. Moreover, by focusing on parents of children eligible over all of the preschool period, we mitigate concerns over treatment effect weighting in staggered designs and anticipatory behavior (De Chaisemartin and d'Haultfoeuille,

[23]Daycare coverage rates are a popular instrument in the literature on childcare enrolment. Havnes and Mogstad (2011b,a, 2015) use differences in childcare expansions to study the impacts of a universal childcare reform in Norway. Cornelissen et al. (2018) and Felfe and Lalive (2018) use municipal variation in Germany, Andresen (2019) do so in Norway, and Yamaguchi et al. (2018a,b) do so in Japan to estimate marginal treatment effects (MTE) on child development and maternal labor supply. To our knowledge, we are the firsts to exploit such variation in a North American setting. However, given the rather small number of regions, our data does not provide us with an instrument of sufficient support to estimate MTE. We thus limit the analysis to comparing low- to high-expansion regions.

2020).[24]

Empirical strategy. We start by analyzing the time-allocation response of parents using waves 1-2-4-5 of the NLSCY following Baker et al. (2008) (henceforth BGM). We first replicate the well-established results of previous studies, notably the large impacts on maternal labor supply. The baseline specification is a standard difference-in-differences estimator, where we estimate, for mother i in province p in year t:

$$Y_{ipt} = \alpha + \beta \text{Elig}_{pt} + \gamma_p + \gamma_t + \delta X_{ipt} + \varepsilon_{ipt} \tag{2.1}$$

where Y_{ipt} is either a parent's labor supply (extensive and intensive margin), childcare use (intensive and extensive margin), or the frequency of reading to the child, our measure of parenting effort. The eligibility dummy Elig_{pt} takes value 1 if the household resides in Québec after the reform. γ_p and γ_t are province and survey year fixed effects. X_{ipt} is a vector of controls including age, age of the child, number of siblings, population of the area of residence, education (both parents), immigration status (both parents), and provincial unemployment rates.

Using our novel data on daycare coverage rates, we then investigate heterogeneity in policy impacts at the administrative-region level. Our empirical strategy employs an intent-to-treat (ITT) difference-in-differences estimator comparing two-parent families with a preschool age child in Québec to similar families in the rest of Canada. We use the same baseline set of control variables and the same sample restrictions (two-parent families) as BGM to ensure that differences in our estimates are solely due to considering local daycare supply and not to differences in design. However, to account for potential changes in composition across regions (within Québec), we also include control variables at the regional level when considering heterogeneous impacts. Our main empirical specification becomes:

$$Y_{iprt} = \alpha + \beta_1 \text{Elig}_{pt} + \beta_2 \text{Elig}_{pt} \times \text{LowExp}_r + \gamma_p + \gamma_r + \gamma_t + \sigma W_{rt} + \delta X_{ipt} + \varepsilon_{iprt} \tag{2.2}$$

where r indicates the administrative region of residence (within Québec only). LowExp_r equals one if region r is in the bottom tercile of the distribution of daycare expansion over the period. This variable aims at capturing regions in which daycare expanded little. The expansion level is defined as the difference between region r's daycare coverage rate in 2003 to its 1997 level. γ_r is a vector of region (within Québec only) fixed effects. Lastly, W_{rt} is a vector of regional-level control variables associated with childcare demand (shares

[24] Ding et al. (2020) find suggestive evidence of strategic placement of children over the implementation period, especially by families with high maternal education. High-educated mothers were significantly more likely to pay for unsubsidised spaces to guarantee a subsidized spot once such new spaces would be open. This strategic response generated a disproportionate increase in childcare use in the province among younger children not yet eligible. Karademir et al. (2023) document similar anticipatory behaviors.

of medium- and high-educated mothers and the number of preschoolers in the region).

Results. The main coefficients of interest are β_1 and β_2, which capture the differential effects of the policy by local daycare availability. In Tables 2.1 and 2.2, we report point estimates of the two specifications above along with results on heterogeneity by availability without regional controls. In columns (2) and (3) of Table 2.1, we find that the labor-supply response of mothers at the extensive margin is much stronger in regions with higher coverage. In regions where daycare expanded more, the policy boosted maternal labor-force participation 40% more on average, well above the average effect of 7.7 percentage points in the entire province. In regions in the bottom tercile of the daycare coverage rate distribution, the increase in maternal employment is substantially lower and this estimate is statistically significant. High-coverage areas thus appear to be the regions that were driving most of BGM's original result (reported in column 1).

Table 2.1: Heterogeneous impacts of the Québec childcare reform on mothers' employment

	Mother works			Mother's work hours		
	(1)	(2)	(3)	(4)	(5)	(6)
β_1 : Eligible$_{pt}$	0.078***	0.128***	0.156***	2.129***	3.667***	4.356***
	(0.007)	(0.018)	(0.014)	(0.298)	(1.000)	(0.654)
β_2 : Eligible$_{pt}$ $\times$LowExp$_r$		-0.053***	-0.063***		-1.770***	-1.751***
		(0.007)	(0.006)		(0.598)	(0.632)
LowExp$_r$		0.042***			1.189**	
		(0.006)			(0.588)	
Region (r) FE			✓			✓
r-level controls			✓			✓
Mean dep. var.	0.532			17.54		
p-value of $\beta_1 + \beta_2 = 0$		0.000	0.000		0.000	0.000
R^2	0.105	0.106	0.107	0.099	0.099	0.102
N	33,758	33,758	33,758	33,637	33,637	33,637

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), immigration status (both parents), and provincial unemployment rates. Odd columns report estimates of equation (2.1) while even columns are regression results of equation (2.2). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In columns (4) to (6) of Table 2.1 and in Table 2.2, we examine the impact of the policy on other components of households' time allocation, namely hours worked by the mother and childcare use. We estimate that, on average, (i) eligible mothers work two additional hours per week and (ii) families with a young child use childcare for more than 5.7 additional hours. Those estimated effects are statistically significant. In Table A.5, we further confirm previous results in that the increase in childcare use is driven by an increase in institutional care and that the labor supply of fathers is unchanged.

For these outcomes as well, average effects mask substantial heterogeneity by the

Table 2.2: Heterogeneous impacts of the Québec childcare reform on childcare use

	Child in care			Childcare hours		
	(1)	(2)	(3)	(4)	(5)	(6)
β_1 : Eligible$_{pt}$	0.138***	0.164***	0.187***	5.736***	6.614***	7.325***
	(0.032)	(0.029)	(0.034)	(1.405)	(0.917)	(0.942)
β_2 : Eligible$_{pt}$		-0.048***	-0.051***		-2.124	-2.276
$\times$LowExp$_r$		(0.014)	(0.016)		(1.443)	(1.556)
LowExp$_r$		0.018			1.182***	
		(0.013)			(0.151)	
Region (r) FE			✓			✓
r-level controls			✓			✓
Mean dep. var.	0.418			13.07		
p-value of $\beta_1 + \beta_2 = 0$		0.001	0.001		0.017	0.012
R^2	0.116	0.116	0.118	0.110	0.110	0.113
N	33,709	33,709	33,709	30,915	30,915	30,915

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), immigration status (both parents), and provincial unemployment rates. Odd columns report estimates of equation (2.1) while even columns are regression results of equation (2.2). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

local daycare supply change. Mean impacts on mothers' hours worked and childcare utilization are indeed less pronounced in regions where childcare supply expanded less. Only for childcare use at the intensive margin do we find an imprecisely estimated β_2, but the coefficient is nevertheless negative as for the other outcomes.

Taken together, our findings suggest that the relief of capacity constraints in daycare at the local level was an important driver of the policy's impacts on households' time allocation. In other words, not only the price decrease but also the increase in capacity at the local level was an incentive for mothers to take-up employment and use childcare.

Impact on parenting practices. The policy increased maternal labor supply and thus mechanically reduced the time mothers spend at home with their children. If maternal care time and parenting effort are substitutes, we might expect mothers to compensate for the reduced time at home by spending more quality time with the child. In Table 2.3, we estimate the heterogeneous impact of the reform on the weekly frequency of reading to the child, our measure of parenting effort, by coverage status.

Consistent with previous results by Molnár (2023), we find that the policy had a positive impact on reading time at the bottom of the reading distribution. Point estimates suggest that parents were 4.4 percentage points more likely to read at least once per week and 5.4 percentage points less likely to never read to the child. We detect no short-run impact at the top of the reading distribution (reading daily). As for the time-allocation outcomes, the average impacts are driven by the most treated regions. For instance, the estimated decrease in the propensity to never read is almost entirely concentrated in high-expansion regions. These results thus suggest that mothers compensated for their increased work hours by exerting more effort parenting when they are home.

Table 2.3: Heterogeneous impact of the Québec childcare reform on weekly frequency of reading to the child by daycare expansion status

	Rarely/never			At least weekly			Daily		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β_1 : Eligible$_{pt}$	-0.054***	-0.112***	-0.095***	0.044***	0.094***	0.046**	-0.015	-0.013	0.020
	(0.009)	(0.006)	(0.012)	(0.017)	(0.019)	(0.022)	(0.018)	(0.036)	(0.030)
β_2 : Eligible$_{pt}$		0.083***	0.079***		-0.079***	-0.074***		0.006	-0.003
$\times$LowExp$_r$		(0.010)	(0.010)		(0.019)	(0.007)		(0.023)	(0.010)
LowExp$_r$		-0.051***			0.037***			0.027	
		(0.010)			(0.010)			(0.021)	
Region (r) FE			✓			✓			✓
r-level controls			✓			✓			✓
Mean dep. var.	0.226			0.748			0.379		
p of $\beta_1 + \beta_2 = 0$		0.084	0.107		0.678	0.410		0.752	0.520
N	33,171	33,171	33,171	33,171	33,171	33,171	33,171	33,171	33,171
R^2	0.170	0.170	0.171	0.053	0.053	0.056	0.165	0.165	0.168

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), immigration status (both parents), and provincial unemployment rates. Even columns report estimates of equation (2.1) and odd columns are regression results without regional-level variables (shares of medium- and high-educated mothers and the number of preschoolers in the region r). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Robustness checks. We additionally perform robustness checks on our baseline results, which we report in the Appendix. First, in Table A.6, we compare our results to using only Ontario, the most similar province to Québec in terms of size and economy, as a control group. The estimates are quantitatively very similar to those of our main specification. Second, we verified that our results are robust to estimating the standard errors with the wild cluster bootstrap procedure of Cameron et al. (2008) accounting for the small number of clusters.[25]

Threats to identification

As with any difference-in-differences strategy, the main concern for identification of the policy's treatment effects is that of differential trends between the treatment and control groups prior to the reform. Many papers on the Québec childcare program have argued and provided robust evidence that Québec and the rest of Canada (RofC) were following similar trends on a wide variety of outcomes prior to treatment (e.g. Baker et al., 2008, 2019; Haeck et al., 2015, 2018; Molnár, 2023). However, we might be concerned that our two treatment groups within Québec (high- and low-expansion regions) were

[25] These results are available in the Research Data Centre of Statistics Canada, and upon request.

evolving differently prior to the policy. While there is no direct test of the parallel trends assumption, we provide graphical evidence in Figure 2.3 that our three groups were on similar trends prior to the reform for our outcome variables.

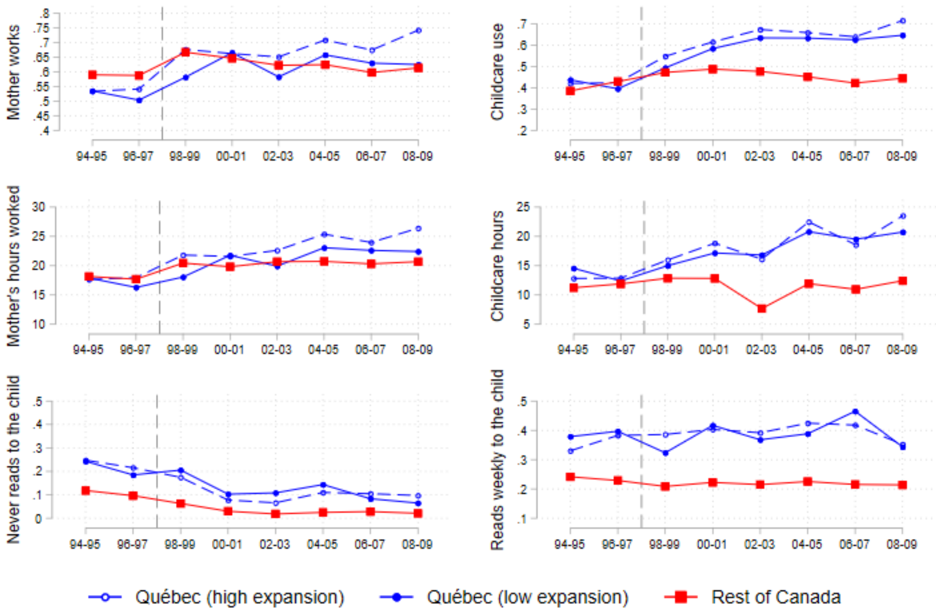


Figure 2.3: Mean childcare use, maternal labor supply, and reading time by daycare expansion status (in Québec) and in the rest of Canada

Note: These figures plot the means of selected outcome variables separately for three groups, namely two groups within Québec (in blue) and the rest of Canada (in red) in the NLSCY. The dotted (resp. solid) blue line represents families living in regions where the expansion of daycare coverage falls into (resp. is above) the bottom tercile in the province over our study period. The data source is the 8 waves of the NLSCY. The sample is restricted to two-parent families with a preschool-age child.

Appendix Figure A.4, which reports estimated coefficients of event-study regressions, further confirms the graphical analysis. Those regression results show that prior to the policy, the mean outcomes of interest were either converging or not statistically different before sharply diverging in post-policy waves. Lastly, Table A.3 reports pre-reform descriptive statistics on household characteristics and selected outcomes by expansion status. Our two treated groups are comparable prior to the policy change on all characteristics, thereby strengthening our confidence that low-expansion regions were not following differential trends.

Heterogeneity by maternal education

Last, we check whether the heterogeneity by local daycare capacity might help explain some intriguing results from previous studies. Lefebvre and Merrigan (2008) and Molnár (2023) found positive impacts of the policy on maternal employment for both the high- and low-educated mothers. These results are somewhat surprising because the financial incentives to take-up childcare were substantially stronger for better-off families. Indeed, to finance the policy, the Québec government abolished a refundable childcare credit that was rapidly decreasing with household income. For low-income families, the difference in the net price of childcare introduced by the reform was thus very small. For the poorest households, the median net price before the reform was actually approximately the same as a subsidized space under the new regime. However, even if the financial incentives were low, it is possible that low-income households responded to increased availability.

In Table 2.4, we estimate equation (2.1) separately for high- and low-educated mothers. Following Molnár (2023), we define high-educated mothers as those who have completed a post-secondary degree. Consistent with the literature, we find that the average employment impact is driven by high-educated mothers (column 3). Using waves 1-2-4-5 of the NLSCY, we find a small and insignificant effect on low-educated mothers (column 1). However, introducing heterogeneity by local daycare supply (columns 2 and 4) reveals that in higher-coverage regions, low-educated mothers do significantly increase their labor supply. This estimated impact is twice as large as that of high-educated mothers in the same regions. Moreover, we find no statistically significant difference in the impact of the policy by coverage status among high-educated mothers. These results are consistent with the financial incentives mentioned above: for high-educated mothers, our results suggest the main incentive to take-up employment was the price reduction; for low-educated mothers, access to a space was key. This also shows up in childcare take-up (columns 5 to 8), where the stronger response in high-expansion regions is again driven by low-educated mothers.

Table 2.4: Heterogeneous impact of the Québec childcare reform on maternal employment by daycare expansion and mother's education

	Mother works				Childcare use			
	Low educ		High educ		Low educ		High educ	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : Eligible$_{pt}$	0.030	0.191***	0.095***	0.084*	0.075**	0.222***	0.158***	0.157***
	(0.019)	(0.029)	(0.079)	(0.045)	(0.014)	(0.078)	(0.027)	(0.016)
β_2 : Eligible$_{pt}$ x LowCov$_r$		-0.099**		0.033		-0.082*		-0.01
		(0.047)		(0.044)		(0.042)		(0.037)
Region (r) FE		✓		✓		✓		✓
r-level controls		✓		✓		✓		✓
p-value of $\beta_1 + \beta_2 = 0$		0.031		0.000		0.046		0.001
N	10070	10070	23688	23688	10048	10048	23661	23661
R^2	0.103	0.103	0.084	0.084	0.093	0.094	0.103	0.103

Note: The data source is waves 1-2-4-5 of the NLSCY. Even columns report estimates of equation (2.1) and odd columns are regression results without regional-level variables (shares of medium- and high-educated mothers and the number of preschoolers in the region r). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2.3.2 Earnings gains

We now turn to analyzing the impact of the reform on earnings, which constitutes the main source of fiscal externalities. Because we find positive effects on maternal labor supply but no significant responses from fathers, we focus on mothers' labor earnings as in Lefebvre and Merrigan (2008). To be consistent with our model estimated in section 2.4, we estimate equation (2.1) where the outcome is mothers' annual labor earnings using the income information in the NLSCY. The point estimate, along with 95% confidence intervals, is reported in Figure 2.4. We find that, on average, mothers of preschoolers in Québec earn an additional \$3,750 (in constant 1997 dollars) per year in the post-reform period compared to other Canadian mothers of young children. Despite using a different dataset and focusing on a different age group than Lefebvre and Merrigan (2008), we reassuringly obtain a point estimate that is quantitatively comparable to their result.[26] To assess the plausibility of the parallel-trends assumption, we also estimate an event-study regression. Appendix Figure A.5 confirms that earnings of Québec mothers are not statistically different in the pre-reform waves of the NLSCY, but start diverging only in the post-policy periods.

Because earnings gains have different fiscal impacts along the income distribution, we move beyond average impacts and investigate how the policy shifts the income distribution. As we are mostly interested in impacts on income in an absolute (unconditional) sense, we use the unconditional quantile regression framework of Firpo et al. (2009). This approach estimates quantile treatment effects by comparing the cumulative earnings distributions in pre- and post-reform periods in treatment and control groups using a recentered influence function (RIF) regression. Point estimates then indicate by how much a given quantile of mothers' labor income in Québec has shifted due to the policy. The analysis reveals that there is a positive effect of about \$2,000 at the 4th and 9th deciles and a larger impact of \$4 to \$5 thousands in between. In our analysis of the fiscal externality in section 2.5, we take into account the impact of this heterogeneity on the government's budget.

2.3.3 Long-run impact on eligible children

Having established that the policy has significant impacts on mothers' labor-market behavior, we now end our reduced-form analysis by investigating long-run effects on eligible children as they age. As mentioned in Section 2.2.2, previous evidence on the

[26]Using the Survey of Labour Income Dynamics (SLID) covering the period 1993-2002, Lefebvre and Merrigan (2008) estimate a positive impact of 2,486\$ (once adjusting to 1997 dollars) on mothers of children aged 1-5. Apart from the different sources and age ranges considered, the difference in the estimates might stem from the inclusion by the authors of 1999, where earnings gains are small and statistically insignificant, as a post-reform period. We refrain from including implementation years in our analysis for the reasons mentioned at the beginning of section 2.3.1.

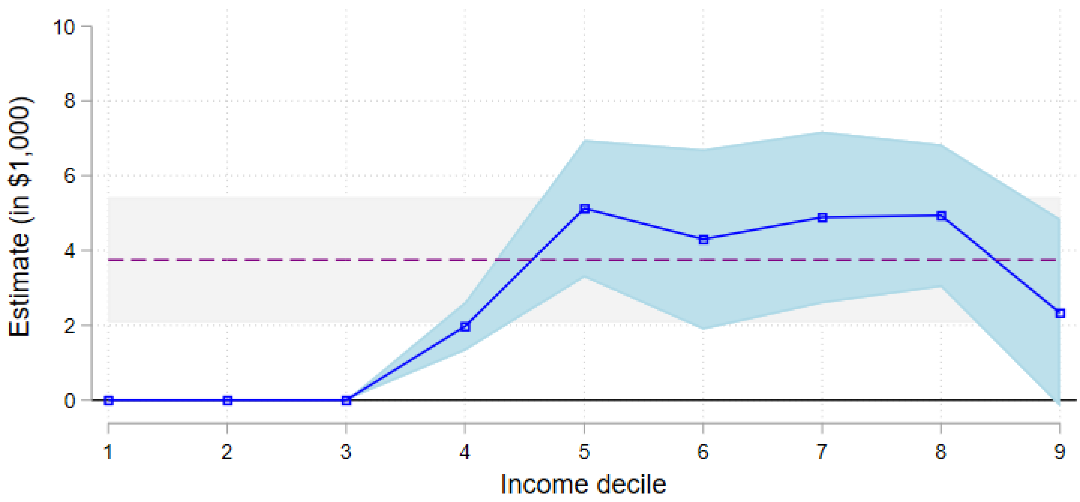


Figure 2.4: Distributional impact of the Québec childcare reform on mothers' labor earnings

Note: This Figure displays the mean impact (dashed purple line) of the Québec childcare reform along with point estimates from RIF unconditional quantile regressions (blue line) of the policy impact on mothers' annual labor earnings (in constant 1997 thousand dollars). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Shaded areas are 95 percent confidence intervals around each estimate.

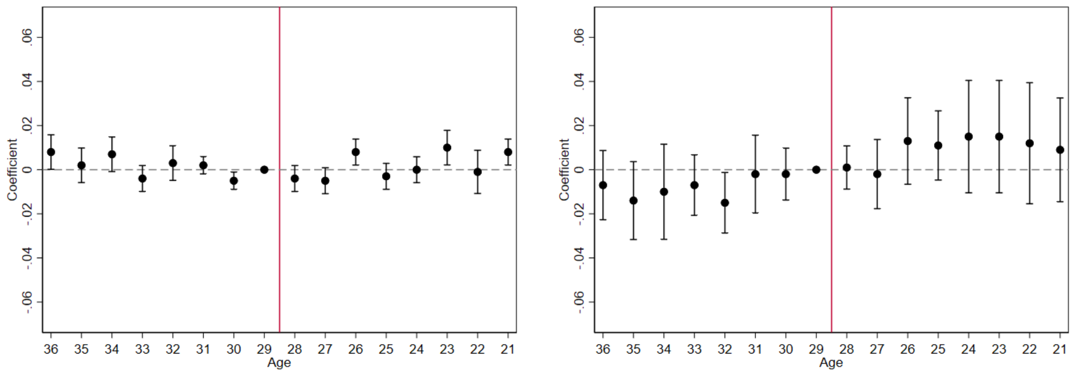
Québec childcare reform documented average negative impacts on children's non-cognitive development in the short run, but evidence is more mixed in the long run (Baker et al., 2008, 2019; Haeck et al., 2015, 2018). In this section, we assess whether those impacts have long-run implications on economic outcomes as eligible children age. Experimental evidence from targeted programs indeed suggests that boosting non-cognitive skills at a young age causes long-term improvements in economic success (Heckman et al., 2013; Algan et al., 2022). It is therefore possible that the short-run negative effects on behavior and health have translated into depressed economic outcomes later in life.

To investigate this possibility, we estimate the long-run impact of the policy on eligible children's educational attainment. Using the Canadian Censuses of 2016 and 2021, we implement a triple-difference estimator, which compares same-age individuals who vary in eligibility status based on the census year and their province of birth. For individual i of age a born in province p observed in census year t, we estimate the following model:

$$Y_{iapt} = \alpha_a + \alpha_p + \theta_1 C_t + \theta_2\, Q_i \times C_t \tag{2.3}$$

$$+ \sum_{a=21}^{36} \{\theta_{3,a}\, Age_a \times C_t + \theta_{4,a}\, Age_a \times Q_i + \beta_a\, Age_a \times Q_i \times C_t\} + X'_{iapt}\delta + \varepsilon_{iapt} \tag{2.4}$$

where Y_{iapt} is educational attainment (completion of a given degree), $Q_i = 1$ is a dummy equal to 1 if the individual is born in Québec. C_t is an indicator of whether the individual is observed in the 2021 Census (= 0 if observed in the 2016 Census). α_a and α_p are age and province fixed effects, respectively. X_{iapt} is a vector of controls (gender, marital status, number of children). Parameters of interest is the vector β_a, which capture the intent-to-treat policy impact.



(a) High-school degree

(b) University degree

Figure 2.5: Long-term effect of the Québec childcare reform on children's educational attainment

Note: These figures plot the regression coefficients on the triple interactions (β_a) from equation (2.3) using the 2016 and 2021 Canadian Census of population. The horizontal axis represents the individual's age. 95 percent confidence intervals shown in brackets.

The results are reported in Figure 2.5. We find no evidence of negative effects on educational attainment of eligible children in the long-run. This pattern is true for each educational level, namely for university, high school, and college/CEGEP completion – the results for the latter outcome, the least comparable across Canadian provinces, is reported in Appendix Figure A.6. The results suggest a positive but statistically insignificant impact on completion of a university degree, the most comparable outcome across provinces, and no impact at lower levels. This null result is consistent with the long-run trends in educational attainment, which have been surprisingly parallel across Canadian provinces over decades (see Appendix Figure A.7).

In another specification, we assess potential heterogeneity in long-run impacts at the regional level. We estimate equation (2.3) separately for our two treatment groups depending on the level of local childcare expansions over the period of the policy. We use individuals' place of residence 5 years before the census year as a proxy for the place of birth, which is not available in the census. Results are reported in Appendix Figure A.8. As in the baseline model, there is no discernible difference in educational attainment by treatment intensity.

In light of the body of evidence documenting long-run effects of early-childhood circumstances on lifetime success, the absence of long-run impacts here might seem puzzling at first glance. However, several reasons could explain these findings. First, while Baker et al. (2019) find negative impacts on health and behavior, they obtain mixed evidence on cognitive skills. For example, they obtain insignificant impacts on some test score measures, but a positive impact on the mathematics component of PISA tests. Second, there is mixed evidence on the persistence of the short-term negative impacts on non-cognitive outcomes. While Baker et al.'s (2019) results suggest such persistence, Haeck et al. (2015, 2018, 2022) find that most negative impacts on children and parental behavior eventually fade away. Third, it is possible that compensating behavior of parents in their children's education (see Molnár, 2023) might have compensated for the impacts of daycare enrolment.

For the remainder of the paper, we treat these results (and findings in the literature) as evidence of no long-run fiscal externality from children's economic outcomes as they age. We nevertheless consider, in a robustness exercise, the long-run fiscal externality stemming from increased youth criminal activity documented by Baker et al. (2019) in our welfare analysis. Before moving to estimating the MVPF of the Québec reform, we describe the economic model used to infer mothers' willingness-to-pay.

2.4 Model

For our preferred MVPF estimator, we use a structural model of behavior to account for parents' behavioral responses and non-pecuniary gains of the policy. We consider a model of the family which follows Chaparro et al. (2020) (henceforth CSW), and that we adapt to our context. Our main departure from CSW is that we introduce rationing in the childcare market. Motivated by the evidence of shortages in daycare presented in Section 2.2, we refrain from assuming that the market is complete.[27] We further leverage our data on regional supply to explicitly introduce local coverage into the household's decision problem.

After describing the model, we briefly explain the numerical procedure to solve it. We then discuss identification of the model, where we exploit the natural experiment generated by the Québec reform to identify some of its key parameters in greater detail.

2.4.1 Setup

The model is that of a time-allocation problem of a mother (a unitary household) with a young child (after the parental-leave period) that has to meet her child's care needs. The mother weights the consequences of her choices on the child's development, the family

[27] As documented in Section 2.2.3, childcare markets in Québec have been characterized by important shortages for several decades. Thus, assuming that childcare is available at any quality (and associated price) to every household as in CSW appears unreasonable. Moreover, in the Canadian context, it is not the case that daycare prices are a strong predictor of quality (Seward et al., 2023).

budget, and her own preferences, thus providing a framework that highlights the key trade-offs families face (Becker, 1965, 1991; Del Boca et al., 2014; Berlinski et al., 2020; Chaparro et al., 2020).

Time constraints

A unitary household decision maker, which we refer to as the mother, with (at least) one preschool-age child makes a static decision on how to allocate her time T between market work L, child care at home T_m, and leisure ℓ. The mother's weekly time budget is given by:

$$T = L + \ell + T_m \tag{2.5}$$

Taking as given her child's baseline skills, denoted h_0, as well as household characteristics X, she jointly decides how to meet a child's care time constraint. While the child is awake (T_c hours), he must be cared for by the mother or in non-maternal care and thus we have:

$$T_c = T_m + T_d \tag{2.6}$$

where T_d is hours of non-maternal-care.[28]

Budget constraint

Each hour devoted to labor-market work is remunerated at a wage rate w. To finance non-maternal care expenses and household consumption C, she can use her own labor income as well as non-labor income I, which includes labor income from a spouse. The mother's budget constraint writes:

$$C + (1 - \tau_d(wL + I))\, p\, T_d = wL + I \tag{2.7}$$

where p is the hourly price of non-maternal care. The household receives refundable provincial childcare credits that account for a share τ_d of childcare expenses. The function τ_d is decreasing in household income, from 0.75 for the poorest households to 0.26 for households with income above \$48,000 (see Appendix Figure A.1 for a graph of this function).

[28]Non-maternal care includes hours in daycare centres, family-based daycare, and care by relatives. We therefore assume that care by relatives is not free. This assumption is often made in the literature because it is otherwise difficult to rationalize not taking-up free care by a relative (e.g. Bernal, 2008; Bernal and Keane, 2010; Apps et al., 2016; Griffen, 2019; Guner et al., 2020). Rather than assuming an *ad hoc* process for how informal care might be available to some families and not others, we follow this stream of literature.

Child skill production technology

In addition to choosing a time allocation (L, T_m, ℓ, T_d), the mother chooses *how* to parent. She cares about her child's achievements, which are determined through a child-development production function H taking as inputs the child's initial skills h_0 as well as time and quality of care in each care mode. We also allow the child's skills at the end of the preschool period h_1 (measured at ages 4-6) to depend on a vector of household characteristics X such as the number and ages of other children in the household and the parents' education.

$$h_1 = H(T_m, T_d, q, e; h_0, X) \tag{2.8}$$

where q is the quality of non-maternal care, which varies across households, and $e \in (0, \infty)$ is the effort devoted by the mother into adopting parenting practices that stimulate the child's development, for which we use reading time to the child as a proxy.[29] To measure non-maternal care quality, we use parents' satisfaction with the interactions the caregiver has with the child, how the caregiver praises the child, and the activities that stimulate learning as a proxy. As for daycare expenses, those variables are only measured for individuals using childcare and are only available in post-reform waves of the NLSCY. Therefore, for each household, we obtain a prediction for non-maternal care quality and the hourly daycare price using characteristics common to all waves (see Molnár, 2023). The variables used for prediction are the age and number of siblings in the household, parents' age, education and immigration status, and the size of the area of residence. Details of the predictions are presented in Appendix A.1.2.

The decision problem

Mothers' utility depend on household consumption, time and effort parenting the child, leisure time, and the child's skill accumulation when he reaches school age (4-6 years old). Additionally, mothers' preferences include a non-monetary cost of childcare use (when $T_d > 0$), intended to capture travel time to the childcare provider or search effort to find a spot. The mother's decision problem is to choose a time allocation $(\ell, L, T_m, \ell, T_d)$, a level of effort e, and a consumption level C to maximize her utility under the four constraints described above. Formally, the decision problem of a mother writes:

$$\underset{\Gamma}{Max} \quad U(C, \ell, h_1, T_m, T_d, e) \quad s.t.\ (2.5), (2.6), (2.7), (2.8) \tag{2.9}$$

where $\Gamma \equiv (C, L, \ell, T_m, T_d, e)$ is the vector of choices.

[29] Compared to Chaparro et al. (2020), who treat effort as unobserved and measure "quality of care", we instead use our data on parenting practices as a proxy of the former. CSW use items of the Home Observation for Measuring the Environment (HOME) index measuring parental Support for Learning and Literacy as their measure of maternal care quality. We do not have such a variable in the NLSCY data, so we proxy for parenting effort using the frequency of reading to the child.

2.4.2 Functional forms

Preferences. The mother's (household's) utility function is given by:

$$U = \gamma_c \ln(C) + \gamma_\ell \ln(\ell) + \gamma_h \ln(h_1) + \gamma_m \ln(T_m) - \gamma_{e,1}\, e\, T_m^{\gamma_{e,2}} - \psi(T_d) + \varepsilon \tag{2.10}$$

where ε is the unobserved component of utility and $\psi(T_d)$ is the disutility of childcare use (defined below). Maternal utility thus depends on consumption, leisure time, time and effort devoted to parenting, the child's human capital accumulation, and entry costs on the daycare market. A parameter of particular interest in this equation is $\gamma_{e,2}$, the non-linearity in the disutility of parenting effort, which in CSW's terms captures an "exhaustion effect" in maternal care. This feature of CSW's model is particularly relevant in our context since reducing exhausting parenting time represents an important source of non-pecuniary utility gain for parents.

Child skills production function. The child accumulates human capital in the preschool period based on the care received and his initial skills h_0. Child skill at age of school entry denoted h_1 is given by:

$$\ln(h_1) = \delta_0 \ln(h_0) + \delta_e \frac{T_m}{T_c} \ln(e) + \delta_d \frac{T_d}{T_c} \ln(q_d) + X'\delta_m + \eta_h \tag{2.11}$$

where η_h is a productivity shock and the δ parameters capture the productivity of various inputs in generating skills. In this specification, quality of care in each care mode is weighted by the share of time the child spends in it so as to ensure that a given care mode impacts the child's development only when the child is actually there.[30]

Disutility of childcare use. We assume that parents incur a fixed utility cost upon entering the childcare market (Berlinski et al., 2020). This cost can represent several unobserved aspects of parents' costs of childcare use such as travel time to the childcare provider or search effort. The non-monetary disutility of childcare use takes the form:

$$\psi(T_d) = \mathbb{1}[T_d > 0]\, (\gamma_{d,1} - \gamma_{d,2}\, CovRate) \tag{2.12}$$

where $\mathbb{1}$ is the indicator function and $CovRate$ is the coverage rate in the administrative region of residence.

We assume that the non-monetary cost of childcare use depends on the local childcare coverage rate. It is intended to capture the fact that increased local availability might

[30] We also estimated a specification allowing the productivity of each care mode to depend on initial skills h_0, thus permitting dynamic complementaries between childcare investments and baseline skills as in Cunha et al. (2010); Attanasio et al. (2020b,a). We, however, find little evidence for such complementarities in our context.

reduce costs associated with travel time or the burden of finding a spot in childcare. Indeed, Bravo et al. (2022) show that the reduced distance to the nearest daycare centre induced by a national expansion in Chile is valued by families. Similarly, De Groote and Rho (2023) find that families in Leuven, Belgium, highly value proximity to daycare providers.

2.4.3 Model solution

Given the potential presence (and importance) of corner solutions in the model, it has to be solved numerically. Combining the budget and time constraints (2.6), (2.5), and (2.7) and plugging them into the objective, we are left with three choice variables. That is, the mother chooses a time allocation (L, T_d) and a level of parenting effort e. The solution algorithm works as follows. We build a grid over the feasible time allocations (applying the time constraints) and the effort level. Then, for each combination of time and effort on the grid, we compute the utility level using equation (2.10) and find the vector yielding the highest utility on the grid.

2.4.4 Identification and estimation

We adopt a transparent multi-step identification strategy following CSW. The key advantage over a joint estimation algorithm is that it better isolates the sources of variation in the data we use for identification of key parameters of the model.

The first steps consist in identifying a set of parameters that do not require additional structure. These are parameters governing the child skills production function (2.11) and the exhaustion effect $\gamma_{e,2}$. Taking these productivity parameters as given, we then estimate the remaining preference parameters using a logit specification.

Child skill production technology

We first consider identification of the productivity parameters of the child skill technology (δ). We observe the time allocation of the child across different care modes as well as proxies for care quality as perceived by the person most knowledgeable (PMK) about the child. Her parenting practices and household characteristics are also observed. In our baseline model, we estimate equation (2.11) by OLS using our measures of child development. We include a set of control variables that might influence child development such as the age and number of siblings in the household and parents' characteristics (age, education, and immigrant status).

In this step, we leverage the panel dimension of the NLSCY, which has rarely been used in the literature, to obtain a measure of h_1 using the children observed longitudinally. We use test scores in the Peabody Picture and Vocabulary Test (PPVT) conducted during the home interview as a measure of endline skills. Development at early ages h_0 is measured using the Motor and Social Development (MSD) score. This MSD score is constructed by Statistics Canada using a series of questions on dimensions of the motor, social, and non-cognitive development of young children. The two indices are standardized by age in months and by 2-month age groups, each with mean of 100 and standard deviation of 15.

The baseline OLS estimator might be biased since childcare choices are potentially endogenous. Several variables in the error term η_h such as the mother's innate parenting ability might be correlated with childcare decisions and child development. Quality of care in each mode might also be subject to measurement error.

To address these identification issues, we also consider an instrumental-variable approach leveraging the policy change to identify δ_e and δ_d. Let $\tilde{X}_i \equiv [\frac{T_m}{T_c} \ln(e), \ \frac{T_d}{T_c} \ln(q_d)]$ denote the row vector of endogenous variables in (2.11). As candidate instruments for $\tilde{X}_i$, we consider the treatment dummy Eligible_{pt} and its interaction with low-coverage status. Sufficient conditions for identification are the usual instruments' relevance and exclusion-restriction conditions. The first identification assumption is that the instruments $Z_i \equiv [\text{Eligible}_{pt}, \ \text{Eligible}_{pt} \ x \ \text{LowCov}_r]$ are correlated with $\tilde{X}_i$. The policy impacts documented in Section 2.3.1 lend support for this assumption: the reform induced a large increase in childcare use and maternal employment, especially so in regions with higher daycare coverage. Second, the exclusion restriction requires that Z_i is uncorrelated with the error term in equation (2.11). The identification assumption is thus that, conditional on initial skills and household characteristics, the policy should impact child development only through childcare choices.

Identifying the exhaustion effect $\gamma_{e,2}$

We next consider identification of the curvature of the (dis)utility of effort $\gamma_{e,2}$. In CSW, this exhaustion effect is identified directly from reduced-form estimates of the IHDP experiment thanks to random treatment assignment. Our natural experiment gives us similar variation to exploit for identification. Specifically, we leverage the optimality condition for the effort choice and our reduced-form estimates of the policy's impacts to identify $\gamma_{e,2}$. The optimal effort level, which must be at an interior solution (it is not defined at 0), is given by the first-order condition of the maximization problem with respect to e:

$$e^* = \frac{\gamma_{h_1}}{\gamma_{e,1}} \, \delta_e \, \frac{T_m^{1-\gamma_{e,2}}}{T_c} \tag{2.13}$$

Taking logs on both sides yields:

$$\ln(e) = (1 - \gamma_{e,2}) \ln(T_m) + \chi \tag{2.14}$$

where $\chi = \ln(\gamma_{h_1}) - \ln(\gamma_{e,1}) + \ln(\delta_e(h_0)) - \ln(T_c)$. Thus, optimal (log) parenting effort is determined by maternal care time and a sum of productivity and preference parameters. As CSW note, a simple OLS estimator of $(1 - \gamma_{e,2})$ would be biased because T_m is likely correlated with preference and productivity parameters in χ. The model, however, assumes that parameters in χ are time-invariant. We could thus identify $\gamma_{e,2}$ with the simple differences in mean effort and maternal care time. Nevertheless, we refrain from using this direct approach because other macro shocks could have affected the productivity of parenting effort δ_e. We therefore consider using the quasi-experimental variation to identify the exhaustion effect. We use the more conservative assumption that potential average *changes* in χ conditional on individual characteristics X are the same in Québec and the rest of Canada.

Under this assumption, we can leverage our difference-in-differences estimates to identify $\gamma_{e,2}$. Given the evidence supporting the parallel-trends assumption (conditional on household characteristics X) for parenting effort and maternal-care time, our DiD estimates identify the Intent-to-Treat effects on those outcomes. Thus, conditioning equation (2.14) on X as well as province ($Q = 1$ for Québec) and a post-treatment dummy and then taking the double-difference yields:

$$\begin{aligned} &\text{ITT}[\ln(e)] = (1 - \gamma_{e,2})\,\text{ITT}[\ln(T_m)] \\ &+\mathbb{E}[\chi|Q=1, post, X] - \mathbb{E}[\chi|Q=1, pre, X] - (\mathbb{E}[\chi|Q=0, post, X] - \mathbb{E}[\chi|Q=0, pre, X]) \end{aligned} \tag{2.15}$$

where ITT$[A]$ is the intent-to-treat impact on variable A. Therefore, assuming that the evolution in χ (conditional on individual characteristics) is the same in Québec and the rest of Canada, the second line in equation (2.15) is null and the ratio of treatment effects on parenting practices and time identifies $\gamma_{e,2}$. This assumption is plausible in our context given the evidence in the literature of parallel trends between the two groups over a wide range of outcomes.

To lend some additional support for this identification assumption, in Appendix Table A.7 we report estimates of the child skill production technology in different Canadian provinces. Reassuringly, we cannot reject the hypothesis that the productivity of parenting δ_e in Ontario, the Western provinces, and Maritime provinces is the same as that in Québec. Parenting productivity thus should not had evolved differently in Québec compared to our control group, the rest of the country.

Preferences

Taking as given the primitives estimated in the previous steps, we estimate preference parameters on pre-reform data using the Québec sample only. We assume that the unobserved component of utility ε follows an i.i.d. Gumbel distribution, which yields a standard logit model for preferences (McFadden, 1974). This distribution for the unobserved component of utility has the well-known advantage of yielding a closed-form expression for choice probabilities. We estimate the preference vector $\gamma \equiv (\gamma_C, \gamma_\ell, \gamma_{h_1}, \gamma_{T_m}, \gamma_{e,1}, \gamma_{d,1}, \gamma_{d,2})$ by maximum likelihood.

2.4.5 Results

Before computing our MVPF estimates using the model in the next section, we begin by presenting the estimates of the model's main components. We then discuss the model fit and contrast the policy impacts estimated in the reduced form to the predictions of the model.

Model parameters

Table A.8 reports model parameters estimated in the first two steps, namely the exhaustion effect ($\gamma_{e,2}$) and productivity parameters (δ). For both specifications, we compare OLS and instrumental-variable models leveraging variation from the policy change.

Child human capital. In Panel A, we report the point estimates for the key inputs of the child human capital production function. These are the productivity of initial skills (δ_0), maternal care (δ_e), and non-maternal care (δ_d). We find that those three inputs are indeed productive of child human capital. Consistent with previous literature on child development, we find that early-age skills are highly predictive of future skills (see Cunha and Heckman, 2007, 2008). The OLS estimate (column 1) suggest that a 10% increase in the Motor-Social Development Score translates into a 2.05% higher PPVT score at ages 4-6.

Childcare time and quality are also positively associated with endline child skills. In the linear regression (column 1), we find that a 1% increase in parenting quality per hour in maternal care is associated with an increase in the child's PPVT score of 0.025% Similarly, a 1% higher quality per hour in non-maternal care increases the child's endline skills by 0.034%. Those parameters, however, should be interpreted with caution because movements in one input involve manipulating several endogenous variables: an increase in maternal-care time (T_m) implies a reduction of non-maternal care (T_d). Additionally, reading time (e) might depend on T_m through the exhaustion effect.

These baseline OLS estimates might suffer from omitted-variable bias through, for example, some unobserved innate parenting ability, which is correlated with childcare choices and reading time. In column (2), we thus report estimates of the instrumental-variable model presented in section 2.4.4. As in the linear regression, we find that both initial skills and quality of care are associated with increased child development. The IV estimates suggest, however, a larger role for both care modes in producing child human capital.

Exhaustion effect. In Panel B of Table A.8, we display the estimation results of the convexity of the parenting-effort cost $\gamma_{e,2}$. Columns (4) to (6) contain the results of the IV-type estimator using the policy change discussed in Section 2.4.4. As derived earlier, the exhaustion-effect parameter is given by $\gamma_{e,2} = 1 - \frac{\text{ITT}(\ln(e))}{\text{ITT}(\ln(T_m))}$. ITT estimates of the policy's impact on log reading time and log maternal-care time are reported in columns (5) and (6), respectively. We find that the reform led mothers to increase reading time by 0.08 log points and to reduce parenting time by 0.09 log points. These results suggest substitution between parenting time and effort, in line with CSW. They also imply a convexity in the cost of parenting effort (column 4) of $\gamma_{e,2} = 1.885$. Given that providing high-quality care is increasingly costly for parents, using childcare can provide some relief. Such reduction in the cost of parenting effort is potentially a key source of non-pecuniary gains for mothers.

Preferences. The final set of parameters is the preference vector γ. Table A.9 shows the estimation results of the discrete-choice model (2.10). As expected, parents derive some utility from leisure time and time spent with the child. Moreover, our parameter estimates suggest that parents assign a significant value to their child's achievements but that parenting (effort) is costly.

Of particular interest from the perspective of the documented heterogeneity in policy impacts by local daycare supply, our estimates reveal that increased daycare coverage substantially reduces fixed costs on the childcare market. In the last two rows of Table A.9, the parameter values indicate that in a hypothetical region with complete coverage (one space per preschool-age child) the entry cost would become negligible. These results again suggest a significant role for local daycare coverage in shaping parents' time-allocation choices. We now turn to a discussion of the model's fit before using our estimated model to compute the MVPF.

Model fit

To assess how accurate the model predictions are compared to the actual data, we use two approaches to test the validity of the model.

In-sample fit. First, in Table A.10, we assess the model in-sample fit by comparing the time-allocation choices predicted by the model to observed parents' behavior in the pre-reform data. Using our parameter estimates of the three steps along with 200 draws of the extreme-value type-1 distribution for each household in the pre-reform Québec data, we create 200 datasets of predicted choices. We then compare key market-share summary statistics from the pre-reform data (column 1) to predicted statistics from our simulated dataset (column 2). In the first three rows, we examine the performance of our simulations at predicting extensive-margin choices. We find that the model is doing a decent job for maternal employment and the share of households reading daily to the child, but over-predicts childcare use. At the intensive margin (last three rows), we find that the model cannot capture the difference between the hours worked by the mother and childcare utilization that is observable in the data. This is likely due to the strong incentives in the model to take-up childcare when the mother works. When the mother works full-time, we assume the child must attend childcare at least part-time. Moreover, when the mother works part-time, she has to sacrifice hours of leisure if the household does not use childcare. The results for the in-sample fit of the model are thus mixed. Nevertheless, as Kaboski and Townsend (2011) argue, the model's ability to reproduce the reduced-form impacts of an intervention is arguably a stronger basis for evaluating a model's usefulness.

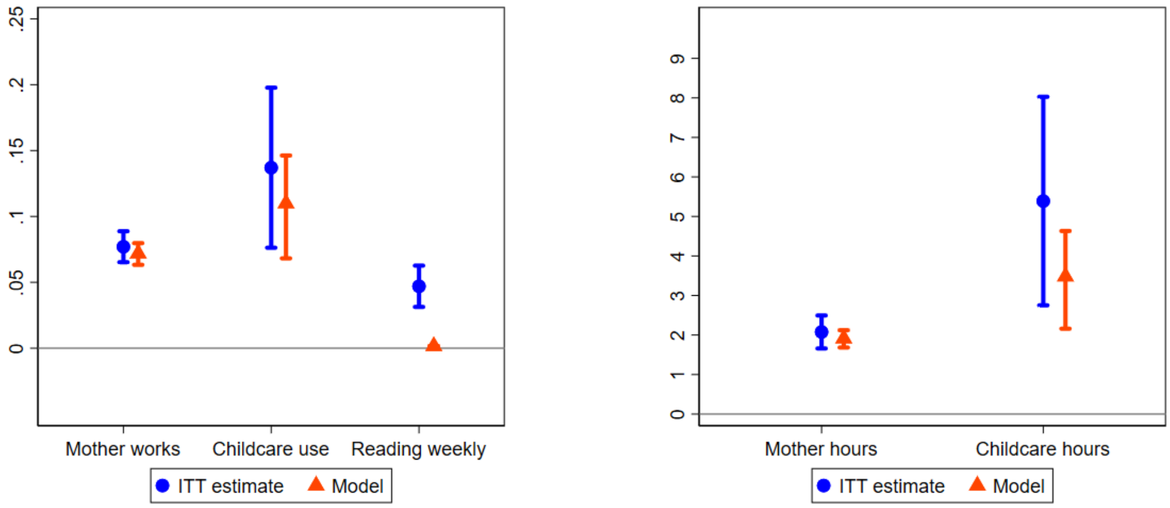
Out-of-sample validation. Thus, second, we perform an out-of-sample validation test by verifying whether the model predicts well the ITT estimates on maternal labor supply, childcare use, and time reading to the child (Tables 2.1 and 2.3). This validation exercise is similar in spirit to Chan and Liu (2018), who study a cash-for-care reform in Norway.[31]

We consider the three main aspects of the policy in our simulations: the price reduction in subsidized spaces, the local increase in supply, and the abolition of the refundable childcare credit. We map these features into the model by *(i)* reducing the hourly price of non-maternal care p_d to \$0.625 (assuming 8 hours of childcare per day) for (almost all) households facing a higher price, *(ii)* setting the credit rate τ_d to 0 for all households, and *(iii)* increasing the regional coverage rates $CovRate$ to their 2003 levels (second year of Cycle 5 data collection). Figure 2.6 summarizes the results of this exercise by contrasting the predicted behavioral responses to this policy experiment in our simulation sample to the ITT estimates from Tables 2.1, 2.2, and 2.3.

For inference, since the model predictions are complex non-linear functions of preference parameters, we compute the standard errors using the simulation procedure of Krinsky and Robb (1986). We draw 1,000 parameter vectors from a multinomial normal distribution and predict behavioral responses for each draw. Confidence intervals for predictions are then obtained from quantiles of the simulated distribution of labor supply and childcare choices.

We find that the model closely replicates the labor-supply response of mothers on both margins and also does a fairly good job for childcare use. Indeed, our simulation of

[31]This type of out-of-sample validation is also conducted by Todd and Wolpin (2006), Kaboski and Townsend (2011), and Chaparro et al. (2020), among others. See Todd and Wolpin (2023) for a review of empirical papers combining program evaluation with structural modelling.



Extensive margin　　Intensive margin

Figure 2.6: Out-of-sample validation

Notes: These figures display the results of our out-of-sample validation exercise, which compares the intent-to-treat estimates of the impact of the Québec reform (results in Tables 2.1, 2.2, and 2.3) with predictions from the policy simulation in the model. Standard errors on model predictions are computed using the simulation procedure of Krinsky and Robb (1986).

the policy predicts a 7.19 percentage points increase in maternal employment, which is very close to the reduced-form estimate of 7.7. Similarly, the model predicts an increase of 1.91 hours at the extensive margin, in line with the positive ITT estimate of 2.08 hours. For childcare use, the model predicts a very similar impact on take-up. On the intensive margin, hours, the model underpredicts the use of childcare, but predictions still lie within the confidence intervals of ITT estimates. The model predicts no response of time reading to the child, in contrast with the positive impact found in the reduced-form analysis. Nevertheless, the good fit of mothers' labor supply and childcare take-up suggests the model is useful to explain key non-marginal responses.

2.5 Welfare analysis

In this section, we turn to the main contribution of the paper, that is, estimating mothers' welfare gains, inclusive of non-pecuniary gains. We compare a benchmark estimator, using earnings gains as a sufficient statistic, to our structural estimator, which accounts for the fact that the policy change is non-marginal. Before delving into the calculations, we first present the welfare framework we consider and how we compute the two estimators in practice. We discuss the estimators via a brief theoretical exposition and through the lens of our model (a more general theoretical analysis is provided in Appendix A.2.2).

2.5.1 Welfare framework

We build on the approach described in Hendren (2016) and Hendren and Sprung-Keyser (2020) where the social welfare impact of a policy change can be measured by its marginal value of public funds (MVPF), which is defined as the ratio of the policy's benefit to its beneficiaries (measured as their willingness to pay for that policy) to the policy's net cost to the government. That is:

$$\text{MVPF} = \frac{\text{Beneficiaries' Willingness to Pay}}{\text{Net Cost to Government}} \tag{2.16}$$

The net cost to the government is given by the difference between the upfront government expenditure on the policy and fiscal externalities (i.e. indirect impacts on the government budget from changes in behavior). In our context, the main fiscal externality is the return to the provincial budget due to mothers' increased labor supply, which takes the form of increased taxes collected and reduced transfers and benefits.

The higher the MVPF is, the larger the welfare gains to the beneficiaries per net dollar spent. A ratio higher than 1 indicates that the marginal benefit of the policy exceeds its marginal costs to taxpayers, suggesting that the policy is welfare-improving. This approach has several desirable features. In particular, this metric can be used to make comparisons of welfare estimates across policy domains, thus permitting to study government policy from a broader perspective (Hendren and Sprung-Keyser, 2020). Additionally, compared to other standard metrics such as the cost-benefit ratio, the MVPF has two important additional advantages. First, while the standard marginal deadweight loss of public expenditure assumes an arbitrary linear income tax rate, the MVPF framework does not make any peculiar assumption on how the government finances the policy. Second, this framework quickly identifies "Pareto-improvements" from net costs.[32]

2.5.2 Willingness to pay

We start by the estimation of the numerator of the MVPF, a key contribution of this paper. As already mentioned, to compute the WTP, it is crucial to distinguish two types of reforms: infinitesimal versus discrete policy changes. For sufficiently small policy changes, it can be shown that, under standard assumptions, the WTP boils down to the treatment effect on beneficiaries' earnings (Hendren, 2016). We illustrate this result

[32]There are recent debates in Economics on the use of the MVPF as a welfare criterion to evaluate social programs. In particular, García and Heckman (2022a,b) criticize the use of this metric and suggest the use of an alternative criterion, namely the net social benefit (NSB). In robustness checks, in Appendix A.3, we compare our MVPF estimates to calculations of the NSB and the standard cost-benefit ratio. We find that, if anything, using one of these alternative criteria reinforces our main conclusion that omitting non-monetary gains for mothers substantially affects the social desirability of the policy.

below in the context of our model outlined in Section 2.4 and of our policy of interest. A more general exposition is provided in Appendix A.2.2.

Environment. Consider the model outlined in Section 2.4 in which the government chooses a childcare-provision policy characterized by a vector $\theta = (p, \tau_d, CovRate)$. As in Hendren (2016) we assume that the labor and consumption-good markets are competitive so that the policy state θ has no impact on prices in those markets.[33] The decision problem thus depends on the policy state θ, which the mother takes as given. Let $\Gamma^*(\theta)$ be the vector of optimal choices under policy state θ.

Substituting the time constraints into the budget, we can rewrite (2.7) as:

$$C^*(\theta) + p(T_c - T_m^*(\theta)) = w(T - T_m^*(\theta) - \ell^*(\theta)) + I(\theta) \tag{2.17}$$

Substituting the child time constraint into the child skill technology, we can rewrite (2.8) as a function of maternal-care time, parenting effort, and initial skills only: $h_1^*(\theta) = \tilde{H}(T_m^*(\theta), e^*(\theta); h_0)$. We are thus left with those two constraints. Let $V(\theta) = U(C^*(\theta), \ell^*(\theta), h_1^*(\theta), e^*(\theta), T_m^*(\theta), T_d^*(\theta); \theta)$ be the agent's indirect utility under policy state θ.

The government now implements a policy change. The reform moves the policy state θ from the status-quo policy θ_0 to some new policy state θ_1. The agent's WTP for this policy change can be measured by the standard equivalent variation (E.V.). That is, the WTP is the variation in income under θ_0 that would make the agent indifferent between the status quo and the new policy state:

$$E.V. = \frac{V(\theta_1) - V(\theta_0)}{\lambda(\theta_0)} \tag{2.18}$$

where λ is the mother's marginal utility of income.

WTP for a small policy change. Let us consider first, as is the case with the sufficient-statistics approach, that the policy change is infinitesimal. For an infinitesimal (marginal) policy change (in θ), at interior solutions, the numerator in (2.18), the difference in indirect utilities, is the total derivative of $V(\theta_0)$ with respect to θ. This

[33] In reality, we might suspect a price response of daycare providers in the private (non-CPE) network in the first years of implementation. For instance, to become more competitive with CPEs, we could expect unsubsidized daycare providers to lower their profit margin to retain some children in the private network. However, data we assembled from ministerial reports show that such pricing behavior is unlikely, at least over the time period considered in our analysis. As Appendix Figure A.9 shows, average real daycare prices in the for-profit network remained relatively constant from 1994 to 1999.

derivative yields:

$$\frac{dV(\theta_0)}{d\theta} = \gamma_{d,2} + \lambda(\theta_0)\,\frac{dI^*(\theta_0)}{d\theta} + \mu(\theta_0)\,\left[\frac{\partial \tilde{H}(\theta_0)}{\partial T_m}\frac{\partial Tm^*(\theta_0)}{\partial \theta} + \frac{\partial \tilde{H}(\theta_0)}{\partial e}\frac{\partial e^*(\theta_0)}{\partial \theta}\right] \quad (2.19)$$

where μ is the Lagrange multiplier on the child development constraint.
Proof. See Appendix A.2.1.

Therefore, the numerator of the WTP is the sum of three terms: the *direct* utility gain from the change in coverage, the *pecuniary* benefits, and the *non-pecuniary* gain stemming from the impact on child human capital. The intuition behind this result is the logic of the envelope theorem, which implies that, at the margin, behavioral responses do not have a direct effect on utility. Thus, if one additionally assumes that the utility gain from the change in coverage and child human capital gains are negligible, which may be reasonable for marginal family reforms, the difference in utilities $(V(\theta_1) - V(\theta_0))$ boils down to the policy's impact on the beneficiaries' budget constraint. The WTP is then simply given by the causal effect of the policy on earnings ($\frac{dI(\theta)}{d\theta}$). This result is powerful because it implies that the treatment effect on beneficiaries' earnings is a sufficient statistic for the numerator of the MVPF (Hendren, 2016; Hendren and Sprung-Keyser, 2020). This framework thus leverages the recent "credibility revolution" in the estimation of causal effects (Angrist and Pischke, 2010) to make transparent welfare statements.

WTP for a non-marginal policy. Consider now a discrete (large or non-infinitesimal) policy change. We refer here to a policy change that has first-order impacts on beneficiaries' utility. In this case, such as with the Québec childcare reform, the previous result no longer holds since envelope conditions only apply to marginal reforms. In particular, behavioral responses, both for labor-market behavior and time-allocation choices, now have direct impacts on utility because the agent re-optimizes behavior. Moreover, for large reforms, non-pecuniary gains (such as child development gains) are likely important. Thus, the treatment effect on earnings of beneficiaries is a biased estimate of the WTP.

Large-policy bias. The first bias, which we label the *large-policy bias*, stems from re-optimization behavior of beneficiaries. In the model above, it is equal to the policy's direct impact on utility through consumption and leisure time choices. Since agents make non-marginal changes in budgetary choices, these no longer have a null direct impact on the difference in utilities $(V(\theta_1) - V(\theta_0))$. This bias is potentially large in our context, given that the literature has documented major impacts of the reform on economic behavior. The large changes in maternal labor supply and child care use have direct impacts on utility through changes in mothers' time allocation, which are not captured by the treatment effect on earnings.

Non-pecuniary gains. Using the treatment effect on beneficiaries' earnings as an estimator of the WTP is subject to a second bias, namely the omission of non-pecuniary benefits of the policy. In this simple model, it is equal to the policy's impact on utility through coverage and child development gains, which is captured by the first and third terms in equation (2.19). This bias, in fact, also applies to small reforms, and Hendren and Sprung-Keyser (2020) themselves acknowledge that it may be important in some cases.[34] We argue that non-pecuniary gains (or losses) are likely to be large in the case of child care policies, perhaps even for small-scale programs. Indeed, preschool reforms may have substantial impacts on parenting time and practices and in turn on child development, which are all valued by parents.

Social willingness-to-pay. To derive the society's WTP, one has to aggregate individual preferences taking into account preferences of the overall society. Assume that there exists of a set of Pareto weights ψ_i for each beneficiary i. The social WTP is then simply the weighted sum of individual WTP for all beneficiaries with ψ_i as weights: $\text{SWTP} = \sum_i \psi_i \text{WTP}_i$. This flexible formulation, by choosing an appropriate set of weights, allows for example for social preferences for redistribution. We, however, focus on the case of a utilitarian planner who sets equal weights to every agent.

Benchmark estimator

As a benchmark, we consider an estimator of the MVPF that assumes the Québec reform is infinitesimal. In the small-policy scenario, in the absence of non-pecuniary gains, as shown above, the WTP is simply the treatment effect on beneficiaries' earnings. To obtain this benchmark estimate, we use our estimates of the pecuniary impacts on mothers and assume that these are sufficient to obtain the WTP. For better comparability with our structural estimator, we focus on short-term gains for mothers.

Willingness-to-pay. We first calculate the numerator of the MVPF under the benchmark estimator, which is the treatment effect of the policy on after-tax income. We thus use our results on the short-run impact on earnings from Section 2.3.2. To obtain the total WTP, we multiply the quantile treatment effects by the number of mothers in each particular quantile, which yields total earnings gains of \$2.469 billion.

Fiscal externality from mothers' short-term earnings gains. The second object we have to calculate is the return to the government stemming from behavioral changes.

[34]For example, in their estimation of the MVPF of admission to Florida International University, changes in effort at school or other forms of utility gains during college time are assumed away (Hendren and Sprung-Keyser, 2020, p. 1230). We discuss many other reforms for which non-pecuniary gains might be important in our survey of MVPF estimates in Appendix A.4.

There is a first fiscal benefit due to the increased labor supply of mothers with young children. At the extensive margin, entry of mothers into the labor market expands the tax base, thus increasing tax revenues for the government. Similarly, at the intensive margin, the government collects tax revenues on additional labor income. Moreover, a second fiscal benefit for the government comes in the form of reduced tax credits and transfer payments to families, since a higher household income decreases eligibility for tax credits.

To compute the net fiscal impact of mothers' responses on the provincial budget, we use the Canadian Tax and Credit Simulator (CtaCS) developed by Milligan (2019). The CTaCS is a comprehensive software that simulates the net fiscal position (at both provincial and federal levels of government) of an individual from a set of raw inputs (e.g. province, year, raw wage income, number of young children). We calculate the fiscal return for the government using our estimates of the impact of the reform on earnings. Given that an increase in earnings has a differential effect on additional taxes paid and reduced benefits along the income distribution, we rely on our quantile regression analysis to get a better sense of which mothers entered the labor market.

We perform the simulation in three steps. First, for each decile of the mothers' income distribution in our sample, we compute the net fiscal position of the average mother in that quantile. Then, we take the average earnings gain in a given quantile and assign it to families in that quantile. We then simulate the net additional taxes (of transfers and benefits) paid by the mother under the post-reform (1998) tax parameters. This procedure yields an estimate of the fiscal externality of $971 million.

For the purpose of our comparison between the benchmark and the structural estimators, we focus on mothers' short-run responses. For completeness of the social-welfare analysis, we further consider two other sources of fiscal externalities which have been identified in the literature.

Dynamic impact on maternal labor supply. First, Lefebvre et al. (2009) find evidence that mothers' earnings gains were lasting in the medium-run. They estimate a positive effect on earnings of mothers of older children whose child was eligible when younger of $1,995 over the period 1999-2004. Such as in our reduced-form analysis, we restrict dynamics in earnings impacts to that period to avoid capturing confounding effects of the Québec parental-leave reform of 2006, which had negative impacts on young mothers' earnings (Karademir et al., 2023). Thus multiplying these average earnings impacts by the number of mothers in two-parent families with children in those age ranges over that period, we obtain total earnings gains of about $1.102 billion for mothers of older children.[35]

[35] The data source for the number of mothers with children in given age ranges is the estimates from T1 Family Files of Statistics Canada (2023).

Youth crime. Second, Baker et al. (2019) find that children exposed to the reform at a young age experienced long-lasting negative consequences on behavior and non-cognitive outcomes. In particular, their results suggest that the policy increased youth crime at ages 12-20 among exposed cohorts as they aged.[36] As a robustness check, we monetize these additional societal costs to verify the sensitivity of our results. For the sake of space, we report the details of the calculations in Appendix A.1.3. Our back-of-the-envelope calculations using estimates of costs of juvenile crime of Cohen (2020) yields an estimate of the WTP to avoid these juvenile crimes of $20.16 million (in 1997 dollars) and a negative fiscal externality of similar magnitude. Thus, these costs are somewhat negligible compared to mothers' earnings gains which amount to billions of dollars and we omit them in the remainder of the analysis. Appendix Table A.11 reports the estimates of the MVPF inclusive of these additional societal costs.

Accounting for re-optimization behavior and non-pecuniary gains

To account for the large nature of the policy, we now refrain from assuming that envelope conditions hold. This poses a key challenge in that one can no longer express the WTP as a single treatment effect parameter. Kleven (2021) shows that practitioners would need to estimate both "policy elasticities" and changes in elasticities along the policy path, which is arguably beyond empirical reach. An alternative approach is to "make the policy marginal" in the spirit of Bravo et al. (2022). Those authors use variation from a national childcare expansion in Chile to evaluate the welfare effect of marginally reducing the distance to a childcare centre. We do not employ this strategy for two reasons. First, as in most of the policy-evaluation literature, we are interested in studying the effect of the reform as it was implemented. Second, and most important, estimating marginal treatment effects would not be possible in our case given that we do not have at hand an instrument (such as coverage rates) with sufficiently large support over the propensity score. In fact, even if we had such an instrument, defining a policy path for a policy changing several features of the economic environment would be difficult.

Instead, we use our estimated model to compute parents' WTP. To do so, we simulate the reform in the model and estimate the WTP by computing parents' equivalent variation as in Brink et al. (2007). The equivalent variation of a parent is given by equation (2.18). The marginal utility of income (λ) in our model is simply given by the inverse of the mother's total income net of childcare expenses. Denoting total income by $Y \equiv wL + I$,

[36] To be sure, prevalent youth crimes are rather "benign" offences such as thefts of small amounts, mischiefs, breaking and entering, failures to appear in court, and cannabis possession (Baker et al., 2019).

our structural estimator of the WTP thus writes:

$$\widehat{E.V.} = \frac{\widehat{V(\theta_1)} - \widehat{V(\theta_0)}}{1/\left[\widehat{Y(\theta_0)} - \left(1 - \tau_d(\widehat{Y(\theta_0)})\right) p_d \widehat{T_d(\theta_0)}\right]} \tag{2.20}$$

where $\widehat{Y(\theta_0)}$ and $\widehat{T_d(\theta_0)}$ are the predicted income and childcare hours under the status quo respectively.

To measure this quantity in the model, we perform the following steps using our simulated sample. First, we obtain the indirect utility in the status quo for each synthetic mother by finding the alternative on the grid yielding the maximum utility. Status quo income net of childcare expenses is then given by the difference between the household's total income and total childcare expenses at that point on the grid. Second, we perform the policy simulation described in Section 2.4.5 where we simulate choices under the key reform parameters (the offer of a \$5/day spot, the increase in coverage locally, and the abolition of the refundable credit). The estimated indirect utility in the post-policy state $\widehat{V(\theta_1)}$ is then given by the new maximum utility on the grid. With all the estimated components in hand, we compute each synthetic mother's WTP using equation (2.20). Last, we take the average WTP over the 200 simulated duplicates of each mother observed in the NLSCY. To obtain a representative sample of Canadian mothers, we use the sample weights provided by Statistics Canada. Because our simulation models the reform using 2003 coverage rates as the new policy state, to maintain comparability with the benchmark estimator that calculates average impacts over the roll-out of the policy, we sum the WTP over the pre-reform data only (two years).

Our structural estimator suggests a WTP exceeding that found using the benchmark estimator by a factor of two. We find that the total WTP amounts to more than 6 billion dollars. This result thus suggests that non-pecuniary gains are important in this context, which we further investigate through counterfactual simulations in Section 2.5.5.

Fiscal externality. The infinitesimal-policy assumption mostly has an important implication for the WTP. However, to obtain an internally consistent structural estimator of the MVPF, we also estimate the fiscal externality within our model. To do so, we calculate mothers' predicted income gains using our simulation of the policy. We then obtain the fiscal externality using the CTaCS calculator. To do so, we divide the sample into quartiles of predicted household income and use average household characteristics and income gains in each quartile as inputs for the calculator. The total fiscal externality is then obtained by multiplying the simulated fiscal impact for each quartile by the number of mothers in that quartile. We obtain an estimate of the fiscal externality that is higher (\$909 million) but comparable to our estimates using the benchmark causal effects. This similarity is reassuring and lends further support for the ability of our model to capture key policy impacts.

2.5.3 Direct cost

We now consider the upfront cost of the Québec reform, which comprises two main changes to the government budget. Those are the new subsidies allocated to the daycare market and the potential savings from the abolition of the refundable childcare credits.

The main fiscal operation is the increased subsidies allocated to the daycare market. These expenditures take various forms: start-up grants, recurring operating grants to daycare centres, special needs, and other subsidies. We first sum the total subsidies over the period covered in our empirical analysis, that is, over fiscal years 1997-1998 to 2002-2003. Figure A.10 shows the evolution of total subsidies to daycare facilities along with the subsidy per space. The graph shows that the rapid expansion of daycare supply over the end of the 1990s and early 2000s induced major increases in government spending. Total subsidies increased from about \$470 million in fiscal year 1998-1999 to \$1.206 billion in 2002-2003. However, a share of these would have most likely been spent by the government absent the reform. As can be seen in Figure A.10, the government was spending nearly \$300 million in the two previous fiscal years. We assume similar subsidies would have been granted absent the policy change. We project these expenses assuming the same conservative growth rate observed from 1996 to 1997 (2.1%) and deduct these counterfactual subsidies from the observed grants. These calculations yield a total amount of new subsidies allocated of \$2.617 billion over our study period.

To lower the impact of subsidized daycare on public finances, the Québec government made simultaneous changes to other family allowances. In particular, for families obtaining a low-fee space, the refundable childcare credit available before the policy was abolished. To obtain an estimate of the savings generated by this fiscal policy change, we compare government expenses in this expenditure item before and after the reform. Total childcare credits allocated to families are retrieved from the Québec Ministry of Finance's annual budget. In 1996, the fiscal year just before the reform, the credit cost \$192 million. In 2001, the same amount was allocated to this program (\$191 million). Given the sharp decrease in the number of subsidized spaces (who became subsidized), this pattern is surprising. Perhaps some families, who were not claiming the credits before the reform, suddenly started doing so. We are thus reluctant here to attempt to impute what would have been spent by the government absent the Québec reform. Those savings would most likely be small in any case compared to the size of the subsidies (Fortin et al., 2013). We thus prefer to consider an upper bound on direct costs and abstract from potential savings from this source.

The reform, as expected for universal preschool subsidies, is costly. In net, abstracting from potential savings from the abolition of the refundable childcare credit for the reasons detailed above, the Québec government spent \$2.617 billion on the policy over our study period. A careful evaluation of the benefits generated by the reform is thus crucial to assess whether the policy yielded a positive return to society.

2.5.4 MVPF estimates

The MVPF is the ratio of the WTP to the net cost of the reform. The net cost to society is the difference between the upfront expenditure and fiscal externalities. We use our estimates from the analyses above to calculate the MVPF of the Québec childcare reform under the benchmark estimator and the structural one. Table 2.5 displays the different components of our MVPF calculations, which we describe below.

Benchmark estimator. First, under the benchmark sufficient-statistic estimator, we obtain an estimate of the net cost of about \$2,617M - \$971M, which yields a net expense of \$1.646 billion. Mothers' willingness-to-pay, captured by their earnings gains under the benchmark estimator, amounts to about \$2.344 billion in after-tax income (substracting the fiscal externality above to the raw earnings gains). Given the absence of evidence of long-run impacts on children from section 2.3.3, the benchmark estimator of the MVPF suggests parents were willing to pay about \$1.42 per net dollar spent on the reform by the government. This is a small MVPF compared to targeted preschool programs studied in Hendren and Sprung-Keyser (2020).

Structural estimator. Second, our structural estimator suggests a similar, but slightly lower fiscal externality from treated mothers of \$909 million. The willingness-to-pay, however, differs sharply. Including non-pecuniary gains for mothers more than doubles the willingness-to-pay. As a consequence, our estimate of the MVPF more than doubles as well and reaches 3.56. This estimate is much closer to MVPF estimates for targeted preschool interventions appearing in the Policy Impacts Library of Hendren et al. (2023).

2.5.5 Policy counterfactuals

Thus far, we have focused on estimating the MVPF of the adopted reform. Our model, however, can also be informative about *(i)* the main mechanism driving mothers' labor-supply response and *(ii)* whether the government could have obtained higher welfare gains under alternative policy schemes. Using our estimated model, we perform counterfactual simulations to shed light on these two questions.

Mechanism. First, we ask which feature of the policy is responsible for the bulk of the welfare gain. We simulate counterfactual scenarios in which we remove each feature of the policy one-by-one and compute the WTP under these alternative scenarios. First, we

Table 2.5: Welfare estimates

MVPF components	Mean values	External sources used
Direct cost	\$2,617M	Québec Treasury Board
Benchmark estimator		
Fiscal externality		
Tax returns and reduced transfers	\$971M	CTaCS
Willingness-to-pay		
Mothers of preschoolers	\$2,213M	CTaCS
Mothers of older children	\$1,102M	Lefebvre et al. (2009)
Taxes and reduced transfers	-\$971M	CTaCS
MVPF	1.42	
Structural estimator		
Willingness-to-pay	\$6,078M	
Fiscal externality	\$909M	CTaCS
MVPF	3.56	
Counterfactual WTP		
No price change	\$5,120M	
No coverage increase	\$362M	
Credits maintained	\$6,178M	

Notes: This table outlines the components of the MVPF under the benchmark and the structural estimators. The last column reports the external sources used for the policy's cost and other sources of fiscal externalities. The acronym CTaCS refers to the Canadian Tax and Credit Simulator of Milligan (2019). The last three rows present the values of the willingness to pay (WTP) under three counterfactual scenarios: (*i*) no price change, (*ii*) no coverage increase, and (*iii*) childcare credits maintained.

implement only the price decrease (or only the increase in coverage) while maintaining the abolition of the refundable credits. Then, we simulate the actual reform, but maintaining the refundable credits.

Our results, displayed in the second part of Table 2.5 suggest that most of the welfare gains are due to increased coverage. Indeed, we find that the WTP for the price reduction only is very small (only \$362M) compared to that of the actual policy. However, the WTP for increasing coverage rates to their 2003 level without decreasing the price is as high as 84% that of the actual reform, further suggesting that the decrease in price is not the main driver. In our model, given that coverage operates through a reduction in entry cost (on the childcare market) and thus does not set a cap on the childcare-use response, this result is not simply mechanical. Last, not abolishing the childcare credits has a negligible effect on behavior, which is not surprising given that, at the reduced fee, obtaining further discounts only slightly lowers the net price.

Our results thus suggest that increasing childcare availability is key for the effectiveness of universal preschool policies. Those results are in line with De Groote and Rho (2023), who find large welfare gains of increasing daycare capacity on a centralized Belgian platform. In particular, they show that even a small increase in daycare capacity is sufficient to compensate advantaged families for their welfare loss under affirmative-action policies.

Alternative policies. Last, we compare the MVPF of the adopted reform to changing the main features of the policy. Specifically, we ask whether the Québec government could have achieved higher welfare gains under different price-coverage combinations.

To provide some insights into this question, we simulate behavioral responses under multiple price-coverage pairs. For simplicity, we assume a uniform coverage rate throughout the province. For each pair, we proceed in three steps as follows. First, we obtain mothers' WTP by calculating the counterfactual equivalent variation (2.20) using our synthetic datasets generated for the simulation of the actual reform. Second, we compute the counterfactual fiscal externality using the CTaCS calculator following the same approach as for the actual reform.

Third, we compute the counterfactual direct costs. Counterfactual government subsidies are given by the difference between the counterfactual societal (total) costs and parents' payments under the counterfactual scenario. These are impacted by both the change in the price paid by families and by the number of spots that need be created to reach the counterfactual coverage rate. We assume that the subsidies paid by the government vary linearly with the expenses made by families. In other words, this means that every additional dollar paid by families reduces government expenditures by one dollar. Next, we have to take into account the fact that, in counterfactual scenarios with high coverage rates, not all spaces are filled. Thus, in such cases, parents do not pay for every existing space, but only for those that they actually use. Consistent with the typical childcare contracts in Québec in that period (see footnote 13), we further assume that daycare centres operate over the maximum number of days (260 days).[37]

The results are reported in Figure 2.7, which shows how the simulated MVPF (where the MVPF of the actual reform is normalized to 1) varies with price and coverage. We find a striking pattern: social welfare gains are generally increasing in daycare coverage but also with the fee charged to families. Together with the large WTP for increased daycare

[37] Counterfactual government subsidies, denoted by G', are then given by the following expression:

$$\begin{aligned} G' &= \text{(Total costs)}' - \text{(Parents' payments)}' \\ &= (G + \$5 \times \text{spaces} \times 260) \times \frac{(CovRate)'}{CovRate} - p' \times 260 \times min\left\{(\text{childcare use})', (CovRate)'\right\} \times pop \end{aligned}$$

where *pop* is the population of preschool-age children in the post-policy period and G is the total government subsidy under the actual reform. Variables with a $'$ sign refer to the values of these variables under counterfactual scenarios.

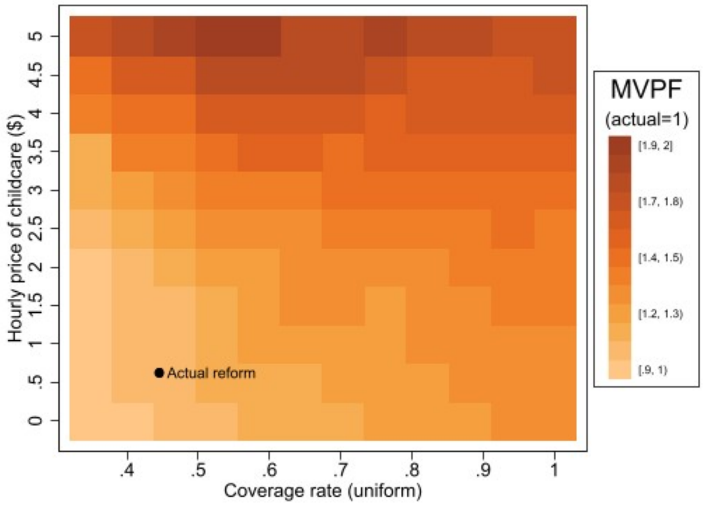


Figure 2.7: MVPF under counterfactual price-coverage combinations

Note: This figure plots the simulated counterfactual MVPF under different price and uniform daycare coverage combinations. The MVPF of the actual reform is normalized to 1. Darker colors represent higher values.

supply, this finding suggests that the government could have achieved larger welfare gains by channelling more resources towards opening spots rather than to lowering prices. We note, however, that there is an important caveat in this analysis. Our empirical model is a partial-equilibrium framework and thus abstracts from general-equilibrium effects. In simulations with large coverage-rate increases, the substantial increase in maternal labor supply that would be predicted should have general-equilibrium implications such as placing downward pressure on wages in reality. Thus, we interpret the results with caution. Nevertheless, this exercise is informative about the direction in the price-coverage space where gains are likely to be larger. For example, compared to the actual reform, counterfactual estimates suggest the MVPF for a reform that doubles the price charged to families and increases the coverage rate by 5 percentage points would be 13% larger.

2.6 Conclusion

Childcare policies may impact social welfare through various channels. Availability of subsidized childcare can reduce the opportunity cost of employment, particularly for mothers, thereby potentially increasing the tax base. The quality and accessibility of childcare options can influence human capital development of children, potentially shaping their future labor-market outcomes. Moreover, increased availability reduces non-monetary costs associated with childcare use, such as time spent commuting to the caregiver and search effort to find a spot when supply is initially limited.

This paper incorporates these various channels into a comprehensive welfare analysis of universal preschool provision. We uncover new patterns regarding the impacts on parental behavior resulting from a universal program implemented in the late 1990s in Québec, leveraging novel data on daycare availability at the local level within the province. We show that the positive impacts on maternal labor supply and childcare use are larger in regions where daycare expanded more. These results suggest that the relief of capacity constraints at the local level, not just daycare affordability, is an important channel through which preschool reforms can boost maternal employment and childcare utilization.

Building on this insight, we estimate the value of the policy for mothers using a structural model of maternal labor supply and childcare use that incorporates the benefits of increased availability. In doing so, we explore the extent to which standard assumptions made to estimate sufficient statistics for social welfare might yield misleading results in empirical welfare analysis when applied to non-marginal reforms. Our study demonstrates how combining a reduced-form causal analysis and estimation of a tractable structural model can provide empirical insights into the magnitude of such bias. For policies with significant costs, overlooking non-pecuniary gains might compromise the conclusions about the welfare implications of a given policy. In the context of the Québec reform our estimates indicate that the benefit-to-net-cost ratio is more than twice as large when these gains are considered.

This study suggests three lessons for empirical welfare analysis of preschool reforms. First, it is the first paper to show that universal preschool reforms can yield substantial welfare gains, in particular in the form of non-pecuniary benefits for mothers. Second, it highlights the limitations of sufficient-statistic methods in welfare analysis, often implicitly used in cost-benefit analyses of large reforms. We show that, when applied to non-marginal preschool policies, this approach might omit key welfare gains that are empirically relevant. Third, it underscores the importance of local daycare supply in shaping policy impacts. Our analysis suggests that mothers have a high willingness to pay for an increase in childcare availability.

This first attempt at measuring the bias of sufficient-statistics metrics when applied to universal preschool reforms raises several questions. Studying non-marginal policies comes at the "cost" of structural assumptions on the economic problem and perhaps realism. While our model predicts the key maternal behavioral responses well, future research could extend the framework to consider labor-force dynamics or general-equilibrium effects. Other potential impacts on the Québec economy, such as gains for firms who hired mothers entering the labor force, could be explored. In addition, for cost-benefit analysis, it is essential to assess whether the sufficient-statistic approximation is reasonable in other policy domains. Given the importance of conducting appropriate cost-benefit analyses for policy, more evidence is needed to better understand which assumptions on economic behavior are reasonable in different contexts.

Appendix A

Supplementary Material for Chapter 2

A.1 Data Appendix and additional results

A.1.1 Data sources

Different sources of Canadian microdata are used in this paper. We provide a brief description of these sources in this Appendix.

The Canadian Census of Population is conducted every five years since 1981 and collects information on all household members from a 20% to 25% sample representative of the Canadian population. It covers a variety of topics, of which we notably extract the province of birth, demographic characteristics as well as educational attainment. We use the 2016 and 2021 waves of the Census for the empirical analysis of children's long-run outcomes.

The NLSCY studies the development and well-being of Canadian children. Children are followed bi-annually from birth to early adulthood and the information collected includes a range of indicators of socio-emotional, cognitive, and behavioral development. The survey series began in 1994 with an initial sample of children under the age of 12, which was followed for 14 years (at two-year intervals) through 2008. In each wave of the survey, a new cohort of children under the age of two was added to focus on early childhood development. On top of including detailed information on child development outcomes, the survey also includes a questionnaire given to the parent who is deemed the most knowledgeable about the child. In that section, the parent provides information on family functioning, parental support, labor supply, and most importantly, the time allocation of care among various options (daycare center, at home by relative, at home by non-relative, and so on). Weekly expenditures on care are also reported by the parent in the last two waves and we use this data to infer the price of private-market care.

The Canadian Labor Force Surveys (LFS) are annual surveys of the working-age population – excluding some specific categories of Canadian households (individuals in Aboriginal reserves, members of the Canadian Forces, and the institutionalized) – that include labor-market and basic demographic variables. Since they include the exact date of birth, we can precisely identify eligible cohorts in the reduced-form analysis. We use the summary Table 37-10-0130-01 of Statistics Canada using the LFS for plots of long-run trends in education across Canadian provinces reported in this Appendix.

A.1.2 Measurement and predictions of variables

To estimate our structural model, we have to impute values for missing data on variables of interest. These include wages and non-labor income, childcare expenses, and quality of non-maternal care. We describe below how we measure these variables and

Wages and non-labor income. To estimate the model, we need to assign wage offers and to measure non-labor income for every household. This includes predicting a wage offer for non-working mothers as well as imputing the wage rate and non-labor income when income information is missing. In the NLSCY, the person most knowledgeable about the child (PMK) reports wages (for both the PMK and the partner) as well as household income. Given the absence of policy impacts on fathers' labor supply, we treat the father's income as non-labor income from the mother's point of view. We thus measure non-labor income as the difference between the reported household income and the mother's labor earnings (wages and self-employment income). We thus estimate Mincer-type models to predict real wages and income for those households. Variables used for predictions are the age and number of siblings in the household, parents' age, education and immigration status, the size of the area of residence, and a set of Census Metropolitan Area (CMA) dummies to capture local labor market variation.

Childcare expenses. The NLSCY contains measures of childcare expenses in the last two cycles. Respondents report their weekly expenses on childcare in cycles 7 and 8. We follow Molnár (2023) and measure the hourly price of childcare by dividing weekly expenses by the number of hours in institutional care. We make sure to remove households who have a subsidized space, which is observed in those waves. We then obtain predictions for childcare expenses in pre-reform data using variables common to all waves. Variables used for predictions are the age and number of siblings in the household, parents' age, education and immigration status, and the size of the area of residence.

Non-maternal care quality. We measure non-maternal care quality by constructing an index from three survey questions available in cycles 3 and 4. These questions concern parents' satisfaction with the interactions the caregiver has with the child, how the caregiver praises the child, and the activities that stimulate learning. They are phrased as follows:

> *How often would you say your caregiver praises and encourages [CHILD'S NAME], and responds promptly when he/she needs help or comforting?*
> *(1) Never (2) Rarely (3) Sometimes (4) Often*
>
> *How often does your caregiver plan activities and use toys and other materials to help [CHILD'S NAME] learn new things?*
> *(1) Never (2) Rarely (3) Sometimes (4) Often*
>
> *How often does your caregiver encourage [CHILD'S NAME]'s language development by talking to him/her and asking questions, as well as using songs and stories for this purpose?*
> *(1) Never (2) Rarely (3) Sometimes (4) Often*

Variables used for predictions are the age and number of siblings in the household, parents' age, education and immigration status, and the size of the area of residence.

A.1.3 Youth crime

In this section, we investigate the robustness of our main results to including long-run costs of juvenile criminal activity. Baker et al. (2019) find that children exposed to the reform at a young age experienced long-lasting negative consequences on behavior and non-cognitive outcomes. In particular, their results suggest a positive impact of the policy on youth crime at ages 12-20.

To be sure, prevalent youth crimes are rather "benign" offences such as thefts of small amounts, mischiefs, breaking and entering, failures to appear in court, and cannabis possession. Through the lens of the MVPF framework, increased criminal behavior can impact welfare through two channels: additional costs to victims and productivity losses for offenders, which reduces the WTP for the policy, and additional costs on the police and criminal justice systems, which is a negative fiscal externality. To take into account these costs to society, we perform a back-of-the-envelope calculation using estimates of costs of crime reported in Cohen (2020). Since these costs appear many years after the enaction of the policy, we apply a discount factor of 3% following Hendren and Sprung-Keyser (2020). However, results are qualitatively robust if we do not discount.[1]

We focus on the estimates from the richest specification (Table 5, column 3) in Baker et al. (2019). They find an average increase in yearly youth criminal activity of 212 crimes per 100,000 inhabitants. Given that crime rates in Québec are very low, this

[1] We obtain a WTP to avoid the committed crimes of \$32.11 million and a fiscal externality of \$30.83 million when we do not discount. Results are available upon request.

Table A.1: Costs of the Québec childcare reform from increased youth criminal activity

Type of crime	Victim costs	CJS costs	Offender productivity	Impact (BGM)	WTP	Fiscal externality
Persons (assaults)	\$9,145	\$3,594	\$524	167 [59]	\$17.71M	\$6.58M
Property (theft)	\$251	\$1,922	\$89	342 [93]	\$1.28M	\$7.21M
Drugs[1]	0	\$4,523	\$786	99 [29]	\$0.85M	\$4.91M
Other[2]	0	\$176	\$86	239 [54]	\$0.32M	\$0.65M
Total					**\$20.16M**	**\$19.36M**

Notes: Costs of crime estimates are taken from Cohen (2020) and are converted in 1997 Canadian dollars using the average exchange rate in 1997 (1.3252CAD/1USD). For each crime category, we use the crime most often committed by Canadian youth (in parentheses) as reported in Baker et al. (2019) (BGM). These cost estimates include crimes committed by adults, which are more costly on average, and should thus be interpreted as upper bounds. The WTP column should be interpreted as the WTP for avoiding the committed crimes and is the sum of the victim and offender productivity costs. The estimated policy impacts are taken from Table 5, column 3 of Baker et al. (2019). Standard errors are reported in brackets.

[1] The most prevalent drug crime is cannabis possession, but the data does not allow us to distinguish between drug possession and sale. These estimates are thus likely to be upper bounds.

[2] The most prevalent crime in the "other" category is failure to appear in court, but the data does not distinguish between types of "other non-traffic violations".

figure represents a rise of 22% of the mean. The authors further break down the crimes into four categories: against persons (rise of 167 crimes), against property (rise of 342), drugs (rise of 99), and other convictions (rise of 239). Since cannabis possession is likely not very costly and now legal in Canada, we focus on the other three categories. For each category, we consider the costs of the most common crime, which are non-aggravated assaults, theft of less than \$5,000, and failure to appear in court.

We first multiply the crime rates impacts by the population of exposed youth in each post-reform cohort considered in the original study to obtain the total yearly impact.[2] Second, we multiply this number by the victimization and offender productivity costs, which enter the WTP, and the government services costs (on the criminal justice system), which imply a fiscal externality. Third, for each post-reform cohort, we discount future costs to obtain the actualized value of increased youth crime. The results are reported in Table A.1. We obtain that these costs, both on the WTP and the fiscal externality, are

[2] The population of Québec residents aged 12 to 20 years old was approximately 850,000 over the years considered for this analysis (Institut de la Statistique du Québec, 2022).

about 20 million dollars. They are thus negligible compared to benefits stemming from mothers' earnings gains.

A.1.4 Appendix Figures

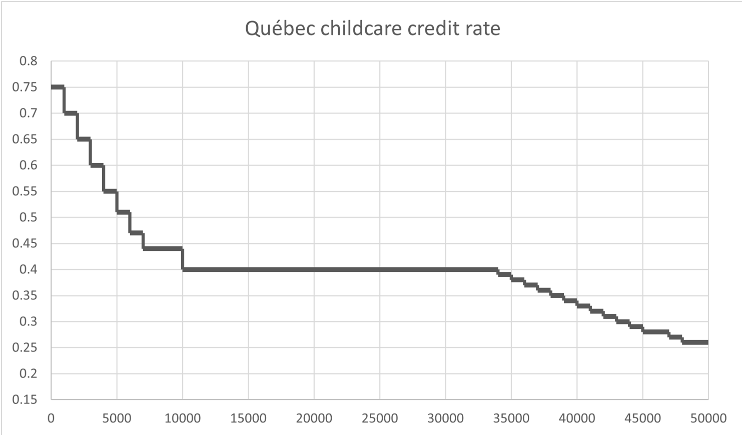
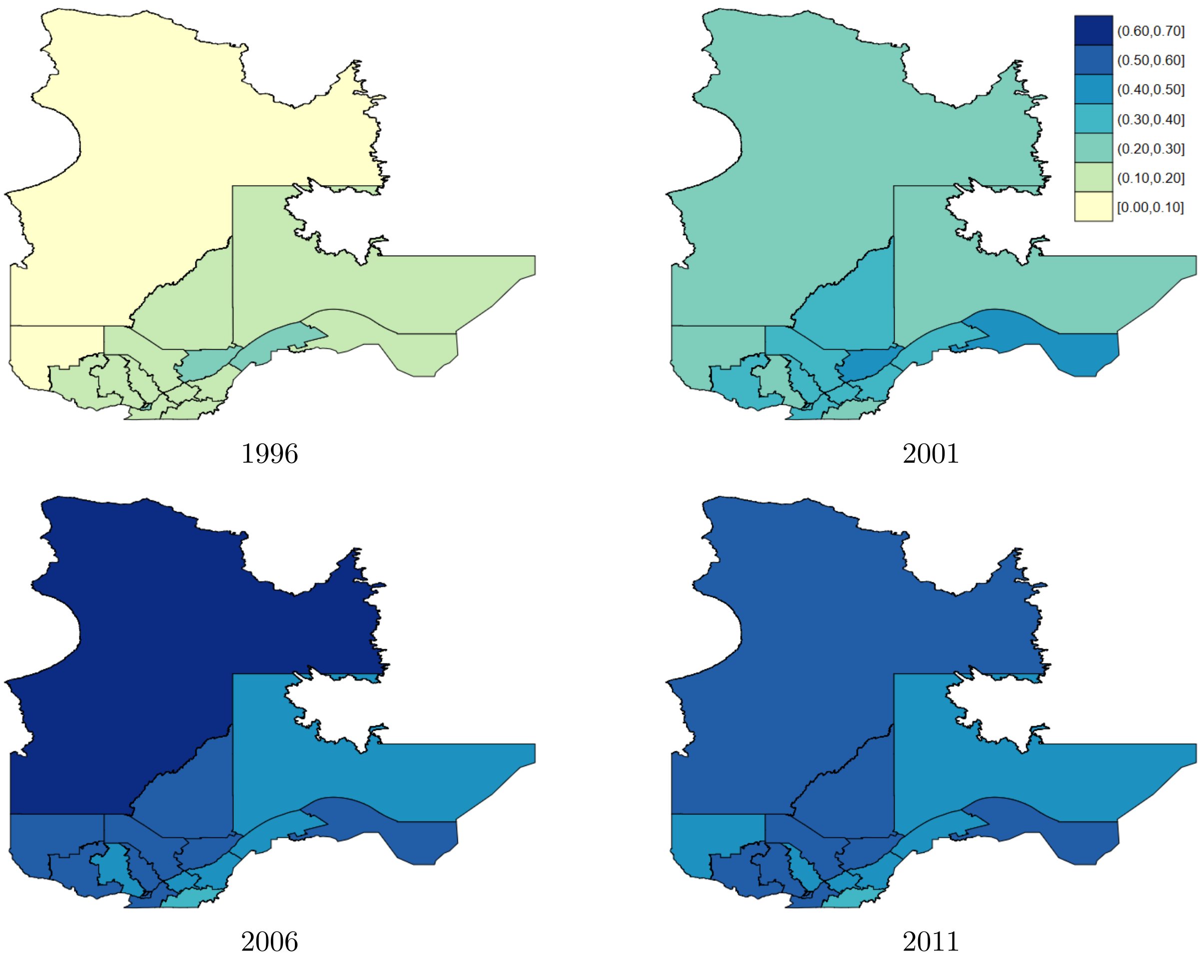


Figure A.1: Refundable childcare credit rate by family income in Québec

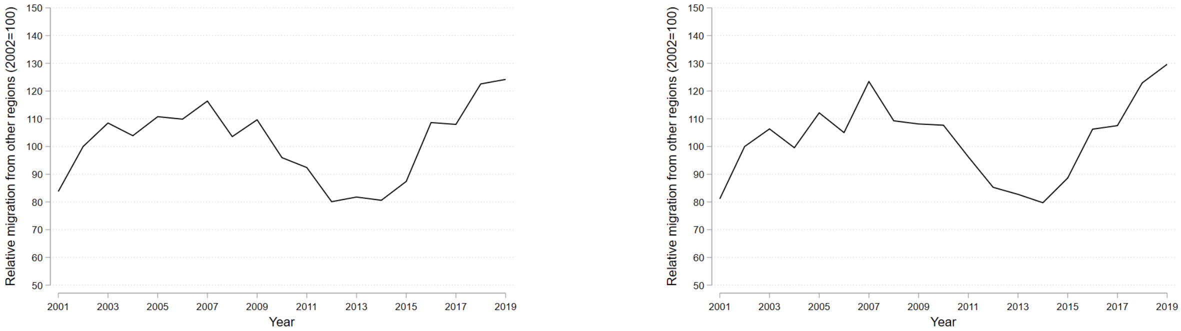
Note: This Figure plots the refundable childcare credit rate as a function of family income. The refundable childcare credit was available to all families who used paid childcare before the 1997 daycare reform. After 1997, families using subsidized childcare were no longer eligible.

Figure A.2: Evolution of the total number of daycare spaces per children aged 0-4 years by administrative region, Québec



Data sources: Ministry of the Family for number of spaces and Institut de la Statistique du Québec for population of preschoolers

Figure A.3: Evolution of inter-regional migration by childcare expansion status



Top tercile Bottom tercile

Data sources: Ministry of the Family for number of daycare spaces and Institut de la Statistique du Québec for population of preschoolers and net inter-regional migration.
Notes: These figures display the evolution of the net inter-regional migration flows of preschoolers (0-4 year olds) in two groups of regions relative to 2002 (normalized to 100). The left panel shows the changes in migration to regions which are part of the top third of regions who experienced the largest childcare expansion (measured as the increase in their coverage rate from 1997 to 2003). The right panel shows the equivalent time series for the bottom third.

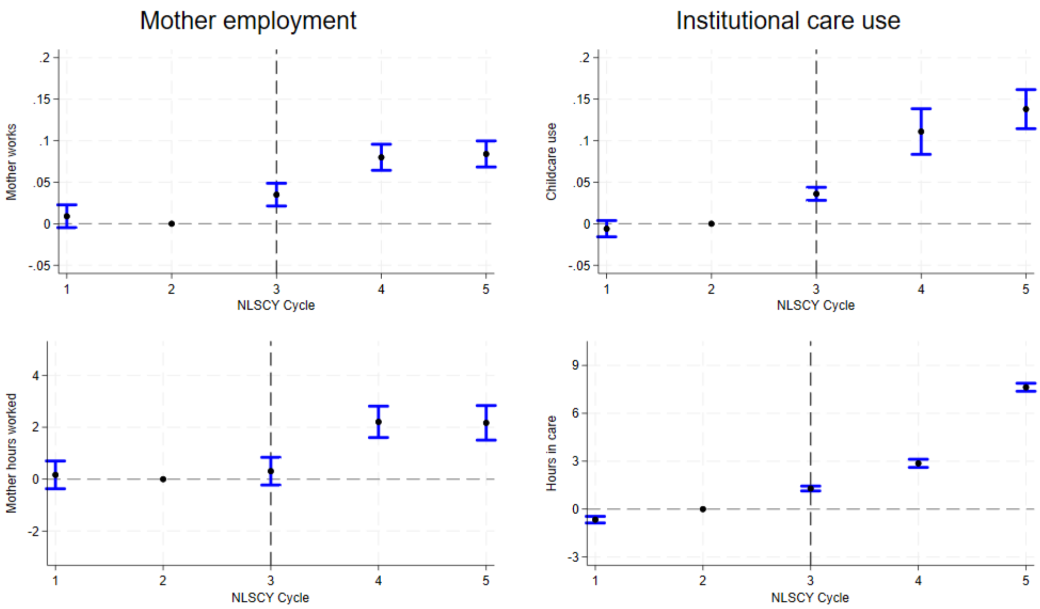


Figure A.4: Dynamic impact of the Québec childcare reform on maternal supply and institutional care use

Note: These figures plot the coefficients of event-study regressions along with 95% confidence intervals. The data source is the first 5 waves of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), and immigration status (both parents). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates.

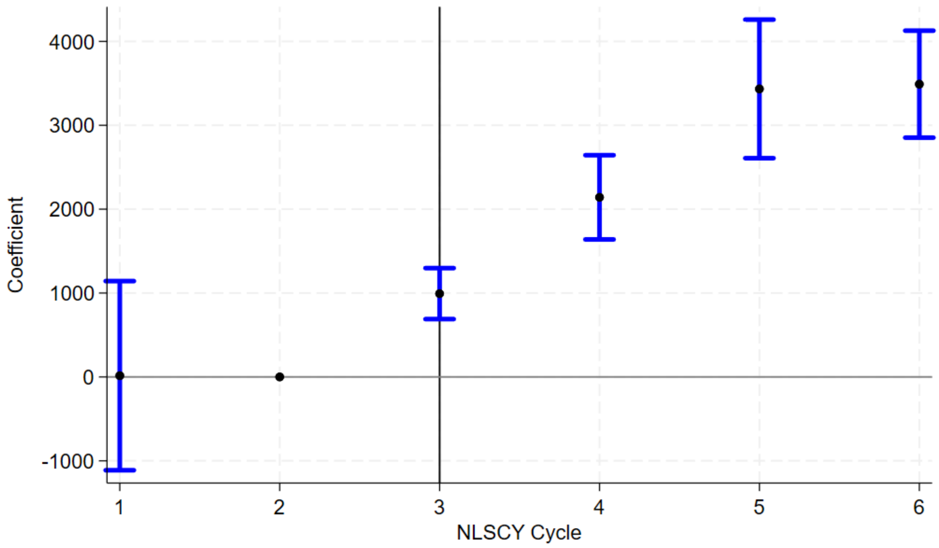
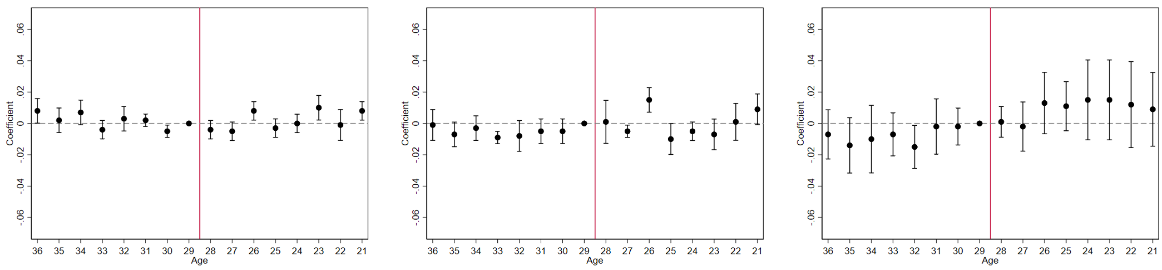


Figure A.5: Dynamic impact of the Québec childcare reform on mothers' labor earnings

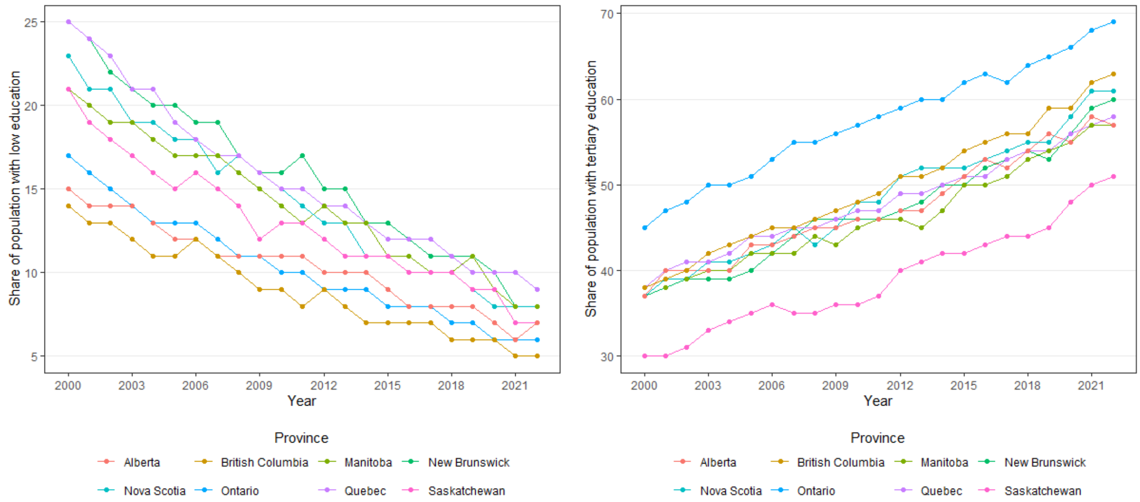
Note: These figures plot the coefficients of event-study regressions along with 95% confidence intervals. The data source is the first 5 waves of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), and immigration status (both parents). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates.



(a) High-school degree (b) College degree (c) University degree

Figure A.6: Long-term effect of the Québec childcare reform on children's educational attainment

Note: These figures plot the regression coefficients on the triple interactions (β_a) from equation (2.3) using the 2016 and 2021 Canadian Census of population. The horizontal axis represents the individual's age. Standard errors clustered at the province level in parentheses.



(a) No education (b) Tertiary education

Figure A.7: Long-term trends in educational attainment across Canadian provinces

Note: These figures plot the shares of low- and high-educated in each Canadian province from 2000 to 2022. The data source is Statistics Canada Table 37-10-0130-01 from the Canadian Labour Force Surveys.

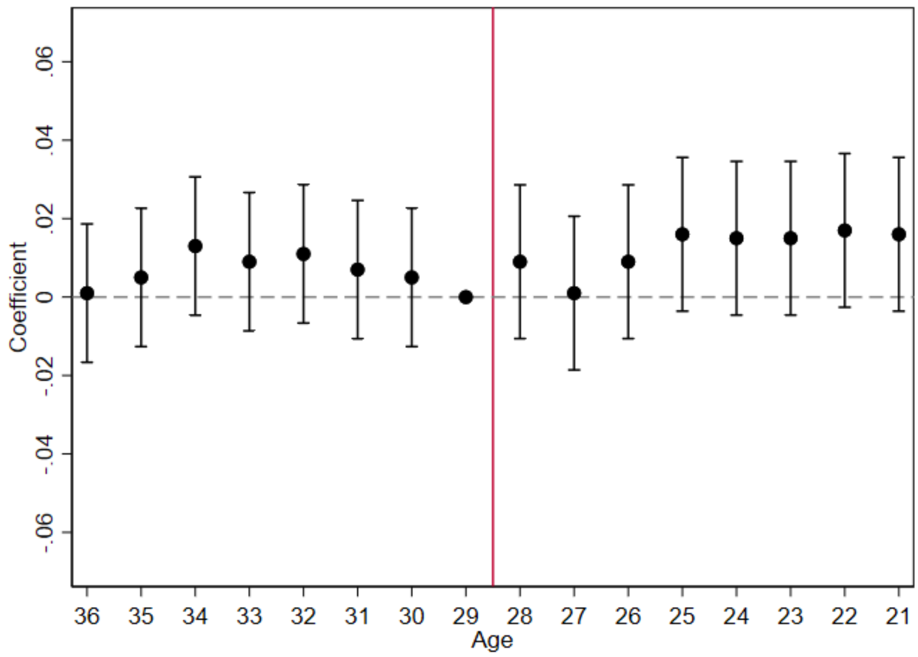
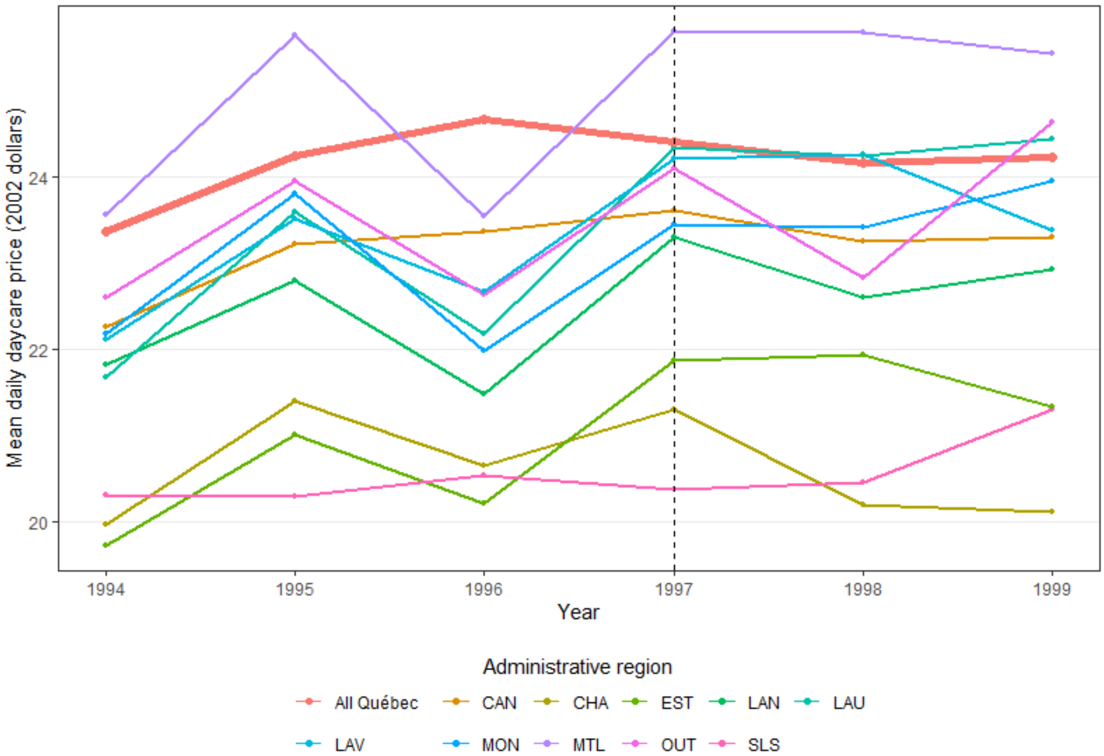


Figure A.8: Long-run policy impact on children's educational attainment in low-expansion regions

Note: This figure plots the regression coefficients of event-study regressions using the 2016 and 2021 Canadian Census of population. The dependent variable is a dummy variable taking the value of one if the individual has completed university studies. The horizontal axis represents the individual's age. Robust standard errors in parentheses.

Figure A.9: Evolution of average daycare prices in unregulated network by administrative region



Data sources: Ministry of the Family

Notes: This figures plots the evolution of average daily daycare prices in constant 2002 dollars in selected administrative regions in Québec. The thickest line is the average in the entire Québec province. CAN = Capitale-Nationale; CHA = Chaudière-Appalaches; EST = Estrie; LAN = Lanaudière; LAU = Laurentides; LAV = Laval; MON = Montérégie; MTL = Montréal; OUT = Outaouais; SLS = Saguenay-Lac-Saint-Jean

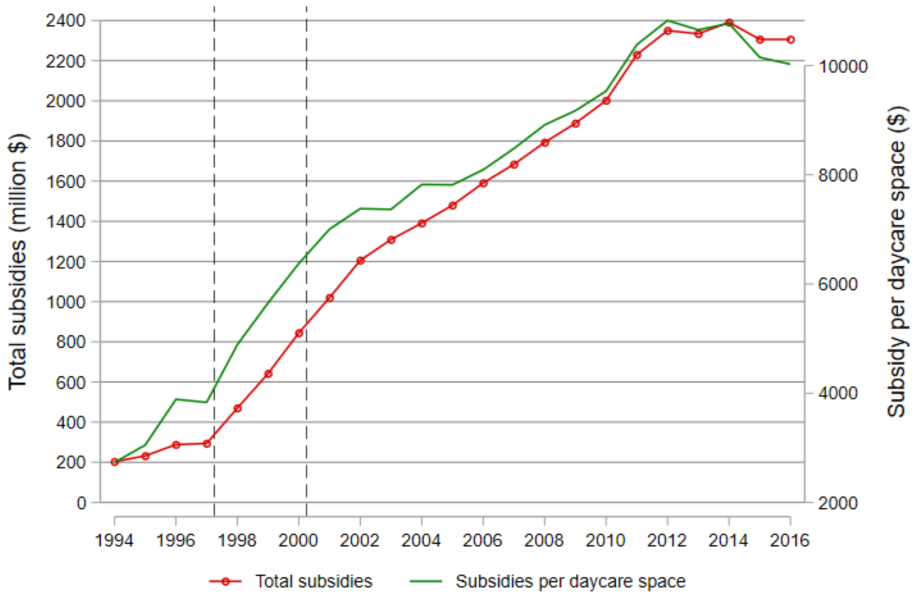


Figure A.10: Evolution of daycare subsidies and subsidies per space in Québec

Note: This Figure displays the total subsidies to daycare facilities and families (red dotted line, left axis) and the subsidy per daycare space (green line, right axis). It is constructed using data from a series of budgetary reports of the Québec Treasury Board for the subsidy amounts and of the Ministry of the Family for daycare spaces. Additional costs of the program include additional administrative costs to operate the program, infrastructure subsidies to CPEs, and retirement pensions payments of daycare providers in CPEs. Those are nevertheless small in magnitude compared to direct subsidies.

A.1.5 Appendix Tables

Table A.2: Descriptive statistics

	Québec		Rest of Canada	
	Pre	Post	Pre	Post
Panel A: household characteristics				
Age of mother	30.893	31.167	31.656	32.226
	(4.86)	(5.38)	(4.994)	(5.27)
Age of father	33.508	33.965	34.065	34.789
	(5.368)	(5.822)	(5.584)	(5.858)
Age of child	2.023	2.019	1.998	2.018
	(1.421)	(1.411)	(1.416)	(1.413)
Number of younger siblings	0.236	0.218	0.235	0.220
	(0.474)	(0.442)	(0.463)	(0.451)
Mother is immigrant	0.088	0.125	0.218	0.245
	(0.284)	(0.331)	(0.413)	(0.43)
Father is immigrant	0.096	0.126	0.209	0.239
	(0.295)	(0.332)	(0.406)	(0.427)
Number of older siblings	0.780	0.792	0.904	0.836
	(0.924)	(0.95)	(1.035)	(0.954)
Child is a girl	0.494	0.479	0.488	0.489
	(0.5)	(0.5)	(0.5)	(0.5)
Mother college degree	0.202	0.270	0.204	0.273
	(0.402)	(0.444)	(0.403)	(0.446)
Father college degree	0.195	0.239	0.215	0.262
	(0.396)	(0.427)	(0.411)	(0.44)
Mother no education	0.133	0.122	0.108	0.093
	(0.34)	(0.328)	(0.311)	(0.291)
Father no education	0.167	0.157	0.138	0.111
	(0.373)	(0.364)	(0.345)	(0.314)
Household lives in rural area	0.153	0.150	0.154	0.105
	(0.36)	(0.357)	(0.361)	(0.307)
Panel B: selected outcomes				
Child in care	0.418	0.630	0.407	0.482
	(0.493)	(0.483)	(0.491)	(0.5)
Hours in care	13.071	17.425	11.571	10.606
	(17.983)	(19.237)	(16.983)	(16.361)
Mother works	0.532	0.644	0.590	0.636
	(0.499)	(0.479)	(0.492)	(0.481)
Mother hours worked	17.541	21.575	17.929	20.250
	(18.176)	(17.982)	(17.84)	(18.448)
Father works	0.867	0.921	0.909	0.954
	(0.339)	(0.27)	(0.288)	(0.21)
Father hours worked	36.374	39.628	39.483	42.264
	(16.264)	(13.854)	(15.119)	(12.766)
Rarely/never reads	0.226	0.086	0.107	0.024
	(0.418)	(0.281)	(0.31)	(0.153)
Reads weekly	0.369	0.395	0.235	0.219
	(0.483)	(0.489)	(0.424)	(0.413)
Reads daily	0.379	0.462	0.645	0.740
	(0.485)	(0.499)	(0.479)	(0.439)
Child PPVT score	98.408	100.462	100.301	102.191
	(16.097)	(15.128)	(14.945)	(15.124)
Child MSD score	99.300	98.418	100.462	101.106
	(15.028)	(14.674)	(15.254)	(14.344)

Note: Pre-reform data is the first two waves (1994-1995 and 1996-1997) of the National Longitudinal Survey of Children and Youth (NLSCY). Post-reform data are waves 4 and 5 of the NLSCY (2000-2001 and 2002-2003). The sample is restricted to two-parent families with a preschool-age child. Standard deviations are reported in parentheses.

Table A.3: Pre-reform descriptive statistics by childcare expansion status

	Low exp.	High exp.
Panel A: Household characteristics		
Age of mother	30.625	31.104
	(4.55)	(5.062)
Age of father	33.889	33.261
	(5.815)	(5.022)
Age of child	1.981	2.060
	(1.42)	(1.424)
Number of younger siblings	0.242	0.233
	(0.484)	(0.467)
Number of older siblings	0.794	0.772
	(0.946)	(0.902)
Child is a girl	0.512	0.483
	(0.5)	(0.5)
Mother is immigrant	0.151	0.044
	(0.358)	(0.206)
Father is immigrant	0.162	0.050
	(0.369)	(0.219)
Mother college degree	0.192	0.212
	(0.394)	(0.409)
Father college degree	0.210	0.186
	(0.407)	(0.39)
Mother no education	0.174	0.104
	(0.38)	(0.305)
Father no education	0.192	0.149
	(0.394)	(0.356)
Household lives in rural area	0.131	0.165
	(0.337)	(0.371)
Panel B: Selected outcomes		
Child in care	0.417	0.422
	(0.493)	(0.494)
Hours in care	13.501	12.852
	(18.468)	(17.666)
Mother works	0.520	0.538
	(0.5)	(0.499)
Mother hours worked	17.122	17.798
	(18.355)	(18.053)
Father works	0.867	0.867
	(0.339)	(0.339)
Father hours worked	36.444	36.321
	(16.401)	(16.145)
Rarely/never reads	0.215	0.232
	(0.411)	(0.422)
Reads weekly	0.388	0.356
	(0.487)	(0.479)
Reads daily	0.373	0.384
	(0.484)	(0.487)
Child PPVT score	96.014	100.120
	(16.887)	(15.338)
Child MSD score	98.412	99.944
	(15.344)	(14.839)

Note: Data: first two waves (1994-1995 and 1996-1997) of the National Longitudinal Survey of Children and Youth (NLSCY). Low-expansion regions are administrative regions (within Québec) in the bottom tercile of the childcare expansion distribution. The sample is restricted to two-parent families with a preschool-age child. Standard deviations are reported in parentheses.

Table A.4: Determinants of local childcare expansions

	(1)	(2)
Initial coverage rate	-0.6756**	-0.2208
	(0.24)	(19.54)
Number of inhabitants (in 1,000s)		-0.0003
		(0.02)
Number of preschoolers (in 1,000s)		0.0055
		(0.39)
Share of medium-educated (high school)		0.0008
		(0.35)
Share of high-educated		0.0019
		(0.28)
Unemployment rate		1.0996
		(74.08)
Female unemployment rate		-1.0255
		(111.29)
Mean wages (in 1,000$)		0.0156
		(1.42)
Mean monthly rent ($)		-0.0005
		(0.03)
Constant	0.3903***	0.2479
	(0.04)	(2.12)
p-value of joint significance	0.0056	1.0000
R^2	0.352	0.515

Note: This table reports coefficients of linear regressions of the change in the childcare coverage rate (number of spaces divided by the population of preschool-age children) from 1997 to 2003 on the initial coverage rate and baseline regional characteristics. The data sources are a series of ministerial reports for daycare spaces (*Ministère de la Famille*, which held various other names) as well as the Québec Statistical Institute and the 1996 Canadian Census for the children population and the regional characteristics. The second-to-last row reports the p-value of the hypothesis that all the coefficients on baseline regional characteristics are jointly zero. Bootstrapped standard errors (1,000 replications) in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Heterogeneous impacts of the Québec childcare reform on fathers' employment and institutional care use by childcare expansion status

	Institutional care		Inst. care hours		Father works		Father's work hours	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : Eligible$_{pt}$	0.131***	0.168***	5.250***	4.926***	0.005	0.043***	0.159	0.614***
	(0.011)	(0.004)	(1.387)	(0.692)	(0.007)	(0.011)	(0.396)	(0.166)
β_2 : Eligible$_{pt}$		-0.073*		-2.448***		-0.048*		-1.559*
x LowExp$_r$		(0.038)		(0.497)		(0.021)		(0.798)
Region (r) FE		✓		✓		✓		✓
r-level controls		✓		✓		✓		✓
R^2	33575	33575	33320	33320	34012	34012	31497	31497
N	0.069	0.074	0.076	0.08	0.161	0.162	0.09	0.093

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), and immigration status (both parents). Odd columns report estimates of equation (2.1) while even columns are regression results of equation (2.2). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.6: Heterogeneous impacts of the Québec childcare reform on mothers' employment and childcare use by childcare expansion status, comparison with Ontario only

	Mother works		Mother's work hours		Child in care		Childcare hours	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : Eligible$_{pt}$	0.076***	0.150***	1.850***	4.065***	0.147***	0.190***	5.481***	6.577***
	(0.002)	(0.015)	(0.138)	(0.656)	(0.033)	(0.035)	(1.547)	(0.946)
β_2 : Eligible$_{pt}$ x LowExp$_r$		-0.060***	-1.738**	-0.046**	-1.972			
		(0.007)	(0.602)	(0.016)	(1.671)			
Region (r) FE		✓		✓		✓		✓
r-level controls		✓		✓		✓		✓
p-value of $\beta_1 + \beta_2 = 0$		0.000		0.000		0.001		0.012
R^2	0.116	0.119	0.105	0.110	0.127	0.130	0.114	0.119
N	15739	15739	15725	15725	15735	15735	14426	14426

Note: The data source is waves 1-2-4-5 of the NLSCY. Control variables are parents' age (in bins), age of the child, number and ages of siblings (in bins), population of the area of residence (in bins), education (both parents), and immigration status (both parents). Odd columns report estimates of equation (2.1) while even columns are regression results of equation (2.2). The sample is restricted to two-parent families with a preschool-age child and with non-missing covariates. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Child skill production technology parameters in different Canadian regions

Parameter	Description	Québec	Ontario	Maritimes	West
		(1)	(2)	(3)	(4)
δ_0	(self-productivity)	0.178***	0.155***	0.136***	0.165***
		(0.015)	(0.012)	(0.013)	(0.011)
δ_e	(maternal care)	0.023***	0.020***	0.019***	0.010***
		(0.004)	(0.004)	(0.004)	(0.004)
δ_d	(non-maternal care)	0.032***	0.012***	0.002	-0.002
		(0.005)	(0.004)	(0.005)	(0.005)
p-value of $\delta_e^{QC} - \delta_e^{p} = 0$			0.765	0.624	0.171
N		3860	5994	4879	7174

Note: This Table reports estimation results for the child human capital production function (equation 2.11) in different Canadian regions. Bootstrapped standard errors in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Production function and exhaustion effect parameters

Panel A: Production function parameters

Parameter	Description	OLS (1)	IV (2)
δ_0	(self-productivity)	0.205***	0.105***
		(0.027)	(0.018)
δ_ρ	(maternal care)	0.025***	0.226***
		(0.007)	(0.02)
δ_d	(non-maternal care)	0.034***	0.264***
		(0.009)	(0.018)

Panel B: Exhaustion-effect estimation

$\gamma_{\rho,2}$ estimate		IV First-stage	
Naive OLS (3)	IV (4)	ITT($\ln(\rho)$) (5)	ITT($\ln(T_m)$) (6)
1.015	1.885***	0.0796***	-0.0899***
(0.022)	(0.205)	(0.028)	(0.022)

Note: This Table reports estimation results for the child human capital production function (equation 2.11) and the exhaustion-effect parameter (equation 2.15) respectively. Standard errors clustered at the province-year level in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.9: Preference parameters

Parameter	Description	Estimate	SE
γ_C	(consumption)	1	(.)
γ_ℓ	(leisure)	0.312***	(0.018)
γ_{T_m}	(maternal care)	1.872***	(0.178)
γ_{h_1}	(child skills)	16.262***	(2.269)
$\gamma_{\rho,1}$	(cost of effort)†	0.227***	(0.0361)
$\gamma_{d,1}$	(childcare use)	1.643***	(0.156)
$\gamma_{d,2}$	(coverage)	1.615*	(0.879)
N	2518		

Note: Bootstrapped standard errors (400 replications) in parentheses. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

† $\gamma_{\rho,1}$ is re-scaled (multiplied) by 10,000 for comparability.

Table A.10: Model in-sample fit

	Observed outcome		Simulated outcome
	Mean	SD	Mean
Extensive margin			
Maternal employment	0.532	0.499	0.557
Childcare use	0.418	0.493	0.482
Reading daily to child	0.379	0.485	0.333
Intensive margin			
Maternal employment	17.54	18.18	14.28
Childcare use	13.07	17.98	14.43
Reading hours	3.67	2.81	3.17

Table A.11: Benchmark welfare estimate including costs of juvenile crime

MVPF components	Mean values	External sources used
Direct cost	\$2,617M	Québec Treasury Board
Fiscal externality		
Tax returns and reduced transfers	\$971M	CTaCS
Youth crime (long-run)	-\$19.36M	Baker et al. (2019) and Cohen (2020)
Willingness-to-pay		
Mothers of preschoolers	\$2,213M	CTaCS
Mothers of older children	\$1,102M	Lefebvre et al. (2009)
Taxes and reduced transfers	-\$971M	CTaCS
Youth crime (long-run)	-\$20.16M	Baker et al. (2019) and Cohen (2020)
MVPF	1.40	

Notes: This table outlines the components of the MVPF under the benchmark estimator including additional societal costs of increased youth criminal activity. The last column reports the external sources used for the policy's cost and other sources of fiscal externalities. The acronym CTaCS refers to the Canadian Tax and Credit Simulator of Milligan (2019).

A.2 Mathematical Appendix

A.2.1 Proof of equation (2.19)

Substituting the time constraints into the budget, we can rewrite (2.17) as:

$$C(\theta) + (1 - \tau_d(w(T - T_m(\theta) - \ell(\theta)) + I(\theta))\, p(T_c - T_m(\theta)) = w(T - T_m(\theta) - \ell(\theta)) + I(\theta)$$

Isolating non-labor income yields:

$$\begin{aligned} I(\theta) = & \underbrace{(C(\theta) + (1 - \tau_d(w(T - T_m(\theta) - \ell(\theta)))\, p(T_c - T_m(\theta)) - w(T - T_m(\theta) - \ell(\theta)))}_{\equiv A(\theta)} \\ & \cdot (1 + \tau_d(T_c - T_m(\theta)))^{-1} \end{aligned}$$

We are thus left with this constraint as well as the modified child skills production technology $\tilde{H}(T_m(\theta), e(\theta); h_0)$. The Lagrangian of the problem thus becomes:

$$\begin{aligned} \mathscr{L} = & \, U(C(\theta), \ell(\theta), e(\theta), T_m(\theta); \theta) \\ & - \lambda \left[A(\theta)\, (1 + \tau_d(T_c - T_m(\theta)))^{-1} - I(\theta)\right] \\ & - \mu \left[h_1(\theta) - \tilde{H}(T_m(\theta), e(\theta); h_0)\right] \end{aligned}$$

where we removed h_1 from the utility function because it is not a choice variable. The first-order conditions for optimality at interior solutions are thus given by:

FOCs: $\dfrac{\partial U(C^*(\theta), \cdot)}{\partial C} = \lambda\, (1 + \tau_d(T_c - T_m(\theta)))^{-1}$ $\qquad \dfrac{\partial U(e^*(\theta), \cdot)}{\partial e} = \mu \dfrac{\partial \tilde{H}(\theta)}{\partial e}$

$$\frac{\partial U(\ell^*(\theta), \cdot)}{\partial \ell} = \lambda\, (1 + \tau_d(T_c - T_m(\theta)))^{-1} \frac{\partial A(\theta)}{\partial \ell}$$

$$\frac{\partial U(T_m^*(\theta), \cdot)}{\partial T_m} = \lambda \left[\frac{\partial A(\theta)}{\partial T_m} \cdot (1 + \tau_d(T_c - T_m(\theta)))^{-1} + A(\theta) \cdot \tau_d\, (1 + \tau_d(T_c - T_m(\theta)))^{-2}\right] + \mu \frac{\partial \tilde{H}(\theta)}{\partial T_m}$$

For an infinitesimal policy change, the difference in indirect utilities (the numerator in (2.18)) is simply the total derivative of $V(\theta_0)$ with respect to θ, which is given by:

$$\begin{aligned} \frac{dV(\theta_0)}{d\theta} = & \, \gamma_{d,2} + \frac{\partial U(C^*(\theta_0), \cdot)}{\partial C} \cdot \frac{\partial C^*(\theta_0)}{\partial \theta} + \frac{\partial U(\ell^*(\theta_0), \cdot)}{\partial \ell} \cdot \frac{\partial \ell^*(\theta_0)}{\partial \theta} + \frac{\partial U(T_m^*(\theta_0), \cdot)}{\partial T_m} \cdot \frac{\partial Tm^*(\theta_0)}{\partial \theta} \\ & + \frac{\partial U(e^*(\theta_0), \cdot)}{\partial e} \cdot \frac{\partial e^*(\theta_0)}{\partial \theta} \end{aligned}$$

where $\gamma_{d,2}$ is the utility gain stemming from the change in coverage. Using the first-order conditions derived above, we have:

$$
\begin{aligned}
\frac{dV(\theta_0)}{d\theta} = \lambda \Bigg(& (1+\tau_d(T_c - T_m(\theta_0)))^{-1} \frac{\partial C^*(\theta_0)}{\partial \theta} + \ (1+\tau_d(T_c - T_m(\theta_0)))^{-1} \frac{\partial A(\theta_0)}{\partial \ell} \frac{\partial \ell^*(\theta_0)}{\partial \theta} \Bigg) \\
& + \lambda \left[\frac{\partial A(\theta)}{\partial T_m} \cdot (1+\tau_d(T_c - T_m(\theta)))^{-1} + A(\theta) \cdot \tau_d \left(1+\tau_d(T_c - T_m(\theta))\right)^{-2} \right] \\
& + \ \mu \left(\frac{\partial \tilde{H}(\theta_0)}{\partial T_m} \frac{\partial Tm^*(\theta_0)}{\partial \theta} + \frac{\partial \tilde{H}(\theta_0)}{\partial e} \frac{\partial e^*(\theta_0)}{\partial \theta} \right) + \gamma_{d,2}
\end{aligned}
$$

In the expression above, the first two lines are equal to the product of the marginal utility of income (λ) and the treatment effect on earnings ($I^*(\theta)$). Therefore, using the budget constraint, we can replace those lines by $\lambda I^*(\theta)$, which yields the result. □

A.2.2 Willingness to pay for a policy change

In this Appendix, we generalize the results highlighting the biases in the estimation of the willingness to pay for a large reform discussed in section 2.5.2. We use a Hendren (2016) framework slightly generalized so as to include non-pecuniary attributes. We first discuss the WTP of a single individual and then aggregation over all beneficiaries to move to social welfare.

Individual willingness-to-pay

Environment. Consider a decision maker $i \in \mathcal{I}$ facing the problem of choosing a vector of J market goods $\mathbf{x_i} = (x_i^1, ..., x_i^J)$, which can include labor-market activity, and K non-market choice variables $\mathbf{z_i} = (z_i^1, ..., z_i^K)$ to maximize a utility function $u_i(\mathbf{x_i}(\theta), \mathbf{z_i}(\theta))$ where θ is a set of policy choices of the government (the tax schedule, the level of public-good provision, the net transfers to the agent, and so on). The government's policy choice θ can potentially affect the agent's choices $\mathbf{x_i}$ and $\mathbf{z_i}$, her after-tax income y_i, and prices of goods. The agent faces a standard budget constraint as well as a set of additional constraints on the non-market choice variables. For simplicity, we suppose this set is a singleton so that the agent has to meet the budget constraint and one constraint on $\mathbf{z_i}$ (for example, a time allocation constraint). The decision problem thus writes:

$$
\max_{\mathbf{x_i}(\theta),\, \mathbf{z_i}(\theta)} \ u_i(\mathbf{x_i}(\theta), \mathbf{z_i}(\theta)) \quad s.t. \ \ p(\theta) \cdot \mathbf{x_i}(\theta) \leq y_i(\theta)
$$
$$
g(\mathbf{z_i}(\theta)) = 0
$$

where $p = (p_1, ..., p_J)$ is a price vector and g is differentiable in each of its arguments. Let $V_i(\theta) = U(\mathbf{x_i}^*(\theta), \mathbf{z_i}^*(\theta))$ be the agent's indirect utility under policy state θ.

The government now implements a policy change. The reform moves the policy state θ from the status-quo policy θ_0 to some new policy state θ_1. The agent's willingness-to-pay (WTP) for this policy change can be measured by the standard equivalent variation (EV), which we denote $\Delta y_i(\theta_0)$. That is, the WTP is the variation in income under θ_0 that would make the agent indifferent between the status quo and the new policy state:

$$\Delta y_i(\theta_0) = \frac{V_i(\theta_1) - V_i(\theta_0)}{\lambda_i} \tag{A.1}$$

where λ_i is the agent's marginal utility of income.

WTP for small policy changes. Let us consider first, as is the case with the sufficient-statistics approach, that the policy change is infinitesimal. For an infinitesimal (marginal) policy change (in θ), the numerator in (A.1), the difference in indirect utilities, is the total derivative of $V_i(\theta_0)$ with respect to θ. Under the additional standard assumption that prices of goods remain unchanged at the margin (i.e. assuming competitive markets for $\mathbf{x}$), we get:

$$\frac{dV_i(\theta_0)}{d\theta} = \lambda_i \, \frac{dy_i^*(\theta_0)}{d\theta} + \mu_i \; g'(\mathbf{z_i}^*(\theta_0)) \cdot \frac{d\mathbf{z_i}^*(\theta_0)}{d\theta} \tag{A.2}$$

where μ_i is the Lagrange multiplier on the second constraint.
Proof. The Lagrangian of the problem writes:

$$\mathscr{L} = U(\mathbf{x_i}(\theta), \mathbf{z_i}(\theta)) - \lambda_i \left[p(\theta)\mathbf{x_i}(\theta) - y_i(\theta)\right] - \mu_i g(\mathbf{z_i}(\theta)) \tag{A.3}$$

and thus the solution satisfies the first-order conditions:

$$\text{FOCs:} \quad \frac{\partial U_i(\mathbf{x_i}^*(\theta), \mathbf{z_i}(\theta))}{\partial \mathbf{x_i}} = \lambda_i \; p(\theta) \qquad \frac{\partial U_i(\mathbf{x_i}(\theta), \mathbf{z_i}^*(\theta))}{\partial \mathbf{z_i}} = \mu_i \; g'(\mathbf{z_i}^*(\theta)) \tag{A.4}$$

where λ_i is the agent's marginal utility of income and μ_i is the Lagrange multiplier on the second constraint. Let $V_i(\theta) = U_i(\mathbf{x_i}^*(\theta), \mathbf{z_i}^*(\theta))$ be the agent's indirect utility under policy state θ.

For an infinitesimal policy change, the difference in indirect utilities (the numerator in (A.1)) is simply the total derivative of $V_i(\theta_0)$ with respect to θ, which is given by:

$$\frac{dV_i(\theta_0)}{d\theta} = \frac{\partial U_i(\mathbf{x_i}^*(\theta_0), \mathbf{z_i}(\theta_0))}{\partial \mathbf{x_i}} \cdot \frac{\partial \mathbf{x_i}^*(\theta_0)}{\partial \theta} + \frac{\partial U_i(\mathbf{x_i}(\theta_0), \mathbf{z_i}^*(\theta_0))}{\partial \mathbf{z_i}} \cdot \frac{\partial \mathbf{z_i}^*(\theta_0)}{\partial \theta} \tag{A.5}$$

Using the first-order conditions (A.4), we have:

$$\frac{dV_i(\theta_0)}{d\theta} = \lambda_i \; p(\theta_0) \; \frac{\partial \mathbf{x_i}^*(\theta_0)}{\partial \theta} + \mu_i \; g'(\mathbf{z_i}^*(\theta)) \frac{\partial \mathbf{z_i}^*(\theta_0)}{\partial \theta} \tag{A.6}$$

Taking the derivative of the budget constraint with respect to θ yields:

$$\frac{\partial y_i(\theta)}{\partial \theta} = \frac{\partial p(\theta)}{\partial \theta} \; \mathbf{x_i}(\theta) + p(\theta) \; \frac{\partial \mathbf{x_i}(\theta)}{\partial \theta}$$

Therefore, assuming that prices are not impacted by the policy change at the margin (for example, assuming competitive markets for $\mathbf{x}$), the first term on the right-hand-side is null and we obtain that the impact of the policy on earnings is simply given by:

$$\frac{\partial y_i(\theta)}{\partial \theta} = p(\theta)\frac{\partial \mathbf{x_i}(\theta)}{\partial \theta} \tag{A.7}$$

that is, the additional spendings induced by the policy. Substituting (A.7) into (A.6) yields the result:

$$\frac{dV_i(\theta_0)}{d\theta} = \lambda_i \, \frac{dy_i^*(\theta_0)}{d\theta} + \mu_i \; g'(\mathbf{z_i}^*(\theta_0)) \cdot \frac{d\mathbf{z_i}^*(\theta_0)}{d\theta}$$

□

Therefore, the numerator of the WTP is the sum of two terms, the *pecuniary* benefits and the *non-pecuniary* gains stemming from the relaxation of the second constraint (e.g. the increase in available time). The intuition behind this result is the logic of the Envelope theorem, which implies that, at the margin, behavioral responses do not have a direct effect on utility (i.e. $\partial V_i(\theta)/\partial \mathbf{x_i} = \partial V_i(\theta)/\partial \mathbf{z_i} = 0$). Thus, if one additionally assumes that non-pecuniary gains are negligible, which may be reasonable for marginal reforms, the difference in utilities boils down to the policy's impact on the beneficiaries' budget constraint. The WTP is then simply given by the causal effect of the policy on earnings ($\frac{dy_i(\theta)}{d\theta}$). This result is powerful because it implies that the treatment effect on beneficiaries' earnings is a sufficient statistic for the numerator of the MVPF (Hendren, 2016; Hendren and Sprung-Keyser, 2020). The MVPF framework thus leverages the recent "credibility revolution" in the estimation of causal effects (Angrist and Pischke, 2010) to make transparent welfare statements.

Large-policy bias. Consider now a discrete (large or non-infinitesimal) policy change. In this case, such as with the Québec childcare reform, the previous result does not hold anymore since Envelope conditions only apply to marginal reforms. In particular, behavioral responses, both for market and non-market choices, now have direct impacts on utility because the agent re-optimizes behavior. Moreover, for large reforms, non-pecuniary gains may be important. Thus, the treatment effect on earnings of beneficiaries is a biased estimate of the WTP.[3] For policies with large direct costs, as is the case of the

[3]To see this, suppose the utility function u is additively separable in $\mathbf{x_i}$ and $\mathbf{z_i}$ such that $u_i(\mathbf{x_i}, \mathbf{z_i}) = u_i^1(\mathbf{x_i}(\theta)) + u_i^2(\mathbf{z_i}(\theta))$. The difference in indirect utilities is given by:

$$V_i(\theta_1) - V_i(\theta_0) = u_i^1(\mathbf{x_i}^*(\theta_1)) - u_i^1(\mathbf{x_i}^*(\theta_0)) + u_i^2(\mathbf{z_i}^*(\theta_1)) - u_i^2(\mathbf{z_i}^*(\theta_0))$$

The first (resp. second) difference captures the overall impact of the policy on utility via behavioral changes in $\mathbf{x_i}$ (resp. $\mathbf{z_i}$). For non-marginal policies, differences in u_i^1 and u_i^2 are no longer the partial derivatives of $V_i(\theta_0)$. The *large-policy bias* is given by: $u_i^1(\mathbf{x_i}^*(\theta_1)) - u_i^1(\mathbf{x_i}^*(\theta_0)) - \lambda_i \, \frac{dy_i^*(\theta_0)}{d\theta}$. The bias stemming from the omission of non-pecuniary gains is simply $u_i^2(\mathbf{z_i}^*(\theta_1)) - u_i^2(\mathbf{z_i}^*(\theta_0))$ since these are

Québec reform, underestimating the WTP might seriously affect the welfare conclusions.

The first bias, which we label the *large-policy bias* (equal to the policy's impact on utility through $\mathbf{x}$), stems from re-optimization behavior of beneficiaries. Since agents make non-marginal changes in market choices, these no longer have a null direct impact on the difference in utilities ($V(\theta_1) - V(\theta_0)$). This bias is likely to be large in our context, given that the literature has documented major impacts of the reform on economic behavior. The large changes in maternal labor supply and child care use have direct impacts on utility through changes in mothers' time allocation, which are not captured by the treatment effect on earnings.

Non-pecuniary gains. Using the treatment effect on beneficiaries' earnings as an estimator of the WTP is subject to a second bias (equal to the policy's impact on utility through $\mathbf{z}$), namely the omission of non-pecuniary benefits of the policy. This bias, in fact, also applies to small reforms and Hendren and Sprung-Keyser (2020) themselves acknowledge that it may be important in some cases.[4] We argue that non-pecuniary gains (or losses) are likely to be large in the case of childcare policies (even for small-scale programs) since they may have substantial impacts on (especially mothers') parenting time and practices. Moreover, early childhood programs have substantial impacts on child development, which is valued by parents.

Social welfare

We now consider aggregation of individual beneficiaries' willingness-to-pay to obtain an estimate of the society's willingness-to-pay. Assuming there exists of a set of Pareto weights ψ_i, for each beneficiary i, social welfare at a given policy state θ is given by:

$$W(\theta) \ = \ \sum_{i \in \mathcal{I}} \psi_i V_i(\theta) \tag{A.8}$$

where W is the social welfare function and V_i is the indirect utility function of beneficiary i. This formulation is very general and can accommodate any social welfare function. It allows, for instance, social preferences for redistribution from richer to poorer individuals (Hendren, 2016).[5]

The society's WTP, which we denote by SWTP, for a reform is then given by (the monetary value of) the difference in social welfare between the the new (θ_1) and the

ignored by assumption.

[4]For example, in their estimation of the MVPF of admission to Florida International University, changes in effort at school or other forms of utility gains during college time are assumed away (Hendren and Sprung-Keyser, 2020, p. 1230). We discuss many other reforms for which non-pecuniary gains might be important in our survey of MVPF estimates in Appendix A.4.

[5]As additional examples, a utilitarian planner sets $\psi_i = 1 \ \forall i$ while a Rawlsian planner sets $\psi_{i'} > 0$ for i' such that $V_{i'} \leq V_i \ \forall i$ and $\psi_i = 0 \ \forall i \neq i'$.

status-quo (θ_0) policy states. Using the equivalent variation $\Delta y_i(\theta_0)$ as a measure of beneficiary i's WTP in dollars, the society's WTP is thus given by:

$$\text{SWTP} \; = \sum_{i \in \mathcal{I}} \psi_i \, \frac{V_i(\theta_1) - V_i(\theta_0)}{\lambda_i} \tag{A.9}$$

where we used equation (A.1). As Hendren and Sprung-Keyser (2020) note, the ratio ψ_i/λ_i is the marginal social utility of individual i's income.

Sufficient-statistics approach As argued in the previous section, if the reform and non-pecuniary gains are sufficiently small, the equivalent variation for a beneficiary boils down to the treatment effect on her earnings. As a naive estimator of the society's willingness-to-pay for a large reform, one can use this powerful result, as if the policy change were infinitesimal. Therefore, using equation (A.2), this estimator can be written as:

$$\text{SWTP}_{naive} = \sum_{i \in \mathcal{I}} \psi_i \, (y_i^*(\theta_1) - y_i^*(\theta_0)).$$

Thus, if the social welfare criterion is utilitarian, this estimator is simply the sum of (weighted) pecuniary gains of all beneficiaries. In other words, the naive estimator of the SWTP is the treatment effect on (weighted) aggregate earnings.

Structural approach For discrete (large) policy changes, the Envelope theorem does not apply. Since beneficiaries do not simply react at the margin, behavioral responses have first-order impacts on utility. Therefore, one has to estimate the utility gains for each individual, which may include gains (losses) from re-optimization and non-pecuniary benefits. These can be estimated using a well-specified structural model of behavior. To obtain the equivalent variation, one can use these estimates to find, for each beneficiary, the amount of additional income that would make her indifferent between the extra cash and the implemented policy. This alternative estimator takes the form:

$$\text{SWTP}_{model} \; = \sum_{i \in \mathcal{I}} \psi_i \, \frac{\tilde{V}_i(\theta_1) - \tilde{V}_i(\theta_0)}{\tilde{\lambda}_i} \tag{A.10}$$

where $\tilde{w}$ denotes that variable w is estimated from the model.

A.3 Comparison of MVPF estimates with other criteria

There are recent debates in Economics on the use of the MVPF as a welfare criterion to evaluate social programs. In particular, García and Heckman (2022a,b) criticize the use of this metric arguing that the MVPF approach *(i)* abstracts from the deadweight loss of taxation and thus from the social marginal value of public expenditure, *(ii)* assumes a fixed government budget and is silent about policies that loosen the government's budget constraint, *(iii)* uses an arbitrary ratio, *(iv)* cannot rank all policies, and *(v)* interprets some welfare-improving policies as "money pumps". They suggest the use of an alternative criterion, namely the net social benefit (NSB). The NSB is simply the difference between the policy's benefits and the cost to society. García and Heckman (2022b) define the NSB as:

$$\text{NSB} = \text{Benefits} - \Omega(\text{Direct cost}) \tag{A.11}$$

where $\Omega()$ is a potentially non-linear function, which notably captures the deadweight loss of public expenditure. In practice, however, a linearity assumption on Ω is often made in the literature and we follow this approach in our comparative exercise below. We compare our MVPF estimates with the NSB and the standard cost-benefit ratio in Table A.12.

Before discussing the results we note, however, that in a reply, Hendren and Sprung-Keyser (2022) show that those critics originate from a misconception about their welfare criterion and that, in several contexts, the MVPF may be preferable to the NSB. For example, one key advantage of the MVPF framework is that it does not assume how the government finances the policy while the standard deadweight loss of taxation assumes an arbitrary linear income tax rate. Hendren and Sprung-Keyser (2020)'s criterion evaluates welfare impacts of budget-neutral programs by comparing two MVPFs: the one of an expenditure policy to the one of a revenue-raising policy. On arguments *(iv)* and *(v)*, the MVPF approach identifies policies that pay for themselves and for which recipients have a positive willingness-to-pay as Pareto improvements (defined as an infinite MVPF, not as a negative one as point *(v)* states). It is thus true that one cannot rank among Pareto improvements, but the message here is that the government should implement all those policies (at no cost) so ranking them is obsolete. Last, a fair criticism of empirical welfare analysis in García and Heckman (2022b) is that, in reality, the welfare costs of raising public revenue are likely non-linear. Such non-linearities in the deadweight loss of public expenditure cannot be accounted for in the MVPF framework, but the critique also applies to other standard criteria for evaluating social programs considered by García and Heckman (2022b). Estimating non-linear welfare costs of raising public revenue is a promising avenue for future research, but is beyond the scope of this paper. We refer the reader to those papers for a more extensive discussion.

Our comparative exercise in Table A.12 reveals that, for the benchmark estimator, the choice of criterion substantially affects social-welfare conclusions. When focusing on earnings gains only, we find that the cost-benefit ratio (CBR) is lower than one (0.97) and

Table A.12: Comparison between MVPF and alternative social welfare criteria

Criterion	Formula	Value
Benchmark estimator		
MVPF	Benefits / Net cost	1.42
NSB	Benefits - $(1+\phi)$ Cost	-$87.1M
CBR	Benefits / $[(1+\phi)$ Cost$]$	0.97
Structural estimator		
MVPF	Benefits / Net cost	3.56
NSB	Benefits - $(1+\phi)$ Cost	$2675.9M
CBR	Benefits / $[(1+\phi)$ Cost$]$	1.79

Note: We assume a deadweight loss of public expenditure $\phi = 1/3$ in this example, as in Hendren and Sprung-Keyser (2022). MVPF = marginal value of public funds, NSB = net social benefit, CBR = cost-benefit ratio

that the NSB is negative (-$87.1M), suggesting that the policy is not socially desirable. However, when incorporating non-pecuniary gains, we find that all three criteria point at the same conclusion: benefits are larger than costs and the policy should be implemented under these criteria. Indeed, we obtain a cost-benefit ratio of 1.79 and a positive NSB of more than $2 billion under our structural estimator. Therefore, we fins that using these alternative criteria reinforce our conclusion that omitting non-pecuniary gains for mothers would lead to a substantial underestimation of social-welfare gains. In fact, under the CBR and NSB criteria, abstracting from mothers' non-monetary benefits would lead one to conclude that the policy should not be adopted. This reinforces our main result that non-pecuniary gains for mothers must be accounted for in welfare analysis of universal preschool policies.

A.4 Survey of MVPF estimates

Some authors compute social-welfare impacts of large policy changes using sufficient-statistics estimators (such as the MVPF) as if the policy were infinitesimal.[6] To assess the prevalence of this practice, we conduct our own survey of MVPF estimates appearing on the Policy Impacts Library (Hendren et al., 2023). Of course, this exercise requires

[6] We note that it is likely not in the authors' intention to make this assumption. They might simply apply the MVPF framework because it is a convenient tool to evaluate welfare impacts of policies using reduced-form causal estimates. For most papers, the welfare analysis is not the main goal and authors may only see this exercise as illustrative of the economic returns of the reform studied.

some judgement calls and we therefore only take the results of our survey as suggestive.

We first have to define a criterion indicating whether a policy change can be considered as infinitesimal. Importantly, envelope conditions allowing one to express the welfare effect of a policy change as a fiscal externality concern individuals' utility maximization problem. The "size" of the policy change should thus not be evaluated using the number of recipients or as a function of how local is the treatment effect estimated, but rather by the size of behavioral responses at the *individual* level.

In line with our discussion of the *large-policy bias* in section 2.5.2, we define a discrete (non-infinitesimal) policy as one that induces significant behavioral re-optimization by recipients. For infinitesimal changes, by the Envelope theorem, recipients do not re-optimize at the margin and only obtain utility gains from the relaxation of constraints in their maximization problem. Ignoring non-pecuniary gains of small reforms may also be reasonable. However, large behavioral responses to a policy such as entry into the labor market suggest agents face a different economic environment and revise their optimal choices, suggesting the policy change is not small. Omitting non-pecuniary gains (or losses) of abrupt changes in behavior may also lead to important biases in welfare estimates.

We have so far surveyed the first 24 papers appearing on the Policy Impacts Library webpage. For each paper, in addition to providing basic information on the reform considered, we assess whether the policy change being studied satisfies the infinitesimal-policy criterion defined above. We also list some potential non-pecuniary gains (or losses) omitted by the authors. Lastly, we briefly discuss whether the large-policy bias or the omission of non-pecuniary gains is likely to affect the authors' welfare conclusions. In particular, we check whether papers finding apparently welfare-improving (resp. welfare-decreasing) policies are omitting utility losses (resp. gains) which biases their estimates upwards (resp. downwards).[7] In our discussions, we focus on policy impacts on outcomes studied by the authors and limit mentions to other fiscal externalities.

The detailed results are reported in Table A.13. Two key findings stand out of this review. First, most papers who apply the MVPF framework do so in the context of a non-infinitesimal policy change. Out of the first 24 MVPF estimates appearing in the Policy Impacts Library, we find that at least 20 cases clearly do not satisfy our criteria for the policy studied be considered as infinitesimal. For example, in many cases, the policy is found to have large impacts on labor supply at the extensive margin. Such employment responses are not marginal. Therefore, unless all beneficiaries are indifferent

[7]One of the stated advantage of the MVPF framework is that it "can be easily compared across programs" (Kline and Walters, 2016, p. 1815). However, for such comparisons across policy domains, which require a precise ranking of policies, obtaining a robust MVPF point estimate is crucial. Therefore, even if the biases do not affect the authors' general welfare conclusions (i.e. whether the policy is welfare-improving or not), omitting potentially large utility gains (or losses) can be problematic for policy comparisons.

between working and staying out of the labor force, the large-policy bias discussed in section 2.5.2 applies.

Second, among those 20 cases, we argue that for at least seven, abstracting from the large nature of the policy change is likely to substantially affect the authors' welfare conclusions. For example, the MVPF of the introduction of old-age pensions in the United Kingdom is estimated to be 0.8 by Giesecke and Jäger (2021). This policy is not infinitesimal as made evident by the impacts on labor-supply and retirement decisions documented by the authors. For this reform, positive non-pecuniary gains for grandchildren such as reduced time taking care of the elderly as well as reduced financial stress related to retirement for beneficiaries are omitted. If these gains are sufficiently large, the MVPF could be higher than one, suggesting that the policy is welfare-improving rather than being a net cost to society.

Table A.13: Survey of MVPF estimates in Policy Impacts Library

Authors	**Causal estimates**	**Policy considered**	**Recipients**	**MVPF**	**Infinite-simal?**	**Agents re-optimize?**	**Non-pecuniary gains/losses**	**Welfare conclusions likely affected?**
Duchini and Van Effenterre (2022)	Duchini and Van Effenterre (2022)	Introduction of primary school on Wednesday morning in France	Parents (mothers) of primary-school age children in France	3.6	No	Large adjustments in week schedule	Parenting time, more regular work schedules, child cognition	No. MVPF is already high and reduced child care contraints should generate even more welfare gains.
Deshpande and Mueller-Smith (2022)	Deshpande and Mueller-Smith (2022)	Conditioning of access to the Social Security Income on re-evaluation of disability at 18 years old	Disadvantaged youth who would have received guaranteed SSI in adulthood	-16.1	No	Impacts on crime found to be concentrated in income-generating criminal offenses. This may suggest normally sub-optimal decisions to catch-up for the income loss. Consistent with this, Deshpande (2016) finds positive impacts of SSI removal on labor supply.	Psychic costs (or benefits) of criminal activity	No. MVPF is high (in abs. value) and reduced income in youth likely to imply additional negative distortions.
Hendren and Sprung-Keyser (2020)	Zimmerman (2014)	Admission to Florida International University	Prospective students at FIU	∞	No	Admission to university likely to induce significant changes in life trajectory, preferences, etc.	utility of attending school, effort costs, moving costs from other States	No. Even if benefits might be overestimated, the net costs are negative.

Authors	**Causal estimates**	**Policy considered**	**Recipients**	**MVPF**	**Infinite-simal?**	**Agents re-optimize?**	**Non-pecuniary gains/losses**	**Welfare conclusions likely affected?**
Paradisi (2021)	Brenøe et al. (2020)	Maternal leave in Denmark (effect of having female employees close to giving birth)	Coworkers and mother's firm	0.74	Arguably yes	Loss of only one employee for a short amount of time. Firms found to compensate by marginal adjustments, but no decrease in labor demand.	Maybe change in working atmosphere	No. MVPF below 1 and no omitted positive impact.
Hendren and Sprung-Keyser (2020)	Currie and Gruber (1996); Cutler and Gruber (1996); Dave et al. (2015); Miller and Wherry (2019)	Medicaid Expansions to Pregnant Women and Infants in US States	Pregnant women and children in the USA	∞	No	Large employment responses of women and probably behavioral adjustments in health habits.	Health insurance likely an important source of stress for families.	No. The policy already pays for itself and there should not be significant losses missing.
Hendren and Sprung-Keyser (2020)	Cohodes and Goodman (2014)	Massachusetts' Adams (MA) Scholarship	High-scoring high-school students	0.72	No	Change college choice and less likely to graduate	-	No. MVPF below 1 and no omitted positive impact.

Authors	Causal estimates	Policy considered	Recipients	MVPF	Infinite-simal?	Agents re-optimize?	Non-pecuniary gains/losses	Welfare conclusions likely affected?
Kuka and Shenhav (2020)	Kuka and Shenhav (2020)	EITC Benefits to Recent Mothers (1993)	Low-income workers (new non-married mothers)	5.6	No	Large effects on employment and wages	Improved mother's health and their children's outcomes	No. MVPF is already high.
Cabral and Dillender (2021)	Cabral and Dillender (2021)	Workers Compensation Benefit Generosity	Injured workers	0.46	No	Medical spending and benefit duration increase	Workers may value the increased medical spending and the consumption-smoothing benefits afforded by more generous coverage (acknowledged by the authors)	Yes. If these benefits are sufficiently high, MVPF could be higher.
Hendren and Sprung-Keyser (2020)	Cornwell et al. (2006)	Georgia HOPE Scholarship	High-school students graduating with at least a B average	4	No	Large effects on college enrollment	Utility of attending college, effort costs	No. MVPF is very high.
Bastian and Jones (2021)	Bastian and Jones (2021)	All post-1990 EITC Expansions for Women	Low-income workers (women)	3.18-4.23	No	Large effects on employment	Improved health, reduced crime, improved child outcomes. (All accounted for only as fiscal externalities.)	No. When accounting for all fiscal externalities, policy already "pays for itself".

Authors	**Causal estimates**	**Policy considered**	**Recipients**	**MVPF**	**Infinite-simal?**	**Agents re-optimize?**	**Non-pecuniary gains/losses**	**Welfare conclusions likely affected?**
Giesecke and Jäger (2021)	Giesecke and Jäger (2021)	Introduction of Old-Age Pensions in the UK	Individuals age 70+	0.8	No	Re-optimization of work / retirement decision. Large monetary transfer (22% of average income)	Time to help taking care of grand-children. Large change in life schedule. Financial stress relief.	Yes. MVPF close to 1. Potential welfare gain for grand-children and relief stress.
Bergolo and Cruces (2021)	Bergolo and Cruces (2021)	Conditional Cash Transfers in Uruguay	Low-income households with children and/or pregnant women. Covering 42 percent of children under 18	0.61	No	Discrete change in behaviour (health and educational conditions). Change in take-up of other social programs	Parenting time, Costs or gains of complying to the program's health and educational requirements	Maybe. Depends if the "complying with conditionalities" aspect has a positive or negative value for individuals.
Wettstein (2020)	Wettstein (2020)	Introduction of Medicare Part D	Individuals age 65+	1.98	No	If full time workers are working to get the private insurance, the policy leads to a complete re-optimization	If agents are liquidity constrained, they may value drugs more than their price, which means the WTP is higher. Gain on health and released stress from not being able to buy drugs	Yes. Some positive effects are neglected. MVPF could be higher

Authors	Causal estimates	Policy considered	Recipients	MVPF	Infinite-simal?	Agents re-optimize?	Non-pecuniary gains/losses	Welfare conclusions likely affected?
Ganimian et al. (2021)	Ganimian et al. (2021)	Early-Childhood Education in India	pre-schoolers	∞	Arguably yes	Not with the pilot. If the policy is enlarged to the whole population, it might change the population of recipients (Parents of lower or higher expected-gain children now enroll because they know that it will improve their kid's outcomes)	Schooling might be more pleasant.	No. MVPF already infinite, only additional gains.
Jácome (2022)	Jácome (2022)	Medicaid Eligibility for Teenagers in South Carolina	19 year olds losing the coverage	1.77 to 14.96	No	Induce significant life decision change. If some individuals have child to get access to the program, it change significantly their behaviours.	Reduced probability of being incarcerated. Reduced opportunities of earnings via crimes. Access to better jobs (better in non-paid compensations)	

Authors	Causal estimates	Policy considered	Recipients	MVPF	Infinite-simal?	Agents re-optimize?	Non-pecuniary gains/losses	Welfare conclusions likely affected?
Deshpande et al. (2021)	Deshpande et al. (2021)	Different evaluation standards (at known age thresholds) for eligibility to the Social Security Income in the USA	3 groups defined by leniency of evaluation standards for SSI eligibility: ¿55, 50-55, ¡50 and spillovers to homeowners	1.04	No	Large impacts on tail events, events that occur infrequently and are associated with large drops in consumption (p.152) like bankruptcy and home sale	Most treated individuals are in great financial distress so potentially large relief and positive metal health impacts of becoming eligible to SSI.	Yes. MVPF around 1 and reduced financial distress and increased access to housing are likely significant gains
Bailey et al. (ming)	Bailey et al. (ming); Hoynes and Schanzenbach (2009, 2012); Hoynes et al. (2016)	Access to Food Stamps (average $4/person/day)	US families (parents and children)	56.25	Probably not	The policy reduced food insecurity, but these are necessity purchases that would happen anyway. However, the employment response of parents suggests that, without food insecurity, some parents would not be working.	Time parents can spend away from work (leisure or child care) thanks to the food vouchers	No. MVPF very high, only additional gains.
Kline and Walters (2016)	Kline and Walters (2016)	Head Start (targeted preschool program)	Disadvantaged children in the US	1.85-2.41	Yes, if focusing only on children, but parents likely to benefit as well.	Children do not choose to attend daycare so no omitted behavioral changes	The value of child development for parents, pareting time, etc.	Possible if negative impacts on parents' utility (eg: time spent with child reduced)

Authors	Causal estimates	Policy considered	Recipients	MVPF	Infinite-simal?	Agents re-optimize?	Non-pecuniary gains/losses	Welfare conclusions likely affected?
Gray et al. (ming)	Gray et al. (ming)	Removal of work requirements for food stamps in Virginia	Potential beneficiaries of food stamps	0.86-1.15	No, which the authors acknowledge. They provide an MVPF estimate including a utility cost of working.	No strong employment responses, but removal of work requirements found to have large impacts on retention.	Contrary to Bailey et al. (2022), the authors consider utility costs of labor.	Yes. The authors themselves show that including utility costs of labor moves the MVPF from below to above 1.
Hyman (2018)	Hyman (2018)	Trade Adjustment Assistance program (retraining incentives and UI) for displaced workers	Displaced workers (due to shifts in production outside the USA)	1.14 in paper; 2.7 on Policy Impacts	No	Important employment responses to the reform. The author acknowledges the caveats of assuming Envelope conditions.	Utility costs of training programs	Possible. The MVPF estimate is slightly higher than one and some utility costs might be omitted.

Authors	Causal estimates	Policy considered	Recipients	MVPF	Infinite-simal?	Agents re-optimize?	Non-pecuniary gains/losses	Welfare conclusions likely affected?
Deshpande (2016)	Deshpande (2016)	Conditioning of access to the Social Security Income on re-evaluation of disability at 18 years old	Disadvantaged youth who would have received guaranteed SSI in adulthood	0.9 to 1.01	No	Changes in labor supply as a response to the income loss	Income stabilization value of SSI included in second estimate. However, utility cost of working (foregone leisure) is omitted.	Yes. When authors include the insurance value of stable SSI payments, MVPF goes from below to slightly above 1. The utility cost of foregone leisure might push the MVPF back below one.
Finkelstein and Hendren (2020) and Baird et al. (2016)	Baird et al. (2016)	School-based deworming program (community-wide) in Kenya	Treated school-age children and spillovers to neighboring schools	Infinite	Arguably yes	Probably not. Children would probably go to school anyway if they were not sick.	-	No. MVPF is infinite and non-pecuniary losses likely small.
Cascio (2023)	Cascio (2023)	Targeted pre-K programs in the US	Preschoolers in US States	0 (no benefit)	No	Substantial evidence from the literature that pre-K programs yield positive economic returns in the long-run	The value of child development for parents, pareting time, etc.	Yes. There should be benefits for both parents and children.

Authors	**Causal estimates**	**Policy considered**	**Recipients**	**MVPF**	**Infinite-simal?**	**Agents re-optimize?**	**Non-pecuniary gains/losses**	**Welfare conclusions likely affected?**
Cascio (2023)	Cascio (2023)	Universal pre-K programs in the US	Preschoolers in US States	1.96-4.27	No	Substantial evidence from the literature that pre-K programs yield positive economic returns in the long-run	The value of child development for parents, pareting time, etc.	No. MVPF already high and mostly additional gains.

Chapter 3

Veiling and the Economic Integration of Muslim Women in France

Antoine Jacquet* Sébastien Montpetit

Abstract

The economic implications of policies limiting the wearing of the Islamic veil for Muslim women in Western countries are still poorly understood. This paper investigates the relationship between veiling behavior and economic participation using the largest sample of Muslim women in France. Firstly, we present new descriptive evidence about Muslim women in France. We demonstrate a significant negative relationship between veiling and economic participation, which contrasts with the existing economic theory of veiling in Muslim-majority countries. Secondly, we show that a model which also accounts for reduced economic opportunities for veiled women is consistent with our findings in the Muslim-minority context. Thirdly, we develop and estimate a discrete-choice model of veiling and labor force participation to disentangle the various motivations behind the joint decision to veil and to be economically active. Our findings indicate that veiled women are less economically active not due to religious preferences, but rather because the benefits of economic participation are lower for women who veil compared to those who do not. This result echoes previous findings in the literature regarding labor-market discrimination against individuals who signal their religious affiliation. Additionally, our results emphasize the significance of personal religious motives in the decision to veil, rather than community-based religious pressure. This calls into question the rhetoric used to justify policies that restrict the wearing of religious symbols in France.

*Sciences Po Paris. Email: antoine.jacquet@sciencespo.fr

3.1 Introduction

Veiling among Muslim women has been at the center of public debates in Western countries for several decades. The Islamic veil is often perceived as a signal of both cultural distance from the majority, and of the subordination of women. It is a particularly burning issue in France, where state secularism (*laïcité*) "constitutes a pillar, even the identity and foundation of the community life."[1] At the heart of the debates lies the idea that Muslim women wear the veil against their own will and must be freed from such oppression.

To be sure, the adoption of this cultural practice entails numerous costs such as reduced employment prospects, discrimination, or physical discomfort (Abdelhadi, 2019; Valfort, 2020). Yet, as many politicians advocate for a strengthening of secular policies, it is crucial to understand the real motives behind veiling, and how it affects the economic participation of Muslim women. First, do women veil willingly despite these costs, or is veiling mostly a result of communitarian pressures? The answer to this question may lead to opposite policy recommendations: if veiling is driven by individual motives, then further restrictions on veiling may inhibit the socio-economic integration of Muslim women even more and reduce social welfare (Carvalho, 2013; Shofia, 2020). But if veiling is community-driven, then those restrictions may help emancipate them (Maurin and Navarrete-H., 2023). Second, are veiled women less economically active because of religious preferences, or because they face more obstacles in the labor market? If the latter is true, policymakers who wish to foster the economic integration of Muslim women should focus on the multiple barriers in the workplace rather than banning the veil.

Despite the considerable media, political, and academic attention, the reasons why women veil in a secular and Muslim-minority country like France are still poorly understood. This is in contrast with the context of Muslim-majority countries, which has received attention both in economics (Carvalho, 2013; Shofia, 2020) and in the wider social science literature. In Muslim-minority countries, most of the empirical evidence on veiling behavior remains based on interviews conducted over small samples of women (or adolescents). Moreover, in France, such interviews are typically conducted in the Parisian region, even though Muslims are increasingly present over the whole territory. In addition to this representativeness issue, this methodology has the inherent drawback that, especially for such sensitive topics, interviewees may be susceptible to social desirability bias. This is even more true because respondents are typically aware that the topic of the interview is veiling behavior. It is thus not clear how individuals' responses reflect true individual preferences for veiling or influences from their community.

In this paper, we make one of the first attempts at analyzing the relationship between veiling and economic participation, using rich survey data over more than 3,000

[1] Andriantsimbazovina et al. (2020), p. 7.

Muslim women in France. This sample constitutes the largest source of data on Muslim women and their veiling practices in France that we are aware of. In addition, its wide geographical coverage arguably improves representativeness compared to interview-based data. The survey also records other detailed information about respondents, providing important controls for our analysis. Overall, these data allow us to study veiling and economic participation among Muslim women empirically on a scale which hasn't been done in a Muslim-minority country before. Furthermore, this paper also extends the existing economic theory of veiling to the context of Muslim-majority countries. This structuring theory notably helps us to disentangle the role of religious motives versus those of economic motives in women's veiling and economic participation decisions.

A second objective of this paper is to unpack the various motives for veiling. By matching our main data with other sources, we are able to measure the influence of the local community on women's veiling and economic participation. We also exploit the richness of the survey to proxy for parental religious transmission, individual religiosity, and the individual's religious environment.

Our study begins with an in-depth descriptive analysis. We provide evidence that in France, veiling is associated with significantly reduced economic participation. Among Muslim women, the practice of always wearing a conspicuous religious symbol in public is associated with a 23 percentage points decline in economic participation (defined as being active on the labor market or studying) in the cross-section. This correlation is large and economically significant: in our preferred specification for instance, it is equivalent to having an additional 1.4 children aged less than 4 years old. This negative relationship is robust to several alternate specifications. In particular, exploiting the information on respondents' employment history, we construct a retrospective panel dataset of economic participation and find a correlation similar in magnitude, robust to the inclusion of year fixed effects and random effects.

In a second step, we develop a model to analyze the joint decision of veiling and economic participation. Our goal here is to provide a conceptual framework to understand the respective roles of religious motives (such as individual religiosity or religious social pressure) and of economic motives (such as employment opportunities and on-the-job discrimination) in this joint decision. The model nests Carvalho's (2013) seminal theory of veiling, where veiling is a response to individual and social religious incentives, acting both as a commitment device to follow religious norms and as a signal of the woman's commitment to her community. Our model extends this theory to fit the French context, based on our descriptive results and on our understanding of the ethnographic evidence. In addition to the religious incentives channel, we introduce economic incentives to (un)veil in the model, which reflect the documented barriers to economic participation that veiled women face. These two mechanisms have distinct implications for how the decisions to veil and to participate economically interact: according to the religious incentives channel, women should veil more when they participate (in order to signal their religious commitment despite their social integration), while according to the economic incentives channel, they should veil less (because veiling directly reduces their economic opportunities).

Finally, we translate our conceptual framework into an empirical static discrete-choice model of veiling and economic participation. We formulate and test direct implications of the theory for the two incentives channels, religious and economic. Within religious incentives, we also distinguish between intrinsic motives, which we measure using multiple indicators of religiosity (both subjective feelings and actual religious practices) available in the survey data; and social motives, for which we develop several proxies. Parental influence is measured using the (self-reported) importance of religion in the education received by the respondent, and religious name-giving. For communitarian pressure, given that data on religious diversity is not available in France, we use the share of Maghrebi immigrants in the local population as well as the local number and size of Muslim places of worship (mosques and prayer rooms).

Our main empirical findings are twofold. First, we find supporting evidence for the economic discrimination channel described in the theory, but not for the religious incentives channel. This result suggests that the impact of religious motives on the economic participation decision is mostly indirect (through the decision to veil), while economic motives seem to have a direct impact on the decision to veil. In other words, the primary reason why veiled Muslim women work less (or, equivalently, that working Muslim women veil less) seems to be that veiling itself reduces their economic opportunities, and not that religiosity disincentivizes working. As such, the lower economic participation of Muslim women could be understood as a demand-side problem on the labor market, more than a supply-side one.

Second, we measure the respective roles of the different religious motives in the decision to veil. While measures of social religious pressures are correlated with veiling behavior, we find that a much larger share of the variation in veiling patterns can be explained by individual religiosity. Our results thus question the rhetoric often used to justify policies restricting the wearing of religious symbols in France. Consistent with our analytical results, we conjecture that regulations which limit the expression of religious faith in public are likely to impede integration of Muslim women into Western societies.

3.1.1 Related literature and contributions

This paper contributes to several strands of the literature. First, it provides novel empirical evidence to the vast literature on Islamic veiling in the social sciences.[2] In this

[2]We review in detail the literature on veiling in France in section 3.2. Recent contributions in other contexts include Harrison (2016) for the United States as well as Aksoy (2017) and Aksoy and Gambetta (2016, 2021) for Turkey.

literature, most of the evidence is based on interviews with Muslim women since veiling behavior is rarely observed in surveys or other standard datasets. While interviews have the potential to dig deeper into specific questions of interest and uncover a large number of potential channels, they often suffer from small sampling and representativeness issues. In a recent contribution, Shofia (2020) measures the veiling rate at the district level to circumvent this problem and provides robust empirical evidence that better economic opportunities for women induce Indonesian women to veil. In contrast, in this paper, we study the case of a secular country in which Muslims form a minority and where wearing the veil is frowned upon rather than encouraged. Similar conclusions to that of Shofia (2020) were reached by Aksoy and Gambetta (2016), the closest study to ours, for the case of Turkey. Aksoy and Gambetta (2016) also attempt to study the determinants of veiling in a Western country, namely Belgium. However, they do not have a direct measure of veiling behavior, but rather a measure of attitudes towards veiling in public. Moreover, the richness of our data allows us to further unpack the relative weight of various incentives that are difficult to measure in the decision to wear the Islamic veil over a large sample. In particular, we can distinguish between private and communitarian incentives to veil, a question which has so far eluded empirical researchers. Another close study is that of Abdelhadi (2019) who finds that the wearing of the veil is associated with lower employment in the United States, but does not investigate the motives for veiling. Her result is consistent with our findings for France, for which we document large differences in economic participation between veiled and non-veiled women.

Second, we bring new evidence on motives for adopting costly cultural practices both theoretically and empirically. In the vast literature on the economics of religion and identity, it is now acknowledged that individuals may choose their identity via rational decision-making even if it requires costly investments or sacrifices (Iannaccone, 1992; Akerlof and Kranton, 2000; Atkin et al., 2021; Jia and Persson, 2021). Though potentially rational, adopting (or transmitting) certain cultural practices can be an impediment to social and economic integration of certain groups. A strand of the literature has investigated the incentives that might justify such choices. Recent examples include foot-binding in China (Fan and Wu, 2022), female genital cutting in Africa (Bellemare et al., 2015; Novak, 2020; Gulesci et al., 2021), and baby-naming choice in France (Algan et al., 2022).[3] We contribute to this literature in three ways. First, we document that in France veiling is associated with reduced economic integration of Muslim women, as opposed to the evidence from Muslim-majority countries (Aksoy and Gambetta, 2016; Shofia, 2020), and we provide detailed descriptive evidence of why Muslim women might wear such a costly signal of religious identity in France. Second, we rationalize this finding

[3]There is also a relevant literature looking at incentives to abandon certain costly cultural traits and adopting less harmful ones. For example, Biavaschi et al. (2017) find important economic payoffs for the Americanization of migrants' names. See also Bisin et al. (2011, 2016) and Drydakis (2013) on economic returns of assimilation for migrants.

by adapting the theory of Carvalho (2013) to a Muslim-majority context, in which the expression of Muslim identity clashes with economic integration instead of facilitating it. Third, we uncover novel empirical patterns concerning the wearing of *discreet* signs of religious affiliation, which have received little attention in the literature. In particular, they appear to be worn by Muslim women who are educated and moderately religious. These patterns might suggest discreet symbols, in the French context, play a similar role to that of the veil in Muslim-majority countries.[4]

Third, our results have implications for State secularisation policies. Of particular interest in our context, two recent empirical studies reach opposite conclusions on the effects of the French headscarf ban in public schools. On the one hand, Abdelgadir and Fouka (2020) find that the 2004 ban depressed schooling outcomes of French girls of North-African origin.[5] On the other hand, Maurin and Navarrete-H. (2023) find that the 1994 ministerial circular asking school principals to prohibit the wearing of the veil in schools had a positive impact on their educational attainment. Even if they are comparing different cohorts of adolescents and different treatments, these contradictory pieces of evidence are puzzling. By focusing on why Muslim women are willing to sacrifice economic opportunities to veil, we can offer a new perspective to this debate. If incentives to veil are mainly *private*, more stringent secular regulations should reduce incentives to integrate for religious women who wish to veil. On the contrary, if *communitarian* incentives prevail, such veil bans may help women emancipate and liberate them from a costly religious norm which limits their economic opportunities. Our results lend support to the former interpretation. The main observed drivers of veiling behavior in France appear to be the woman's religiosity as well as non-religious identity such as her origins. Religious pressures from women's close community are also correlated with veiling behavior, but turn out to explain only a small share of variation in veiling behavior in our regressions. Proponents of French secular regulations often base their arguments on the idea that Muslim women simply do not want to veil and are forced to do so by other Muslims. Our analysis thus casts serious doubts on this assumption and suggests that the French secular regulations most likely inhibit social and economic integration of Muslim women in France rather than facilitating their emancipation.

The rest of the article is structured as follows. Section 3.2 provides the institutional context. Section 3.3 describes the data sources and provides a detailed descriptive analysis of veiling patterns in France. Section 3.4 outlines our theoretical framework. Section 3.5 translates this framework into an empirical model and covers its estimation. Finally, section 3.6 concludes.

[4]We, however, have little statistical power to test this hypothesis because few Muslim women wear only discreet symbols in our sample.

[5]In a similar spirit, Benzer (2022) finds that the re-introduction of Islamic schools, which do not prohibit the headscarf, had positive impacts on girls' educational attainment in Turkey.

3.2 Historical and sociological background

The wearing of the Islamic veil has been a burning issue in France since at least three decades. In 1989, the "*affaire des foulards*" (headscarf affair) garnered nationwide attention when three girls were expelled from their middle school for refusing to remove their headscarves. The incident sparked heated debates but eventually culminated in the highest French administrative court ruling in favor of the expelled girls (Scott, 2009). Despite this ruling, in 1994 the Ministry of Education issued a circular asking school principals to prohibit conspicuous religious symbols worn by students. This controversial position was later enshrined in a 2004 law, whose supporters argued that headscarves "infringed on the liberty of conscience of other pupils and represented the triumph of communitarian pressures" (Abdelgadir and Fouka, 2020, p. 4). The debate then shifted to other public spaces, with a nationwide ban of full-face veils (*burqa*) in 2010, and later with several city bans of the *burkini* in swimming areas and beaches.[6]

Despite the significance of these policies for Muslim women and girls, they have largely been excluded from the conversation. In fact, this "one-sided debate"[7] has revealed a lack of understanding among policymakers about the realities and constraints faced by the Muslim population (Scott, 2009; Nordmann, 2004). Nevertheless, considerable research in sociology and anthropology has been dedicated to understanding the experience of Muslims in France, and particularly the reasons for women to wear the veil. In the following paragraphs we focus on two factors which have been shown to be significant in that decision: balancing religious and family expectations with societal integration, and the potential impact of veiling on economic participation due to discrimination.

Why do women veil? France's secular policies against veiling have been justified by the idea of a "silent majority" of Muslim women who are forced to wear the veil by their families or communities. According to this idea, the benefits of helping this silent majority outweigh the harm imposed on other female Muslims who truly want to veil (Maurin and Navarrete-H., 2023). However, existing evidence on the motives behind veiling behavior contradicts this argument. In fact, interviews and surveys conducted in France suggest that the vast majority of Muslim women who wear the veil do so by individual choice and not out of coercion (IFOP 2019, Institut Montaigne 2016). Even within the Muslim community, the motives behind veiling seem to be misinterpreted. For instance, non-veiled Muslim women are more likely to believe that veiling is done out of coercion or imitation (IFOP 2019). This discrepancy highlights a key limitation of interview data: it is unclear whether "individual choice" reflects the preferences of

[6] The question of veiling in public resurfaced for instance during the debates surrounding the adoption of the "law on separatisms" of August 2021, with some Senators suggesting a complete ban of all religious symbols in public spaces (see Sénat, 2021).

[7] Gresh (2020).

the women themselves, or the internalization by these women of the preferences of their social networks.

In a series of interviews with Muslim girls and women,[8] Gaspard and Khosrokhavar (1995) identified three broad categories of veiled women: "veiled immigrants," i.e. middle-aged women who arrived in France veiled and kept the practice; adolescent girls born in France who wear the veil either by force or by choice; and young women who wear the veil willingly to reconcile their religious duties and integration into French society. The veil worn by first-generation immigrants is well tolerated by French society. Animosity is instead directed towards the veils worn by adolescents and young women born in France, which is perceived as a symbol of failed integration – "a sign of inherent non-Frenchness" (Scott, 2009, , p. 15).

When asked why they wear the veil, Muslim women mostly invoke religious duty (76%) and issues of safety (35%) (Institut Montaigne, 2016). Young women in particular mention "the difficulty to reconcile their families' demands with those of the society" (Khosrokhavar, 2004, p. 90). Familial pressures typically discourage them from engaging in activities that favor their integration, such as going out with friends or finding a job. In this respect, veiling can be a tool which allows them to "exempt themselves from the constraints that traditionally weigh on women" (Gaspard and Khosrokhavar, 1995, , p. 37) and to resolve the tension between religious duty, families' demands, and integration.[9]

This interpretation of veiling as facilitating integration is in line with research in economics which has explored veiling practices in relation to economic participation (Carvalho, 2013; Shofia, 2020). The theory of Carvalho (2013) considers veiling as a technology available to Muslim women in order to alleviate the intrinsic and social costs of their integration. By providing a practical protection against opportunities to engage in religiously prohibited behaviors, veiling acts both as a commitment to oneself and as a signal of this commitment to others. This commitment aspect of veiling is confirmed by survey evidence and interviews conducted in France and elsewhere.[10] Furthermore, Shofia (2020) provided evidence for this mechanism in a study of veiling among Indonesian

[8] Gaspard and Khosrokhavar (1995) conducted around one hundred interviews with Muslim girls and women in the Paris and Dreux suburbs.

[9] The following interview excerpts collected by Atasoy (2006) in Canada also illustrate this tension well:

> "It is hard as a young woman not to have a boyfriend in this society. [...] The veil reminds you that this isn't allowed [in Islam]."
>
> Sarah believes the veil keeps her away from doing "stupid things like dating a guy."
>
> "The veil reminds me that I submit to Allah... If I don't wear it, people might take it as I'm doing something wrong."
>
> "If you are not covered, you feel isolated from other Muslim girls. They don't socialize with you. They think you are doing bad things."

[10] See for example Atasoy (2006) for Canada and Read and Bartkowski (2000) and Droogsma (2007) for the United States.

schoolgirls.

Veiling and economic participation. The sociological and anthropological record documents the challenges faced by veiled women in France when trying to integrate into the workforce (Adida et al., 2010, 2016; Jouili, 2020). Alongside the policies restricting religious expression in public areas, veiled women encounter various constraints in the workplace. For example, French civil servants have an obligation of religious neutrality – a strict application of *laïcité*, the French conception of state secularism. This obligation prohibits the expression of religious beliefs while on duty, including the wearing of conspicuous religious symbols. Breaching this obligation is considered a serious offense that can lead to sanctions or even dismissal.

Veiled women also encounter obstacles in the private sector (Ajbli, 2011). First, private-sector workers providing a public service are also subject to neutrality requirements. Second, since August 2016, private firms can introduce neutrality requirements in their internal rules of procedure. The law states that it is allowed "as long as these restrictions are justified by the exercise of other liberties and fundamental rights or by the necessity of the good functioning of the firm, and if they are proportionate to the pursued goal."[11] Famous cases of firms who introduced neutrality requirements include a private kindergarten and a recycling factory. Third, studies have shown that Muslims, particularly those who display higher levels of religiosity (a trait associated with wearing the veil), face discrimination when seeking employment. Valfort (2020) uses a correspondence-test method to demonstrate that while signalling religiosity increases call-back rates for Christian applicants, it significantly reduces them for Muslim applicants in France.[12] Similar discriminatory hiring practices have been reported in other European countries.[13]

Employers claim that discrimination against Muslims is due to religious expression causing conflicts, and accommodating religious practices is viewed as a challenge (Adida et al., 2016; Cintas et al., 2012). Muslims, in particular, face discrimination as some of their religious practices, such as daily prayers and fasting, are perceived as reducing productivity (Bouzar and Bouzar, 2009; Maillard, 2017).[14] In its yearly surveys of French managers, the *Observatoire du Fait Religieux en Entreprise* documents a rise in observed religious behaviors requiring managerial intervention, with Islam being by far the most cited religion (Institut Montaigne, 2021).[15]

[11] https://www.legifrance.gouv.fr/codes/id/LEGIARTI000033001625/2016-08-10

[12] Valfort (2020) uses extra-curricular activities (volunteering for a Christian or a Muslim Scout association) as a signal of religiosity.

[13] Weichselbaumer (2020) and Fernández-Reino et al. (2022) also use correspondence tests to confirm the existence of discrimination against veiled women in Germany, the Netherlands, and Spain.

[14] Hu and Wang (2021) provides empirical evidence suggesting that Ramadan fasting does not in fact reduce productivity.

[15] The *Observatoire du Fait Religieux en Entreprise* conducts surveys on religious behaviors in the

Of course, Muslim women report wearing the veil for various other reasons, including signaling piety to potential husbands, or even fashion (Patel, 2012). Worth mentioning are identity motives that are not necessarily religious. For some Muslim women, the veil is a means to affirm their distinction with the rest of society and to feel closer to their community of origin (Silhouette-Dercourt et al., 2019). For instance, adolescents who want to distinguish themselves from their peers may use the veil as a visible sign of difference from the "rooted French" (Khosrokhavar, 2004; van der Hasselt, 2019). In some cases, wearing the veil is a form of rebellion against a society that claims to defend liberty of choice but discriminates against Muslims, as evidenced by studies on "identity backlash" (Abdelgadir and Fouka, 2020).[16]

3.3 Data and descriptives

In this section we start to explore empirically the relationship between veiling behavior and economic participation. We present our main data sources, and we describe them along several dimensions of interest. We first provide novel descriptive evidence on French Muslim women's living conditions. The data suggest a strong negative correlation between veiling behavior and economic participation in this population.

3.3.1 Data

Our primary data source is the cross-section from the *Trajectoires et Origines* survey (henceforth TeO; Beauchemin et al. 2016). Conducted in 2008–2009 by the French National Institutes for Demographic Studies and for Statistics and Economic Studies (INED and INSEE), the TeO survey targeted adults between 18 and 60 years old residing in metropolitan France. Purposefully oversampling immigrants and minorities, it includes 3,033 women who identify as Muslim. To our knowledge, this is the largest sample of this kind in France.[17] When including Muslim men and other religious groups, the entire survey contains more than 21,000 observations.

The TeO dataset is a comprehensive source of information on various aspects of respondents' lives, including living conditions (such as employment, education, housing,

workplace. Islam is most frequently associated with observed religious behaviors (73% in 2021), and the proportion of observed religious behaviors requiring managerial intervention has risen from about 25% in 2014 to over 50% in 2021. Of those cases requiring intervention, 19.5% resulted in conflicts in 2021, compared to 6% in 2014. When discriminatory situations in hiring are observed, they involve Muslims in 70% of cases, according to manager reports. In addition, 10% of managers feel overburdened by religious behaviors in their company (Institut Montaigne, 2021).

[16] See also Fouka (2020) and Sakalli (2019) for evidence of cultural backlash against assimilation policies in other contexts.

[17] Two surveys conducted by private firms, namely Institut Montaigne (2016) and Institut Français d'Opinion Publique [IFOP] (2019), have much smaller sample sizes (slightly above 1,000 individuals of Muslim origin, both genders included) and do not have a similarly deep content as that of TeO.

commune of residence, and health), social life (such as migration history, language use, family, and children), and public life (such as political views, experiences of discrimination, and social relationships). Of particularly value for this study is the religion section, which is a unique inclusion in a French survey of this scale since the collection of individual information on religion is closely monitored in France. This section includes variables such as religious affiliation, measures of religiosity, religious symbols worn, and intergenerational religious transmission.

We also use the TeO survey to create a panel dataset of respondents' lifetime education and labor-market status. The dataset is constructed by analyzing respondents' retrospective accounts, year by year, of their work status including salaried work, self-employment, unemployment, studying, staying at home, inactive for other reasons, or out of metropolitan France.

Our second data source is the *Annuaire des mosquées de France* (La Boussole, 2004), a comprehensive directory of mosques and Muslim praying rooms in France. This is a novel data source in the literature, which we digitized manually. Compiled by a Muslim association in 2003–2004, the directory provides for each worship facility at the time its full address and estimated capacity by gender.

3.3.2 Measurement

Alongside standard metrics of economic activity, our empirical analysis relies on measures of religious practice and religious social pressure which we describe here.

Veiling. We use the following question from the TeO survey:

> *In your daily life, do you wear in public a piece of clothing or jewelry that might evoke your religion? (1) Never (2) Sometimes (3) Always*

If applicable, respondents were subsequently asked to report which religious symbols they wear. Answers were later sorted by the survey institute into four categories: Jewelry, Clothing, Headcoverings, or Others. Because they visibly signal religion and are the ones usually targeted by secular policies, we group the Clothing and Headcoverings categories together as *conspicuous symbols*. Among Muslim women this is an excellent proxy for veiling, since headcoverings represent 93% of these conspicuous symbols. In contrast, we group Jewelry and Other symbols, which can usually be hidden, as *discreet symbols*.[18] We then cross these categories with the initial answer on frequency of wearing religious symbols. Thus, in our measure of veiling each respondent is categorized as wearing either (1) *no symbol* (if they answered *Never* to the initial question), (2) *sometimes discreet*

[18] A respondent who wears both discreet and conspicuous symbols is categorized as wearing conspicuous symbols.

symbols, (3) *always discreet symbols*, (4) *sometimes conspicuous symbols*, or (5) *always conspicuous symbols*.[19]

Individual religiosity. The TeO survey includes several questions which relate to individual religiosity. Our preferred measure is the frequency of attendance of religious ceremonies, a standard measure of religiosity which focuses on religious practice (Iyer, 2016). To analyze incentives for veiling we combine this measure with other questions related to individual religiosity: the self-reported importance of religion in the respondent's life, whether she uses her religion to self-identify, the respect of religious dietary restrictions, and religious marriage. In order to aggregate the answers to these questions into a single measure of individual religiosity, we use a measurement system, as in Heckman et al. (2013) or Bolt et al. (2021), to construct a latent index of individual religiosity. The advantage of this method is that we are able to leverage the variation on several survey questions while keeping the convenience of a single, continuous measure of religiosity. (In Appendix B.1.1 we provide details on the procedure and on the survey questions.)

Family and community pressures. As discussed in section 3.2, religious social pressures play a role in women's decisions to integrate socio-economically and to veil. Drawing on insights from the literature on cultural transmission (Bisin and Verdier, 2000), and particularly on the distinction between vertical transmission (from parents to children) and horizontal transmission (between peers), our measures of social pressure aim to disentangle the respective influences of women's families and of their larger communities on their decisions.

To capture vertical religious pressure by parents, our preferred measure is a question on the self-reported importance of religion in the respondent's education. We also use whether or not the respondent has a religious first name.[20] We then combine these measures into a single index.

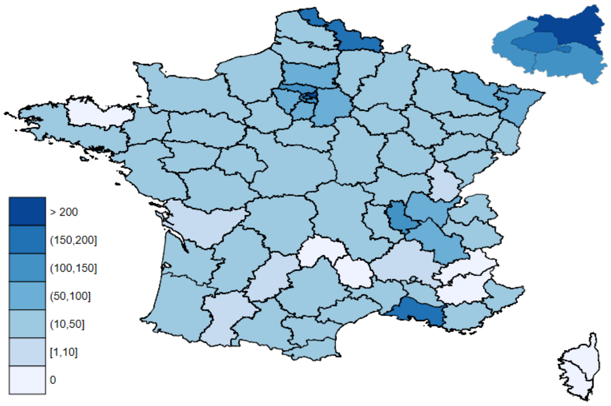
For social pressure stemming from the local community, our preferred measure is the share of Maghrebi immigrants in the neighborhood.[21] We also use a second measure,

[19] A limitation of this data is that appreciations like "sometimes" or "always" remain subjective. For instance, a woman who removes her veil in the workplace by obligation might still consider that she "always" wears it – when she is able to. In our data, a few Muslim women do report veiling "always" even though they work in the public sector, where conspicuous religious symbols are prohibited (cf. section 3.2).

[20] Name-giving has been recognized as an important cultural transmission channel (Fryer and Levitt, 2004; Abramitzky et al., 2020; Algan et al., 2022). We classify as religious the names of the Islamic prophet's wives, Khadija, Sawda, Aicha, Hafsa, Zainab, Hind, Juwairiya, Safiya, Ramla, and Maimuna (Morsy, 1989); and of his daughter Fatima. Variations in spelling are permitted. For male first names, we follow Sakalli (2019) by considering a name as religious if it is a variation of the prophet's name (Mohamed in French) or if it begins with "Abd-" ("servant of. . . " in Arabic).

[21] The precise geographical unit is the IRIS level. Having a parent (especially a father) born in Maghreb is a strong predictor of Muslim affiliation in France (Abdelgadir and Fouka, 2020).

Figure 3.1: Geographical distribution of Muslim women in the TeO survey.



Note: Number of places of residence of Muslim women in the TeO survey per *département*. Some *départements* are collapsed together when counts are low due to confidentiality reasons. The top-right subfigure zooms in on Paris and its suburban area.

the local worship capacity per thousand inhabitants, which we construct from our novel data on Muslim worship facilities in France by combining information on the place of residence of TeO respondents with the addresses and estimated capacity of these worship facilities. Since these measures are already continuous, we use them as they are and do not aggregate them into an index.

3.3.3 Descriptive evidence

Using the TeO data, we provide novel summary statistics on Muslim women in France, and especially new empirical evidence for the negative relationship which exists between veiling and economic participation among Muslim women in France, thus echoing the ethnographic evidence outlined in section 3.2. To illustrate the magnitude of this relationship, our preferred specification suggests that consistently wearing a conspicuous religious symbol is associated with a decrease in economic activity equivalent to having an additional 1.4 preschool-age children or 5.4 fewer years of experience in the labor market.

Geographical coverage

The representativeness of the ethnographic studies discussed in section 3.2 is limited due to their predominant focus on the Parisian suburbs, some of which are distressed areas that may not accurately reflect the living situations of Muslim women as a whole. In contrast, the TeO survey includes Muslim women from a diverse range of locations, as illustrated in Figure 3.1. Although some respondents remain concentrated in major urban centers such as Paris, Marseille, and Lille, the survey has a wide geographical coverage across the country.

Summary statistics

Table 3.1 presents summary statistics for our main variables of interest, disaggregated by veiling behavior. Panel A examines demographic characteristics and economic outcomes, such as employment and educational attainment. The data reveals that veiled Muslim women have significantly worse economic outcomes compared to those who wear no symbol or discreet ones. On average, they are much less educated, less likely to be employed, and have fewer years of work experience, despite being older. Particularly striking is the sharp difference in activity rates (activity being defined as either working, looking for a job, or studying). Almost two-thirds of women who always veil are inactive, compared to less than 20% for non-veiled women, indicating significant barriers to integration linked to veiling.

Panel B examines our primary measures of religiosity and religious social pressure. We observe a positive link between both individual religiosity and veiling, and religious social pressure. On average, veiled Muslim women attend religious ceremonies more frequently, received an education which stressed the importance of religion more, and they now live in neighborhoods with higher proportions of Maghrebi immigrants. Our other measures of religiosity and religious social pressure confirm these patterns (Appendix Table B.1).

Veiling is negatively correlated with economic participation

Our summary statistics provide some preliminary evidence of the negative link between veiling and economic participation, which we now investigate further using regression analysis. We perform two regression exercises, which complement each other.

First, we explore the relationship between Muslim women's active status and veiling in the cross-section. With this approach, we are able to include a rich set of controls by using the wide range of information on respondents available in the TeO survey. We also check the robustness of our results by restricting attention to particular subsamples and by conducting placebo tests on populations other than Muslim women.

Our second approach is to explore this relationship in a panel dataset that we construct from respondents' retrospective accounts of their studies and professional trajectories. Since this retrospective account focuses on a few questions only, our set of controls is more restricted. However, the panel dimension does allow us to verify that the relationship between veiling and economic activity is not merely due to the particular timing of the survey. Timing might indeed be a concern since the survey was conducted around the time of the Great Recession, which may have affected veiled women disproportionately, e.g. if they faced stronger discrimination. Together, the two exercises thus provide a robust assessment of the correlation between veiling and economic participation.

Table 3.1: Summary statistics by veiling status, Muslim women.

		By veiling behavior				
	All Muslims	No symbol	Sometimes discreet	Always discreet	Sometimes consp.	Always consp.
Panel A: demographics and economic outcomes						
Demographics						
Age in 2008	34.59	35.55	28.40	25.06	35.94	36.00
First-gen. immigrant	0.62	0.61	0.24	0.51	0.68	0.78
Second-gen. immigrant	0.38	0.39	0.66	0.49	0.32	0.22
Number of children	1.87	1.78	1.11	0.63	2.26	2.79
Has a partner	0.61	0.59	0.49	0.48	0.68	0.74
Not a French speaker	0.10	0.07	0.02	0.01	0.14	0.32
Labour-force status in 2008						
Employed	0.46	0.54	0.43	0.36	0.44	0.22
Unemployed	0.17	0.18	0.23	0.27	0.12	0.09
Inactive	0.28	0.19	0.15	0.24	0.30	0.65
Student	0.09	0.09	0.20	0.13	0.14	0.03
Has never worked	0.28	0.19	0.29	0.48	0.31	0.50
Schooling attainment and work experience						
Completed high school	0.57	0.61	0.66	0.42	0.51	0.43
Higher education degree	0.21	0.22	0.24	0.10	0.20	0.19
Years of schooling	14.59	15.30	17.41	15.69	12.86	11.11
Years of work experience	5.85	7.06	3.93	3.44	5.75	2.66
Panel B: religious characteristics						
Attends religious ceremonies						
Familial ceremonies only	0.27	0.29	0.32	0.29	0.30	0.18
Religious feasts only	0.22	0.20	0.34	0.21	0.32	0.27
Once or twice a month	0.04	0.03	0.05	0.02	0.08	0.09
At least once a week	0.06	0.02	0.01	0.03	0.11	0.19
Importance of religion in education received						
A little important	0.14	0.18	0.11	0.05	0.04	0.04
Quite important	0.29	0.31	0.30	0.29	0.33	0.15
Very important	0.51	0.47	0.58	0.64	0.63	0.81
Percentage of Maghrebi immigrants in neighborhood						
Fourth quintile	0.22	0.27	0.31	0.19	0.32	0.32
Top quintile	0.71	0.43	0.40	0.54	0.44	0.47
Observations	3,033	2,017	166	151	148	516

Note: This table reports means of variables of interest by veiling status as defined by the type of symbol and the frequency at which they are worn.

Table 3.2: Veiling and economic participation, Muslim women.

	Woman is active (= 1 if active, = 0 if inactive)					
	(1)	(2)	(3)	(4)	(5)	(6)
Veiling behavior						
Sometimes discreet symbol	0.029	-0.052	-0.054	-0.054	-0.028	-0.036
	(0.041)	(0.037)	(0.035)	(0.037)	(0.036)	(0.034)
Always discreet symbol	0.117***	0.019	0.036	0.028	0.038	0.055*
	(0.028)	(0.028)	(0.029)	(0.028)	(0.029)	(0.031)
Sometimes conspicuous symbol	-0.107*	-0.090**	-0.072	-0.083*	-0.053	-0.055
	(0.055)	(0.046)	(0.044)	(0.047)	(0.038)	(0.037)
Always conspicuous symbol	-0.441***	-0.351***	-0.300***	-0.328***	-0.268***	-0.232***
	(0.048)	(0.032)	(0.030)	(0.033)	(0.031)	(0.030)
Demographics						
Number of children			-0.051***			-0.028**
			(0.012)			(0.011)
Number of children below age 4			-0.165***			-0.171***
			(0.022)			(0.020)
Lives in a couple			-0.065**			-0.065*
			(0.031)			(0.033)
Educational attainment and work experience						
Years of schooling					0.011***	0.010***
					(0.003)	(0.002)
Completed high school					-0.017	-0.021
					(0.025)	(0.024)
Higher education degree					0.059**	0.040*
					(0.024)	(0.021)
Years of work experience					0.041***	0.039***
					(0.004)	(0.004)
Experience squared					-0.001***	-0.001***
					(0.000)	(0.000)
Constant	0.812***	0.514***	0.341**	0.481***	0.055	-0.144
	(0.016)	(0.138)	(0.140)	(0.157)	(0.146)	(0.157)
Other demographic controls			✓			✓
Religious controls				✓		✓
Birthyear dummies		✓	✓	✓	✓	✓
Age of arrival in France dummies		✓	✓	✓	✓	✓
Birthplace dummies		✓	✓	✓	✓	✓
Region of residence dummies		✓	✓	✓	✓	✓
Observations	2433	2433	2433	2433	2433	2433
R^2	0.147	0.358	0.428	0.374	0.450	0.511

Note: This table reports results of linear regressions on a dichotomous variable taking the value of 1 if a woman reports being in the labor force or studying. The other demographic controls are dummies indicating whether the individual is a first-generation immigrant, has an Arabic-sounding first name, has a partner working, has a parent born in France, as well as levels of feelings of French identity. Also included in each regression is a set of dummies capturing the conditions in which the survey took place (whether the partner was present, whether parents were present, survey month dummies, age group of surveyor dummies, and surveyor's gender). The religious controls include measures of individual religiosity (levels of importance of religion in own life and of religious practice as well as a dummy indicating whether the woman uses religion to self-identify) and of religious influences from the community (dummies for whether each parent is Muslim, has a religious first name, has a Muslim partner, most of her friends are Muslims, shares of Muslims in the neighborhood, and for levels of importance of religion in her education as well as the number of seats in places of worship in the local area of residence.) The last regression also includes a dummy for whether the individual has right-wing political opinions. The sample is restricted to Muslim women with no missing covariates. Observations are weighted using the weights provided in the TeO survey. Robust standard errors in parentheses. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Cross-sectional analysis. Table 3.2 shows the results of linear regressions where the outcome variable is the activity status (0 if inactive, 1 if active), and the main explanatory variable is the respondent's veiling behavior. Other important explanatory variables include our measures of individual religiosity and religious social pressure, economic characteristics which are usual predictors of labor market participation such as education and experience, and other demographic predictors. The sample is restricted to Muslim women with non-missing covariates, yielding 2433 observations.

Column (1) includes veiling behavior as the only predictors of active status. Veiling behavior alone is an important predictor of the activity rate, explaining 13.5% of the variation in the activity status. In columns (2) to (6) we add more controls, including dummy variables for birth year, age of arrival in France, birthplace, and region of residence. We further include a set of dummy variables capturing the conditions in which the survey took place (whether the partner was present, whether parents were present, survey month dummies, age group of surveyor dummies, and surveyor's gender), which gives us confidence that social desirability bias is minimized in our regressions.

We include additional groups of control variables one by one to investigate the relative contribution of different mechanisms. The last column reports the results of a regression controlling for all of the covariates. In this last specification, the only significant predictors of the activity status are the wearing of conspicuous symbols, the number of children, age, birthplace, and the education level. The magnitude of the main coefficients of interest is reduced compared to specifications with a sparser choice of controls, but it remains statistically and economically significant. The point estimates indicate that Muslim women who always wear a conspicuous symbol are 23 p.p. less likely to be active compared to those who never wear any symbol. Even in this most parsimonious specification, the estimated effect is substantial: it is equivalent to having an additional 1.4 preschool-age children.

Robustness checks. Overall, the regression results of Table 3.2 confirm a strong negative association between veiling and economic participation. We further verify the validity of this statement through a series of robustness checks, the results of which are summarized in Table 3.3. The first three columns correspond to re-estimations of our preferred specification (column 6, Table 3.2) in different subsamples. The goal of this exercise is to verify that our results are not driven by particular observations or simply capturing something else apart from the potential impact of veiling. The first row excludes students to use a more conventional measure of economic participation, that is, labor-market participation. The second row excludes individuals born outside France, since summary statistics suggested an important difference in immigration status between veiled and non-veiled women. The third row excludes women whose religious symbol is categorized as Other (i.e. neither Clothing, Headcoverings, or Jewelry). Restricting

Table 3.3: Robustness checks, cross-sectional data

	Excl. students (1)	Born in France (2)	Excl. "other" symbols (3)	Other religious groups (placebo): Muslim men (4)	Other religious groups (placebo): Excl. Muslims and Catholics (5)	Other religious groups (placebo): All non-Muslims (6)
Veiling status						
Sometimes discreet	-0.033	0.035	-0.040	0.021	0.017	-0.017
	(0.040)	(0.032)	(0.034)	(0.013)	(0.018)	(0.012)
Always discreet	0.072	0.049	0.058	0.040***	-0.008	-0.022*
	(0.037)	(0.034)	(0.031)	(0.012)	(0.019)	(0.012)
Sometimes conspicuous	-0.063	0.075*	-0.058	-0.050*	0.022	0.012
	(0.044)	(0.036)	(0.037)	(0.029)	(0.068)	(0.044)
Always conspicuous	-0.234***	-0.246***	-0.226***	0.016	0.080	0.066
	(0.031)	(0.052)	(0.030)	(0.078)	(0.154)	(0.141)
Controls	✓	✓	✓	✓	✓	✓
Observations	2,158	1,199	2,427	2,197	1,756	5,744
R^2	0.510	0.411	0.517	0.204	0.245	0.196

Controls included in the regressions are the full set of variables included in Table 3.2, column (6). In column (1), we exclude students so that the dependent variable becomes labor-market participation. In column (2), the estimation sample is restricted to second-generation immigrant Muslim women (born in France of foreign parents). In column (3), individuals reporting to wear a religious symbol that is neither jewelry, a headcovering, or clothing (symbols labelled as "other") are excluded from the sample. Columns (4) to (6) estimate the same regression on other religious groups. Robust standard errors in parentheses. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

attention to these subsamples yields point estimates for the effect of conspicuous-symbol wearing which are of similar magnitude to those obtained on the complete sample.

Columns four to six of Table 3.3 re-estimate the same specification, this time on groups other than Muslim women, thus providing a form of placebo test. We find that wearing a religious symbol has no significant association with economic participation for Muslim men, nor for women and men with different religious affiliations. These results confirm the unique place of the Islamic veil among other religious symbols, as evidenced by the debates mentioned in section 3.2. Whether it is because of individual preferences, social pressure, legal restrictions on veiling at work, or discrimination, veiling seems to be the only widespread religious symbol which is strongly associated with decreased economic participation.

Panel analysis. We perform another robustness check in order to control for timing effects, in particular in the event that veiled women's employment prospects were differentially affected by the 2008 economic crisis (which coincided with the time of the survey). To investigate this possibility, we use the retrospective panel dataset, where we exclude observations for which individuals report multiple activities as well as periods in which the respondent was out of metropolitan France. This empirical strategy allows us to control for time-varying observables and time fixed effects, to substantially increase the number of observations, as well as to include random effects. For the sake of space, we present this analysis in Appendix B.1.3. The results overall confirm the findings obtained in the cross-sectional analysis, with the wearing of a conspicuous symbol being associated with a significant decline in economic participation that is similar in magnitude.

3.4 A model of veiling and labor supply

In the previous section, we have shown that veiling displays a strong negative association with economic participation in France. Our discussion of the literature on veiling from section 3.2 suggests that such an association can originate from two sources of incentives, namely religious (women who veil are more religious and therefore less likely to engage with an environment they perceive as dangerous) and economic (women who veil face discrimination on the labor market).

In order to structure our empirical analysis of these motives, in this section we model Muslim women's joint decision of economic participation and veiling. This model builds on the theory of Carvalho (2013), who considered the veil as a tool available for women to mitigate the socio-religious cost of their integration. We expand on this model by proposing a general analytical framework which remains agnostic as to the reasons why women veil. We then show that this general framework can be specified to accommodate together both religious motives in the spirit of Carvalho (2013), as well as economic motives stemming from anti-veil discrimination on the labor market.

3.4.1 General model

We consider a static model in which an agent must simultaneously decide on her labor supply and her veiling behavior. For her labor supply, she allocates her total time budget $T = 1$ between time worked, t, and time devoted to leisure, $1 - t$. In addition, she chooses what degree of veiling to adopt at work, v_1, and what degree of veiling to adopt during her leisure time, v_0. The flow utility that the agent derives from work and leisure then depends on her degree of veiling in each of these activities. The model remains agnostic about whether veiling has a positive or negative effect on the flow utility of working or leisure. In this way, it is able to account for a wide range of mechanisms linking veiling behaviors and labor supply decisions, from the religious stigma faced by working Muslim women to identity-based discriminations at and outside work.

Formally, the utility that the agent derives from working, u_1, and the utility that she derives from leisure, u_0, take the form

$$u_j(v_j) = a_j + b_j\, v_j - c(v_j), \tag{3.1}$$

where a_j and b_j are constants which may be positive or negative and are specific to activity j. The parameter a_j represents the baseline return to activity j when not veiling. It could account for motivations as diverse as the agent's baseline wage rate, the religious social pressure that she might face against her working, or how much she appreciates her colleagues.

The parameter b_j represents how veiling affects this baseline return to activity j. This could be a combination of positive effects, such as alleviating the religious stigma faced by working Muslim women (as in Carvalho 2013); and negative ones, such as triggering discriminations or hostile reactions from peers.

Finally, there is an intrinsic cost $c(\cdot)$ to wearing the veil, which is the same across activities j. Following Carvalho (2013), this cost can for instance be interpreted as physical discomfort. We assume that the cost function $c(\cdot)$ is convex, with $c'(0) = 0$ and $\lim\limits_{v \to 1} c'(v) = \infty$.

With this, we can write the complete utility function of the agent:

$$\begin{aligned} U(t, v_1, v_0) &= t\, u_1 + (1-t)\, u_0 - d(t) \\ &= t\, \big[\underbrace{a_1 + b_1\, v_1 - c(v_1)}_{\text{flow utility from work}}\big] + (1-t)\, \big[\underbrace{a_0 + b_0\, v_0 - c(v_0)}_{\text{flow utility from leisure}}\big] - d(t). \end{aligned} \tag{3.2}$$

The component $d(t)$ represents a disutility of working, which we assume is increasing and convex, with $d'(0) = 0$ and $\lim\limits_{t \to 1} d'(t) = \infty$.

3.4.2 Optimal choices

The problem of the agent is to find the time allocation t and the degrees of veiling v_0 and v_1 which maximize her utility (3.2). This problem can be solved sequentially. First, the agent determines for each activity the degree of veiling which maximizes her flow utility. Second, she chooses her labor supply based on those optimized flow utilities.

Veiling. Call v_j^* the optimal degree of veiling in activity j. If the agent has negative returns to veiling in activity j, i.e. $b_j \leq 0$, then she has no incentive to veil and her optimal degree of veiling is $v_j^* = 0$. Otherwise, if $b_j > 0$, then her optimal degree of veiling maximizes the utility (3.1) that she derives from activity j:

$$c'(v_j^*) = b_j, \tag{3.3}$$

so that v_j^* is positive, increasing in the agent's return to veiling b_j.

Thus, in this model, differences in veiling behavior between work and leisure time are reflective of different returns to veiling for the agent across these activities. We summarize this result in the following lemma, which will become useful later on.

Lemma 1. The agent veils more at work than during leisure time if and only if $b_1 > b_0$.

We now move on to the labor supply problem. In what follows, we denote by $u_j^* = u_j(v_j^*)$ the indirect utility that the agent derives from activity j.

Labor supply. Call t^* the optimal labor supply. If her indirect utility of working is less than that from leisure, i.e. $u_1^* \leq u_0^*$, then the agent has no incentive to work and her optimal labor supply is $t^* = 0$. Otherwise, if $u_1^* > u_0^*$, her optimal labor supply t^* solves the first-order condition

$$d'(t^*) = u_1^* - u_0^*, \tag{3.4}$$

so that t^* is positive, increasing in the indirect utility of working u_1^*, and decreasing in the indirect utility of leisure u_0^*.

In the case whereby the agent has equal returns on veiling for both activities, i.e. $b_1 = b_0$, veiling has no impact on the labor supply decision. Indeed, in this case the agent chooses the same degree of veiling at work and during leisure time: $v_1^* = v_0^*$. Therefore the difference in indirect utilities is simply $u_1^* - u_0^* = a_1 - a_0$, which depends only on the baseline return to each activity. Thus, veiling has an impact on the agent's labor supply decision only if it distorts the returns to work and leisure in distinct ways. In particular, if veiling motives are purely personal and do not interact with the environment, the veiling and labor supply decisions are orthogonal.

3.4.3 Mechanisms

In this section we provide two concrete examples of theoretical mechanisms which may underpin the relationship between the veiling and labor supply decisions, based on the discussion from section 3.2. We use these examples to provide micro-foundations to the generic parameters a_j and b_j that we have introduced in our general framework above.

Religious motives: the Carvalho model. The theoretical mechanism studied by Carvalho (2013) relates to social norms and expectations. In some communities, women may face social pressure to limit their labor supply in order to conform to gender role expectations and maintain social approval. This social pressure can be amplified for religious women who may themselves feel reluctant to integrate into a work environment they perceive as religiously unsafe. Here, veiling can serve a dual purpose as a self-commitment to religious beliefs and as a signal to their community of their religious intentions. As a result, veiling can help mitigate the social cost of women's employment, making it a useful tool for their economic integration.

Let us show that the Carvalho model of veiling is nested in the framework that we have developed above. In this model, the incentive to veil stems from a combination of the individual religiosity of the agent, r, and of the religious social pressure, R. Together, these religious factors determine the penalty that the agent suffers if she engages in religiously-prohibited behavior. This penalty, equal to $-(r+R)$, is both self- and socially-imposed, reflecting personal regret on the one hand, and social stigma on the other hand. It is steeper if the agent herself has higher religiosity, and if religious social pressure is more intense. Note that in this context both r and R can be negative, meaning individual or social approval for religiously-prohibited behavior.

Each activity j, working or leisure,[22] is then characterized by an exogenous risk of engaging in religiously-prohibited behavior, p_j. Crucially, the agent is able to attenuate that risk by veiling (see footnote 9): if she chooses a degree of veiling v_j, then the probability that she engages in religiously-prohibited behavior becomes $p_j(1-v_j)$. Veiling also entails a cost $c(v_j)$ (e.g. physical discomfort).

Finally, there is a material reward m_j associated with each activity j. As a result, the expected utility that the agent derives from activity j is

$$u_j(v_j) = -p_j(1 - v_j)(r + R) - c(v_j) + m_j. \tag{3.5}$$

This utility function is a particular case of equation (3.1), which is obtained by taking $a_j = -p_j(r + R) + m_j$ and $b_j = p_j(r + R)$.

In the Carvalho model, the exogenous risk of engaging in religiously-prohibited behavior is assumed to be greater at work than during leisure time: $p_1 > p_0$. This assumption implies that a woman will always choose a higher degree of veiling at work than during leisure time. Indeed, recall that for the agent to veil at all, she must have positive returns to veiling, i.e. $b_j > 0$. For this to hold here, the agent must have $r+R > 0$, and as a consequence $p_1(r+R) > p_0(r+R)$, i.e. $b_1 > b_0$. Thus, according to our lemma 1, a woman will always veil more at work than during leisure time in the Carvalho model.

Regarding the choice of activity, Carvalho considers a discrete choice $j \in \{0, 1\}$. Again this is a particular case of our framework, obtained by ignoring the disutility of working: $d(t) = 0$. Following our analysis of the labor supply decision, the agent will work if her indirect utility from working is greater than that from leisure, $u_1^* > u_0^*$. This happens if and only if the material reward for working m_1 is large enough.

Here Carvalho shows an interesting result, namely that within a range of values of this material reward m_1, (i) low-religiosity women choose to work, (ii) high-religiosity women choose not to work, and (iii) low-religiosity women veil more than high-religiosity

[22] Carvalho gives a broader interpretation of this decision as a choice between *integration* or *segregation*.

ones. This happens provided that the surrounding population approves of the veil, i.e. $R > 0$, because in this case low-religiosity working women choose to attenuate the social penalty associated with working by veiling. Shofia (2020) finds evidence for this pattern of veiling among women in Indonesia.

Economic motives: labor market discrimination against veiling. We now consider a mechanism which relates to the role of discrimination. Veiled women may face discrimination in the workplace due to negative stereotypes or biases held by their employers or colleagues. This discrimination may limit their opportunities for employment or career advancement, and could ultimately lead them to reduce their labor supply. We predict that women with higher wage potential, who face a greater opportunity cost of unemployment or limited career advancement, will incur higher costs associated with veiling.

This veiling-based discrimination on the labor market can also be accounted for by our general model. Consider a simple consumption–leisure set-up: the agent has quasilinear utility $U(x, t) = x + g(1 - t)$ where x is her consumption of a numeraire good and $1 - t$ is her leisure (the function $g(\cdot)$ is increasing and concave). Consumption is the only source of spending, so that the default budget constraint is $x = w\,t$, where w is the agent's wage rate.

Assume now that discrimination against veiling has a direct negative effect on the agent's wage, such that an agent with wage potential w who also chooses a veiling level v gets the effective wage $w(1-v)$. This assumption broadly reflects that since discrimination typically makes it more difficult for women who wear the veil to secure and keep a job or to advance in their career (cf. section 3.2), the associated cost of veiling should be greater for women with higher earning potential. For instance, we could expect the opportunity cost of job loss or slower career progression to be proportional to one's earning potential. With this assumption, the budget constraint of the agent becomes $x = w(1 - v)t$.

Aside from the cost associated to discriminations, suppose that veiling at work provides a return y to the agent (maybe through the religious incentive mechanism discussed above), and entails an intrinsic cost $c(v)$. In this case, her utility function is

$$U(t, v) = \big[w + (y - w)v - c(v)\big]\, t + g(1 - t). \tag{3.6}$$

Again, this is a particular case of the utility function (3.1), obtained by taking $a_1 = w$, $b_1 = y - w$, $a_0 = b_0 = 0$ (so that $v_0 = 0$), and $d(t) = -g(1 - t)$ (so that $d(\cdot)$ is increasing and convex). This model predicts that women with a higher wage potential w should work more and veil less than those with a lower wage. This result is a direct consequence of veiling having a negative, proportional impact on the agent's effective wage.

The two mechanisms above mostly play in opposite directions. According to the first mechanism, women who are religious or who face religious pressure from their family or community have an incentive to veil at work in order to mitigate the social penalty associated with working. But according to the second mechanism, discrimination at work provides an opposite incentive to unveil at the workplace. In the next section we pool these two motives together in a unified empirical model. We then use data on veiling behaviors and employment of Muslim women to quantify the various effects at hand.

3.5 Empirical analysis

3.5.1 Econometric model

Our econometric specification is derived by pooling together the two motives for (un)veiling described in the previous section, religious and economic. To capture these motives, we focus on three main individual characteristics: individual religiosity r_i, the religious social pressure faced by the individual R_i, and earning potential w_i. We obtain a unified expression for the utility that woman i receives in activity j by choosing the degree of veiling v:[23]

$$u_{ij}(v) = \underbrace{-p_j(1-v)(r_i+R_i)}_{\text{religious motives}} + \underbrace{\mathbf{1}_{\{j=1\}}w_i(1-v)}_{\text{economic motives}} - c(v). \tag{3.7}$$

Our empirical approach relies on measures of the individual characteristics r_i, R_i, and w_i. We use the data and constructed measures that we described in section 3.3.2. Regarding individual religiosity, we use our index measure aggregated from six different survey questions, $\mathsf{Religiosity}_i$. Regarding religious social pressure, we use our index measure of vertical pressure, $\mathsf{VertiReligiousPressure}_i$, and two measures of horizontal pressure, $\mathsf{ShareMaghrebi}_i$ (the share of Maghrebi immigrants in the individual's neighborhood) and $\mathsf{MosqueCapacity}_i$ (the local capacity for Muslim worship). Regarding the earning potential, we use measures of both the individual's educational attainment using her years of schooling, $\mathsf{Education}_i$, and her years of professional experience, $\mathsf{Experience}_i$. To summarize, we use the following proxies for the individual characteristics of woman i:

$$r_i \sim \mathsf{Religiosity}_i \tag{3.8}$$

$$R_i \sim \mathsf{VertiReligiousPressure}_i + \mathsf{ShareMaghrebi}_i + \mathsf{MosqueCapacity}_i \tag{3.9}$$

$$w_i \sim \mathsf{Education}_i + \mathsf{Experience}_i. \tag{3.10}$$

[23] We let the material payoffs m_j in equation (3.5) to be also individual-specific by taking them equal to $\mathbf{1}_{\{j=1\}}w_i(1-v)$, thus combining the religious and economic motives described in section 3.4.3.

Table 3.4: Correspondence between estimated parameters and theoretical model

Explanatory variable	Parameter	Proportional to...	Varies with v	Varies with j
Religiosity variables				
$\mathsf{Religiosity}_i$	β^1_{jv}	$-p_j(1-v)$	$+$	$-$
$\mathsf{VertiReligiousPressure}_i$	β^2_{jv}	$-p_j(1-v)$	$+$	$-$
$\mathsf{ShareMaghrebi}_i$	β^3_{jv}	$-p_j(1-v)$	$+$	$-$
$\mathsf{MosqueCapacity}_i$	β^4_{jv}	$-p_j(1-v)$	$+$	$-$
Economic variables				
$\mathsf{Education}_i$	γ^1_{jv}	$\mathbf{1}_{\{j=1\}}(1-v)$	$-$	$+$
$\mathsf{Experience}_i$	γ^2_{jv}	$\mathbf{1}_{\{j=1\}}(1-v)$	$-$	$+$

Next, we formulate an econometric model informed by the theory which is based on these variables. We use a multinomial logit model to explain the joint decision of activity and veiling, (j, v), with two activity statuses $j \in \{0 = \text{Inactive}, 1 = \text{Active}\}$ and three levels of veiling $v \in \{0 = \text{None}, 1 = \text{Intermediate}, 2 = \text{Always conspicuous}\}$. Due to the low sample sizes in the three intermediate veiling categories, we group discreet-symbol wearing and sometimes wearing a conspicuous symbol into an Intermediate category. Adapting equation (3.7) into an econometric discrete-choice model which uses the proxies described above,[24] the utility for woman i to jointly choose activity j and veiling level v is given by

$$\begin{aligned} u_{ijv} = \alpha_{jv} + \beta^1_{jv} \times \mathsf{Religiosity}_i + \beta^2_{jv} \times \mathsf{VertiReligiousPressure}_i \\ + \beta^3_{jv} \times \mathsf{ShareMaghrebi}_i + \beta^4_{jv} \times \mathsf{MosqueCapacity}_i \\ + \gamma^1_{jv} \times \mathsf{Education}_i + \gamma^2_{jv} \times \mathsf{Experience}_i + X_i' \theta_{jv} + \varepsilon_{ijv}. \end{aligned} \tag{3.11}$$

Here X_i is a set of individual-level controls, and ε_{ijv} is the unobserved part of the utility. The coefficients β_{jv}, γ_{jv} and θ_{jv} are estimated with respect to the baseline $(j, v) = (0, 0)$ (i.e. being inactive and not veiling). We assume that the unobserved components of utility ε_{ijv} are distributed i.i.d. Gumbel, giving rise to a standard multinomial logit model in which the probability for i to choose alternative (j, v) is

$$\frac{\exp u_{ijv}}{\sum_{(j',v')} \exp u_{ij'v'}}. \tag{3.12}$$

3.5.2 Implications of the model

The religious and economic motives channels from the model have separate but clear implications regarding how the estimated parameters should vary with j and v. Table

[24] As mentioned above, moving from continuous to discrete choice is simply achieved by taking $d(t) = 0$.

3.4 outlines the correspondence between the parameters of our estimating equation (3.11) and the theoretical components of the model. To interpret the associated implications, we categorize our explanatory variables into two groups: "religiosity variables," which are associated with the religious motives behind the joint decision of economic participation and veiling and are linked to the β_{jv} parameters (β^1_{jv}, β^2_{jv}, β^3_{jv}, and β^4_{jv}); and "economic variables," which are associated with economic motives and are linked to the γ_{jv} parameters (γ^1_{jv} and γ^2_{jv}).

We describe below the empirical implications of the religious and economic motives of the model for our parameter estimates. Since the same implications apply to β^1_{jv}, β^2_{jv}, β^3_{jv}, and β^4_{jv} on the one hand, and to γ^1_{jv} and γ^2_{jv} on the other hand, we drop the superscripts and make statements about the generic parameters β_{jv} and γ_{jv} instead.

Implication 1. Within activity,

(a) religiosity variables have a milder (negative) impact on utility for women who veil more:
$$\text{at } j \text{ fixed}, \quad \beta_{j0} < \beta_{j1} < \beta_{j2},$$

(b) economic variables have a milder (positive) impact on utility for women who veil more:
$$\text{at } j \text{ fixed}, \quad \gamma_{j0} > \gamma_{j1} > \gamma_{j2}.$$

Implication 2. For a given degree of veiling,

(a) religiosity variables have a stronger (negative) impact on utility for women who participate economically:
$$\text{at } v \text{ fixed}, \quad \beta_{0v} > \beta_{1v},$$

(b) economic variables have a stronger (positive) impact on utility for women who participate economically:
$$\text{at } v \text{ fixed}, \quad \gamma_{0v} < \gamma_{1v}.$$

To interpret these implications of the model, let us focus on the meaning of the parameters to estimate. For instance, the parameter β^1_{jv} indicates how own religiosity impacts the probability of choosing the alternative (j, v). According to the theory, this impact is negative since religiosity implies more limitations on acceptable behavior and a higher intensity of regret. In magnitude, the impact should be milder for women who veil – this is the purpose of veiling in the Carvalho model – hence β^1_{jv} should be increasing in v (Implication 1a). Furthermore, the impact should be greater for working women – because the work environment is more risky than the home environment – hence β^1_{jv} should be decreasing in j (Implication 2a). Similar predictions apply for β^2_{jv}, β^3_{jv} and β^4_{jv}, which relate to the social religious pressure.

Next, the parameter γ^1_{jv} indicates how education impacts the probability of choosing the alternative (j, v). In the model education plays a role by increasing the working wage. Therefore the impact of education should be lower for women who veil more – they have lower expected wage because of discrimination (Implication 1b); and it should be greater for women who work compared to those who do not (Implication 2b). Similar predictions apply to γ^2_{jv}, which relates to professional experience.

Implications 1 and 2 above focus on veiling and economic participation choices independently. However, our main interest is to understand how veiling and economic participation choices interact, and in particular whether religious and economic motives are relevant mechanisms in this interaction. These mechanisms will be captured by studying the signs of double differences in the parameters β_{jv} and γ_{jv}.

First, according to the religious motives mechanism, the religious benefits of veiling are greater for women who integrate economically. This is stated formally as follows:

Implication 3: Religious motives channel. The religious returns on utility to increasing one's degree of veiling are larger for women who participate economically, compared to those who don't:

$$\text{for } v < v' \text{ fixed}, \quad \beta_{1v'} - \beta_{1v} > \beta_{0v'} - \beta_{0v}.$$

Second, according to the economic discrimination mechanism, the economic losses induced by veiling are greater for women who integrate economically. This is stated formally as follows:

Implication 4: Economic discrimination channel. The economic returns to being economically active are smaller for women who veil, compared to those who don't:

$$\text{for } v < v' \text{ fixed}, \quad \gamma_{1v'} - \gamma_{0v'} < \gamma_{1v} - \gamma_{0v}.$$

Having established these empirical implications of the model's different mechanisms, we now turn to the estimation and to testing the model implications 1–4.

3.5.3 Results

Table 3.5 presents the results for the estimation of equation (3.11). Recall that all parameter estimates are relative to the baseline of an inactive woman who never wears religious symbols. This estimation is performed without controls – in Appendix B.1.4 we perform the same exercise while including controls, and observe that results remain sensibly similar.

The parameter estimates suggest two main findings. To ease interpretation, we focus on the predicted marginal effects (panel B in Table 3.5). First, individual religiosity is a strong and significant predictor of changes in veiling behavior, but the same observation does not hold for social pressures. For example, we estimate that a 1 standard deviation increase in individual religiosity decreases the probability of not wearing any religious symbol and being active (resp. inactive) by 17 percentage points (resp. 9 p.p.). On the contrary, it increases the probability of wearing a conspicuous symbol and being active (resp. inactive) by 8 percentage points (resp. 13 p.p.). Social religious pressure (both vertical and horizontal) is also associated with higher degrees of veiling, although most parameter estimates are not significantly different from 0 at the conventional levels. For instance, a 1 s.d. increase in vertical social pressure is associated with an 11 p.p. increase in the probability of wearing a conspicuous symbol and being inactive, while an extra 1 Muslim worship seats per 100 inhabitants is associated with an 8 p.p. increase in the same probability. Overall, both the magnitude of the estimates and their significance level suggest that individual religious motives are the strongest predictors of veiling behavior, above (and conditional on) other social religious pressures.

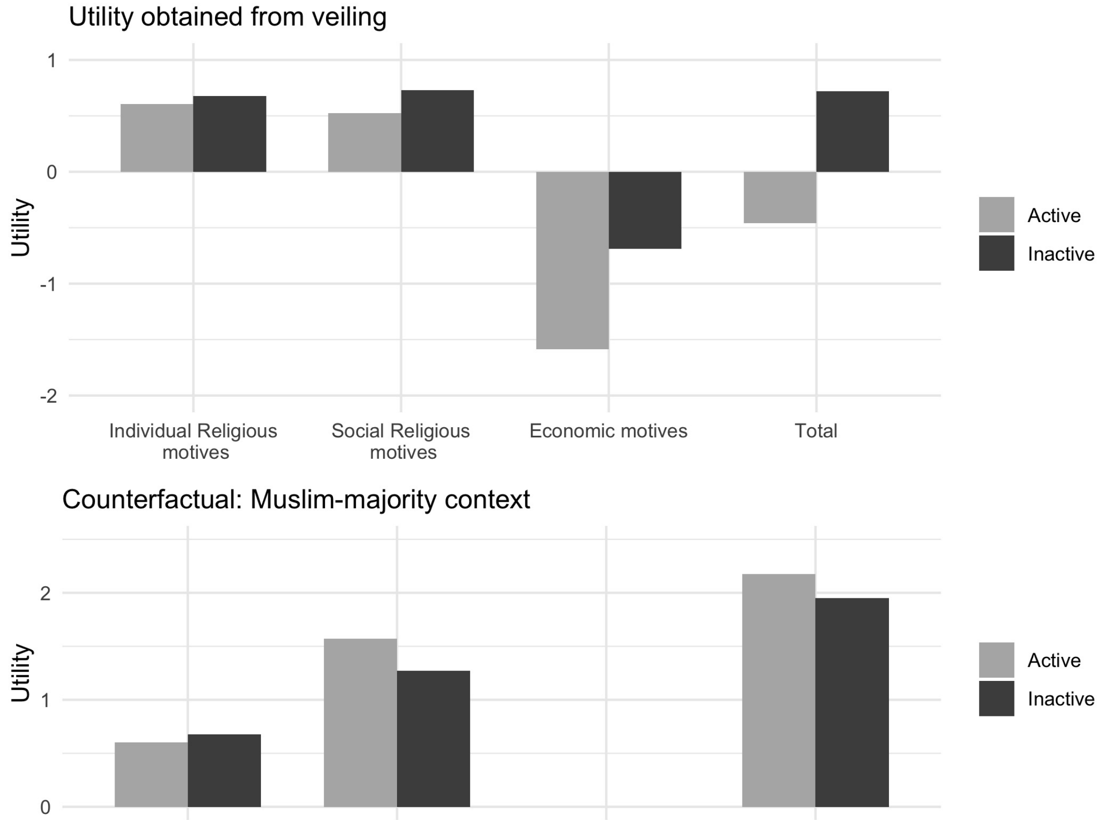
Second, both schooling and work experience substantially increase the probability of being active and decrease the probability of veiling. For instance, an additional year of schooling is associated with a 1.8 p.p. increase (resp. 0.5 p.p.) in the probability of being active and wearing no symbol (resp. wearing a discrete symbol). Interestingly however, these human capital factors are not associated with an increase in the probability of being

Table 3.5: Determinants of joint employment and veiling decision, multinomial logit.

Activity choice (j)	Inactive ($j = 0$)			Active ($j = 1$)		
Veiling choice (v)	None (*baseline*)	Discreet (1)	Conspicuous (2)	None (3)	Discreet (4)	Conspicuous (5)
Panel A: Parameter estimates						
Indiv. religiosity (β^1_{jv})	0	**0.78** (0.21)	**2.26** (0.28)	0.17 (0.16)	**1.00** (0.21)	**2.18** (0.34)
Vert. pressure (β^2_{jv})	0	–1.74 (2.97)	1.54$^+$ (0.81)	0.05 (0.69)	0.56 (0.88)	0.96 (0.94)
Horiz. pressure						
ShareMaghrebi$_i$ (β^3_{jv})	0	4.14 (3.23)	0.68 (1.20)	0.18 (0.85)	0.28 (1.01)	2.14 (1.39)
CapacityMosques$_i$ (β^4_{jv})	0	–0.15 (0.12)	**0.10** (0.04)	0.02 (0.02)	–0.03 (0.02)	0.05 (0.03)
Schooling (γ^1_{jv})	0	0.03$^+$ (0.02)	–0.03$^+$ (0.02)	**0.15** (0.02)	**0.15** (0.02)	**0.06** (0.02)
Work experience (γ^2_{jv})	0	–0.11$^+$ (0.07)	–0.06* (0.03)	**0.12** (0.02)	**0.08** (0.02)	**0.06** (0.02)
Panel B: Average marginal effects						
Indiv. religiosity (β^1_{jv})	**–0.09** (0.01)	0.00 (0.01)	**0.13** (0.02)	**–0.17** (0.02)	**0.05** (0.02)	**0.08** (0.02)
Vert. pressure (β^2_{jv})	–0.03 (0.07)	–0.08 (0.12)	0.11* (0.05)	–0.08 (0.11)	0.04 (0.08)	0.03 (0.04)
Horiz. pressure						
ShareMaghrebi$_i$ (β^3_{jv})	–0.08 (0.09)	0.13 (0.14)	–0.00 (0.08)	–0.12 (0.14)	–0.03 (0.09)	0.09 (0.07)
CapacityMosques$_i$ ($\beta^4_{jv} \times 10$)	–0.02 (0.02)	–0.06 (0.05)	**0.08** (0.03)	0.03 (0.04)	–0.05* (0.02)	0.02 (0.01)
Schooling ($\gamma^1_{jv} \times 10$)	**–0.10** (0.01)	**–0.02** (0.01)	**–0.09** (0.01)	**0.18** (0.02)	**0.05** (0.01)	–0.02$^+$ (0.01)
Work experience ($\gamma^2_{jv} \times 10$)	**–0.07** (0.02)	–0.06$^+$ (0.04)	**–0.10** (0.02)	**0.21** (0.02)	0.01 (0.01)	0.01 (0.01)
Observations	2802					
Sampling weights	✓					
Pseudo R^2	0.159					

Note: This table reports estimates of the parameters of the econometric model (3.11). The baseline category is the choice of inactivity and not wearing any religious symbol. Individual religiosity and vertical religious pressures are measured as indices (with mean zero and variance 1) constructed from multiple proxies available in the TeO data (see Appendix B.1.1 for details). ShareMaghrebi$_i$ is the proportion of the local population that is of Maghrebi origin. CapacityMosques$_i$ is the estimated capacity in Muslim places of worship in the area of residence. Robust standard errors in parentheses. Point estimates in bold are significant at the 1% level ($p < 0.01$), * $p < 0.05$, $^+$ $p < 0.1$.

Figure 3.2: Utility obtained from veiling



Note: Difference of utility between choosing $v = 0$ and $v = 2$ according to the estimates of Table 3.5, and based on an 'average' woman in our sample (cf. footnote 25). The lower panel is obtained by counterfactual, shutting down the economic discrimination channel and modifying some environmental characteristics of this average woman to reflect a Muslim-majority environment (cf. footnote 26).

active while wearing a conspicuous symbol. This result might suggest that veiling at work offsets the benefits of human capital on economic activity, an expected consequence of the labor-market discrimination channel.

We illustrate these results in Figure 3.2 by plotting the utility obtained by veiling for an 'average' woman in our sample, according to our estimates.[25] We observe that this average woman has a disincentive to veil overall if she is active, which is a consequence of the economic motives being stronger than the religious ones. On the contrary, an inactive woman has an incentive to veil, because she is less affected by economic motives.

We then compute the same utilities in a counterfactual, Muslim-majority environment in which there is no economic discrimination against wearing the veil at work.[26] In this

[25] We set the following values for this 'average' Muslim woman: Individual Religiosity: 0.1, Vertical Religious Pressure: 0.1, Local share of Maghrebi immigrants: .10, Muslim worship seats per thousand inhabitants: 2, Schooling: 15 years, Work experience: 4 years. One can compare those values with the summary statistics of Tables B.1 and B.2 to verify that this roughly corresponds to an average Muslim woman in our sample.

[26] To compute this counterfactual, we shut down the economic discrimination channel, and set the share of local Maghrebi immigrants to 0.6 (instead of 0.1) and the number of worship seats to 4 (instead

case, we see that active and inactive women have somewhat equivalent incentives to veil, which sharply contrasts with our findings in the French setting. Active women benefit slightly more from veiling than inactive ones overall, a finding which is consistent with the religious channel of the Carvalho (2013) model and with the evidence from Shofia (2020) on Indonesia, although the difference here is small.

In the rest of this section, we verify these results formally using the tests formulated in Implications 1–4. Detailed results for these tests are available in Appendix B.1.5.

Baseline implications. Implications 1 and 2 concern the direction of variation for the coefficients β_{jv} and γ_{jv}, respectively with the veiling level v and the activity j. Tests of these implications should indicate whether our joint outcomes react to our predictors in the direction expected by the model.

Implication 1. Our first model implication concerns the relationship of our predictor variables with veiling behavior, within a given economic activity. Consider for instance our measure of individual religiosity. We can see clearly from Table 3.5 that individual religiosity is associated with an increase in the degree of veiling, both for active and inactive women. Indeed, at activity j fixed, our estimates for β^1_{jv} increase across veiling levels v, indicating that higher measures of individual religiosity are associated with an increased propensity to wear the veil.

To verify this formally, we conduct hypothesis tests of the form $\beta^1_{jv'} - \beta^1_{jv} > 0$ for the different possible combinations of j, v and v' such that $v' > v$. (We present the detailed results in Figure B.1, Appendix B.1.5.) In this case, we find that Implication 1 holds at the 95% confidence level for all possible combinations of v and v', thus confirming the positive association between individual religiosity and veiling.

We then perform similar tests of Implication 1 for our five other main predictors. Most of our point estimates for the tests associated with the different predictors agree with Implication 1, although several tests do not reach statistical significance. Regarding vertical social pressure, five estimates out of six fall in the predicted region. For our first measure of horizontal pressure, i.e. the percentage of people from Maghrebi origin in the neighborhood, again five out of six point estimates fall in the predicted region. For our second measure of horizontal pressure, i.e. the local number of seats in religious facilities per 1000 inhabitants, four out of six point estimates fall in the predicted region, with two of those being significant at the 95% confidence level. Finally, both for our work experience variable and for our schooling variable, five out of six point estimates fall in the predicted region, with three of those being significantly different from zero.

of 2).

Furthermore, if we ignore the 'discreet symbols' veiling category for which we have few observations, then our point estimates systematically fall in the half-space predicted by the model, with a majority of the tests yielding statistically significant predictions.

Put together, we interpret these results as providing partial evidence for Implication 1. Although a majority of the tests do not hold at the 95% level, the overall pattern of point estimates falling in the predicted region suggests some validity for the statement of Implication 1. Notably, statistical power might be an issue here, as we observed by discarding the estimates linked to the 'discreet symbols' category, for which we have few observations: doing so decreases the rejection rate for our tests. Overall, the tests of Implication 1 thus confirm that our religiosity variables are broadly associated with an increased propensity to veil, while our economic variables are associated with a decreased propensity to veil.

Implication 2. Our second implication concerns the relationship of our predictor variables with economic activity, holding the degree of veiling fixed. As we did with Implication 1, we perform tests of Implication 2 for our six main predictors, the results of which are presented in Figure B.2. First, regarding our four religiosity variables, there does not seem to be much support for Implication 2. There is no systematic pattern for point estimates as we observed for Implication 1, and all tests fail at the 95% confidence level. Therefore, we do not find any evidence for our religious variables being associated with an increased or decreased propensity to be economically active.

On the contrary, we find that our economic variables are strongly associated with economic activity. Indeed, Implication 2 holds for both our work experience and schooling variables. This indicates a strong positive association between these economic variables and the propensity to be economically active.

Since we do not find that religiosity variables are strongly associated with the propensity to be economically active, the 'religious motives channel' is already undermined by the tests of Implication 2. This is because this channel predicts that, when holding the degree of veiling constant, women who are more religious or who face more external religious pressure should be less economically active. However, this is not what we find here: our results suggest that the religiosity variables do not have a direct effect on economic participation, but only an indirect one through the practice of veiling. We discuss this further with the test of Implication 3 below.

Mechanisms. We now move on to the tests of Implications 3 and 4, which are more directly related to the two mechanisms that we highlighted above: the religious motives channel, and the economic discrimination channel.

Implication 3. Our third implication can be interpreted as a formal test for the religious motives channel, since it examines whether veiling has higher religious returns for women who are economically active, compared to those who are not. Our results for these tests are presented in Figure B.3. In this case, neither test significance nor point estimates suggest that the formal statement of Implication 3 holds. As such, we do not find evidence for this mechanism.

This result is in line with those of the tests for Implication 2, which already suggested an absence of association between our religious variables and economic participation among Muslim women. Taken together, these results point towards religious motives having an effect on economic participation only through the practice of veiling. This supports the idea that the negative correlation between veiling and economic participation that we observed in the descriptive analysis may be mostly due to veiling having a cost on the labor market, as opposed to religious women having different preferences from non-religious women regarding economic participation.

Implication 4. Finally, our fourth implication can be interpreted as a formal test for the economic discrimination channel, by examining whether economic participation has higher returns for women who do not veil, compared to those who do. Results are presented in Figure B.4.

Regarding our first economic variable, work experience, we do not find support for the statement of Implication 4: the tests reject the hypothesis at the 95% confidence level, and there is no pattern of point estimates mostly belonging to the predicted region. This is perhaps because, on average, work experience does not substantially differs by veiling status (see Table B.1) since veiled women are older and thus had more time to accumulate experience. However, we find some support in the tests associated with our second economic variable, schooling, which most women in our sample had time to complete. In this case, all point estimates fall within the predicted region. Furthermore, the test which ignores the 'discreet symbols' category suggest statistically significant differences (although those which involve these categories do not hold at this level).

This second result offers support the economic discrimination channel: higher-educated women are less likely to integrate economically if they veil, even if we hold religiosity variables constant. In other words, the utility returns on schooling are lower for women who veil compared to those who do not. We have seen in our discussion of Implication 2 that this seems to be unrelated to an underlying preference towards economic participation linked with individual religiosity or social religious pressures. Therefore, this result seems to support the idea that there is an economic cost to veiling, in the sense that veiled women face weaker economic opportunities than those who do not veil.

To sum up, our results suggest that the interaction between the decision to veil and that of economic participation is mostly driven by economic concerns. First, both religious motives and economic ones play important roles in the decision to veil. Second, while economic motives are strong drivers of economic participation, the same is not true for religious motives, suggesting that the veil itself (and not underlying religious preferences) is linked to decreased economic participation. Third, non-veiled women seem to enjoy higher economic returns on their education compared to veiled women (holding individual religiosity and social religious pressures fixed), as evidenced by their higher propensity to be economically active.

Overall, those results suggest that the religious mechanism suggested by Carvalho (2013) cannot fully explain veiling and economic participation patterns in France. Instead, the interaction between veiling and the economic incentives to economic participation, such as the discrimination against veiled women on the labor market, seems to play an important role in this context. Furthermore, and of particular importance for the French debate, we note that individual religious motives turn out to be at least as important as communitarian influences in the decision to veil.

3.6 Conclusion

Theoretical and empirical studies of veiling in economics have so far mainly focused on Muslim-majority countries, perhaps because of the paucity of data on veiling in Western countries. With the rising immigration flows of Muslims to secular countries, getting a better understanding of why women veil is nonetheless crucial as many countries, of which France is maybe the most emblematic, limit the expression of religious faith in public.

In this paper, we tackle this question using rare rich observational data on Muslim women in France. The richness of the data notably allows us to distinguish between private and communitarian incentives to veil. We first document that in France, wearing conspicuous religious symbols is associated with a much lower economic integration for Muslim women. The magnitude of this relationship is large, comparable to having a child less than 4 years old for instance. Second, we find that, among the main incentives for veiling highlighted in the economic literature, the wearing of conspicuous symbols appears to be strongly driven by private religious motivations. Third, we find that the joint decision to veil and being economically active can be mostly explained by economic (dis)incentives. Our results thus suggest that the veiling mechanism proposed by Carvalho (2013) and evidenced in the context of Indonesia by Shofia (2020) may be second-order in a non-Muslim-majority country such as France. Instead, when choosing whether to work and to wear the veil, Muslim women seem to be more sensitive to incentives related to how veiling impacts their economic opportunities.

Because they underline the role of private religious motives instead of community pressure ones, our results question the rhetoric often used to justify policies restricting the wearing of religious symbols in France. In the media and in political spheres, journalists and politicians almost always defend veiling restrictions on the basis that Muslim women are being forced to veil by their husband and community. If these claims were true, it is believed that secular policies could have the potential to "free" Muslim women from religious pressures and promote gender equality (e.g. Maurin and Navarrete-H., 2023). Actually, even in this case, Carvalho (2013) shows that banning the veil in public spaces might lead to *more* segregation because women would lose the ability to signal their piety to their community. However, consistent with existing evidence from qualitative interviews with Muslim women, we find that the main incentives for veiling appear to be private. In other words, Muslim women who veil do so for personal reasons linked to their own beliefs, first and foremost. Therefore, further restricting the wearing of conspicuous religious symbols is likely to lead to even poorer integration of Muslim women if these private benefits are high and discreet symbols are imperfect substitutes. Our complementary analysis of the Turkish case, a country which also imposed secular constraints in the public sphere, is consistent with this argument.

Furthermore, our results call attention to the importance of the discriminations that women who wear the veil face on the labor market. For instance, hiring discriminations against people who signal their Muslim affiliation were already documented by Valfort (2020). Because we find that individual religiosity and other religious factors seem to be associated with the decision to be economically active mainly through the act of veiling, a possible interpretation is that women who veil are less economically active not because of underlying preferences linked with their religiosity, but rather because the veil represents an obstacle to economic participation.

Our empirical approach in this paper is descriptive and should not be interpreted as causal. Still, our results suggest that veiling in France entails significant costs to economic integration, and is driven by private incentives before social ones. Given the importance of better integrating Muslim populations in developed countries, future work could provide more robust assessments of the patterns uncovered in this paper. For example, if larger databases on Muslim women become available, one could evaluate the effect of external shocks to the local religious composition, such as exogenous migration waves, on veiling patterns. We finally note that data limitations inherent to studies of this type call for more initiatives like the TeO survey to better document the experiences of minority populations in a context of increasing global migrations.

Appendix B

Supplementary Material for Chapter 3

B.1 Data and additional results

B.1.1 Measurement of individual religiosity and communitarian pressures

The TeO dataset contains rich information on respondents' religious life. We first describe the variables we use to proxy for individual religiosity, vertical religious influence (from parents), and horizontal pressures (from Muslim peers). We then detail how we combine those multiple measures into meaningful indices through a measurement system.

Individual religiosity. In TeO1, we measure individual religiosity using survey questions on the frequency of attendance of religious ceremonies, the self-reported importance of religion in the respondent's life, whether she uses her religion to self-identify, the respect of religious dietary restrictions, and religious marriage. In TeO2, an additional variable is available, that is, the frequency of praying. We list details of these variables below:

Vertical religious pressure. We measure vertical religious pressures using two variables, namely the self-reported importance of religion in the respondent's education and religious name-giving.

Horizontal religious pressure. We measure horizontal religious pressures (from Muslim peers) using two variables, namely the share of Maghrebi immigrants in the respondent's neighborhood (IRIS) and the local capacity in Muslim places of worship. In TeO1, the share of Maghrebi immigrants is reported in deciles of the distribution across France. We select the middle point of each bin, except for the extremes – zero or above

Variable name	Values	Question	Type
attendance of religious ceremonies	never; for familial ceremonies only; for religious feasts only; one or twice a month; weekly	"How often do you attend religious ceremonies?"	ordinal
importance of religion in respondent's life	no importance; a little; quite important; very important	"What importance do you give to religion in your life today?"	ordinal
uses religion to self-identify	yes; no	"Among the following characteristics, which ones define you best? [...] Your religion?"	indicator
respect of dietary restrictions	never; sometimes; always; none (coded as a dummy if "always")	"In your daily life, do you respect your religion's dietary restrictions?"	indicator
religious marriage	yes; no	"Did you and your husband do a religious wedding?"	indicator

Variable name	Values	Question	Type
importance of religion in education	no importance; a little important; quite important; very important	"What importance did religion have in the education you received in your family?"	ordinal
religious first name	yes; no	constructed by authors using respondent's first name	indicator

40%, where we set the value of the variable to 0 and 0.4 respectively. Our second proxy of local Muslim presence is the estimated capacity (by the Muslim association who produced the inventory) in Muslim places of worship at the local level. In TeO1, this is measured at the *commune* (municipal) level of residence for all French cities except Paris, Lyon, and Marseille, for which we observe the *arrondissement*.

Measurement system. For the first two concepts above, since there is no natural way to combine the ordinal and indicator variables into meaningful indices, we formulate a measurement system. We are interested in two latent variables, *individual religiosity* and *vertical religious pressure*, which we assume load into their respective proxies listed above. We interpret those proxies as noisy measures of the associated unobserved, underlying concept. Denote by Z and W the vectors of proxies for individual religiosity and for vertical pressure respectively. We assume ordinal relationships between measures $\{Z, W\}$ and underlying factors $\text{IndivReligiosity}_i$ and VertPressure_i:

$$Z_{i,j} = \mu^z_{1,j} + \lambda^z_j \text{ IndivReligiosity}_i + \varepsilon^z_{i,j} \tag{B.1}$$

$$W_{i,j} = \mu^w_j + \lambda^w_j \text{ VertPressure}_i + \varepsilon^w_{i,j} \tag{B.2}$$

where ε are measurement errors assumed to be i.i.d. and to follow a logistic distribution. As the latent factors do not have a natural scale or location, to simplify interpretations, we normalize the means of $\text{IndivReligiosity}_i$ and VertPressure_i to zero, and their variances to one. We then predict the latent factors for each individual by calculating their empirical Bayes means (Skrondal and Rabe-Hesketh, 2009).

B.1.2 Summary statistics (TeO)

We present some novel summary statistics of Muslim women by veiling status in Table B.1. We distinguish between four categories for the wearing of religious symbols, which depend on (1) whether the symbol is "discreet" or "conspicuous", and (2) whether it is worn "sometimes" or "always". Since there is very little variation in the number of symbols worn (most women report only wearing one), we do not use that information and focus on the extensive margin. Along with the outside option of not wearing any symbol, we thus compare five veiling levels. In terms of the theoretical model, we interpret the veiling level (v) as being increasing in the following order: no symbol ($v = 0$), sometimes worn, and always worn. Overall, Muslim women wearing conspicuous religious symbols differ from other Muslim women in many respects. For example, they are on average older, have more children, and are more likely to live in a couple. Moreover, while most Muslim women wearing a discreet symbol are second-generation immigrants, the vast majority of women who wear a conspicuous symbol are first-generation immigrants. In line with a potential learning of the French social norms by women wearing discreet symbols compared to those wearing the veil, the former are more likely to report being discriminated against for non-religious reasons, not to trust the French institutions, and to believe that racism is widespread in France.

In Table B.1, we report summary statistics of all religion-related variables by veiling status. As expected, as we move toward "higher" veiling status, individuals report higher degrees of religiosity and live in more religious environments. For example, 79% of women who always wear conspicuous symbols report that religion is very important in their life, while less than half of women not wearing a religious symbol do so. Women wearing discreet symbols appear to be moderately religious, but still report higher degrees of religiosity than women without any symbol. Women who wear conspicuous symbols also seem to live in more religious environments: they are more likely to have a Muslim partner and to report that most of their friends are Muslims. Moreover, they live in communes (and neighborhoods) populated by a larger Muslim community (proxied by Maghrebi immigrants and Muslim places of worship). Veiled women also seem to be subject to stronger parental religious pressures. They are significantly more likely to report that religion was very important in their education and to be given a religious first name. In short, all of the core potential mechanisms mentioned so far display some association with veiling behavior in the expected directions (see Table 3.4).

The main fact that motivates the first part of our analysis is that women wearing religious symbols, in particular those who always do so, have much poorer labor-market and schooling outcomes than the rest of the sample. Indeed, women who always wear conspicuous religious symbols are much less economically active on average. Our measure of economic activity is the activity rate, that is, whether the woman is either working, studying, or looking for a job (unemployed) at the time of the survey. While less than 20% of women not wearing conspicuous symbols are inactive at the time of the interview, this proportion increases to 30% for women who sometimes wear a conspicuous symbol and up to 64% for women who always do. Moreover, while 20% of women not wearing a symbol report having never worked in their life, almost half of women who always veil indicate having never entered the labor force. In terms of schooling outcomes, Muslim women who wear a conspicuous symbol are less likely to have any schooling degree. They have completed, on average, 2 to 7 fewer years of schooling than Muslim women who wear discreet symbols or none. Overall, the data suggests that wearing the veil seems to be strongly associated with a decline in economic integration, but this correlation may be due to many other factors over which veiled women differ from other Muslim women. We therefore provide a more thorough regression analysis of this pattern in our empirical approach.

B.1.3 Analysis of panel data

Exploiting the respondents' employment history available in the TeO data, we construct a retrospective panel dataset of economic activity to test the robustness of our results to the timing of the survey. We restrict the sample to adults, meaning that we remove observations for which an individual is aged less than 18 years old. This sample selection is made because it can be plausibly assumed that the veiling decision, on average, is made before adulthood.[1] We estimate random effects models using this data and report results in Table B.3. In column (1), we regress the activity rate on veiling status and year fixed effects. In columns (2) and (3), we include, in turn, time-varying observables and time-invariant controls. The time-invariant controls are all covariates and dummies included in the cross-sectional analysis that are not likely to have changed over time (at least after age 18). These include the mother's and father's religion (Muslim or other), whether the individual has an Arabic-sounding name, attendance of religious ceremonies (proxy for religiosity), self-reported feelings of French identity, the importance of religion in the respondent's education, birthplace dummies, and a set of survey fixed effects. In these regressions, we cluster standard errors at the individual level to account for

[1] In the case of the Islamic veil, ethnographic evidence shows that the decision is usually made between the age of reaching puberty and around 20 years old (Gaspard and Khosrokhavar, 1995). According to Islamic prescriptions, girls are supposed to dress modestly (including covering their hair) when reaching puberty so as to reduce men's temptation. In reality, in France, many adolescents or young women choose to veil a few years after reaching puberty, that is, around adulthood. We also verify that our results are not sensitive to the 18 years old threshold. In a robustness check, we restrict the sample to individuals aged at least 25 years old and find similar results.

Table B.1: Summary statistics by veiling status, Muslim women

Veiling status:	No symbol	Sometimes discreet	Always discreet	Sometimes consp.	Always consp.	Diff (C-D)
Demographics						
Age in 2008	35.55	28.40	25.06	35.94	36.00	8.62***
First-gen. immigrant	0.61	0.24	0.51	0.68	0.78	
Second-gen. immigrant	0.39	0.66	0.49	0.32	0.22	-0.46***
Number of children	1.78	1.11	0.63	2.26	2.79	1.88***
Lives in a couple	0.59	0.49	0.48	0.68	0.74	0.34***
Not a French speaker	0.07	0.02	0.01	0.14	0.32	0.26***
Labour-force status in 2008						
Employed	0.54	0.43	0.36	0.44	0.22	-0.17***
Unemployed	0.18	0.23	0.27	0.12	0.09	-0.10***
Inactive	0.19	0.15	0.24	0.30	0.65	0.44***
Student	0.09	0.20	0.13	0.14	0.03	-0.15***
Has never worked	0.19	0.29	0.48	0.31	0.50	0.16***
Schooling attainment and work experience						
Completed high school	0.78	0.85	0.58	0.68	0.61	-0.22***
Higher education degree	0.22	0.24	0.10	0.20	0.19	-0.06**
Years of schooling	15.30	17.41	15.69	12.86	11.11	-6.09***
Years of work experience	7.06	3.93	3.44	5.75	2.66	-0.61*
Social life and integration						
Participates in household's food shopping	0.49	0.39	0.34	0.59	0.69	0.30***
Often meets her family	0.89	0.89	0.89	0.89	0.93	0.03
Often meets her friends	0.88	0.90	0.94	0.87	0.90	-0.03
Meets with neighbors	0.41	0.45	0.50	0.52	0.62	0.13***
Meets with work colleagues[1]	0.32	0.36	0.33	0.22	0.11	-0.11**
Visits some recreation sites	0.67	0.78	0.76	0.53	0.42	-0.32***
Refuses to visit some recreation sites	0.09	0.12	0.15	0.06	0.04	-0.08***
Belongs to an association	0.17	0.18	0.21	0.18	0.12	-0.07**
Brings the children to school most of the time[1]	0.78	0.88	0.78	0.83	0.82	-0.02
Opinions on discrimination and French institutions						
Victim of racism due to religion	0.36	0.50	0.56	0.51	0.66	0.09***
Victim of racism due to origins	0.79	0.84	0.84	0.83	0.75	-0.07**
Victim of discrimination in past 5 years	0.28	0.41	0.34	0.40	0.28	-0.07**
Believes that racism happens often in France	0.49	0.60	0.68	0.45	0.38	-0.25***
Does not trust the French justice system	0.23	0.28	0.32	0.20	0.20	-0.10***
Does not trust the French police	0.29	0.40	0.50	0.28	0.25	-0.19***
Does not trust the French school	0.07	0.10	0.15	0.07	0.06	-0.06***
ID controlled by the police at least once	0.18	0.28	0.31	0.28	0.12	-0.14***
Observations	2,017	166	151	148	516	

Note: The data source is the Trajectories and Origins (TeO) dataset of 2008. Veiling status is measured using the respondents' answers to the wearing of religious symbols. We distinguish four categories depending on (1) whether the symbol is "discreet" or "conspicuous", and (2) whether it is worn "sometimes" or "always". In the last column, we report differences in means between individuals wearing conspicuous and those wearing discreet symbols where we pooled individuals along the first dimension (salience) as well as significance levels of those differences. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[1] Meeting with work colleagues is conditional on employment and bringing children to school is conditional on having children. Thus, these variables are measured over restricted samples.

Table B.2: Religious environment and religiosity by veiling status, Muslim women

Veiling status:	No symbol	Sometimes discreet	Always discreet	Sometimes consp.	Always consp.	Diff (C-D)
Religious environment						
Muslim partner	0.56	0.49	0.53	0.74	0.76	0.33***
Muslim father	0.94	0.95	0.68	0.96	0.98	0.05***
Muslim mother	0.94	0.95	0.75	0.99	0.97	0.06***
At least half of friends are Muslims	0.719	0.783	0.675	0.838	0.919	0.17***
At least half of work colleagues are immigrants[1]	0.43	0.37	0.42	0.46	0.55	0.14**
Had conflicts on religion with parents when 18 years old	0.15	0.17	0.19	0.07	0.11	-0.04*
Individual religiosity						
Importance of religion in one's life						
A little important	0.18	0.11	0.05	0.04	0.04	-0.06***
Quite important	0.31	0.30	0.29	0.33	0.15	-0.14***
Very important	0.47	0.58	0.64	0.63	0.81	0.20***
Attends religious ceremonies						
Familial ceremonies only	0.290	0.329	0.247	0.284	0.198	-0.07**
Religious feasts only	0.216	0.348	0.273	0.372	0.283	-0.01
Once or twice a month	0.036	0.061	0.047	0.088	0.099	0.05***
At least once a week	0.027	0.006	0.047	0.088	0.155	0.11***
Other indicators of religiosity						
Always respects the religious dietary restrictions	0.826	0.898	0.901	0.946	0.975	0.07***
Religious marriage	0.390	0.307	0.298	0.527	0.657	0.33***
Share of children with a religious first name[1]	0.030	0.013	0.096	0.172	0.186	0.06***
Uses her religion to self-identify	0.13	0.21	0.12	0.25	0.22	0.05*
Parental influence and communitarian religious presence						
Religious first name	0.09	0.08	0.04	0.18	0.13	0.05***
Local Front National vote share	0.098	0.100	0.099	0.102	0.106	0.005***
Importance of religion in education received						
A little important	0.173	0.115	0.139	0.068	0.074	-0.06***
Quite important	0.303	0.265	0.231	0.225	0.198	-0.05
Very important	0.468	0.566	0.543	0.674	0.708	0.14***
Percentage of Maghrebi immigrants in IRIS of residence						
(5.9%, 10.7%]	0.086	0.066	0.093	0.095	0.045	-0.02
(10.7%, 16.7%]	0.150	0.199	0.166	0.088	0.130	-0.06***
(16.7%, 27.3%]	0.289	0.295	0.265	0.304	0.275	0.00
More than 27.3%	0.418	0.398	0.417	0.473	0.510	0.09***
Presence of Muslim places of worship in commune (or arrond.)						
Places of worship (/1000 inh.)	0.053	0.047	0.050	0.055	0.069	0.01***
Capacity in a place of worship (/1000 inh.)	12.249	8.882	11.498	12.582	17.243	5.42***
Capacity for women in a place of worship (/1000 inh.)	2.061	1.600	2.197	2.041	3.095	0.94***
Observations	2,017	166	151	148	516	

Note: The data source is the Trajectories and Origins (TeO) dataset of 2008. Veiling status is measured using the respondents' answers to the wearing of religious symbols. We distinguish four categories depending on whether (1) the symbol is "discreet" or "conspicuous", and (2) it is worn "sometimes" or "always". In the last column, we report differences in means between individuals wearing conspicuous and those wearing discreet symbols where we pooled individuals along the first dimension (salience) as well as significance levels of those differences. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

[1] The composition of work colleagues is conditional on employment and names of the respondents' children is conditional on having children. Thus, these variables are measured over restricted samples.

serial correlation. However, we cannot include individual fixed effects because we do not have panel data on veiling. We thus implicitly assume that the veiling decision is permanent, which we argue is a reasonable assumption because "unveiling" is a relatively rare phenomenon in France.[2]

The results from these regressions overall confirm the findings obtained in the cross-sectional analysis. Indeed, the wearing of a conspicuous symbol is associated with a significant decline in economic participation. Once more, the estimated effect is much stronger when the individual always wears the symbol. The estimates are smaller in magnitude then those obtained in the cross-section, but are still statistically and economically significant. The results indicate that women who always veil are 20 percentage points less likely to be active than women not wearing any religious symbol in a given year. Other important determinants of the activity rate, as expected, are the number of young children, marital status, and the number of years of schooling. These results suggest that those obtained in section 3.3.3 are not merely due to the timing of the survey and portray a more general phenomenon about Muslim women in France.

[2]Two surveys conducted over (rather small) representative samples of the French Muslim population suggest that between 8 and 10 percent of women of Muslim faith declare having worn the veil in the past and are no longer doing so (IFOP 2019, Institut Montaigne, 2016). Out of the total number of women not currently wearing the veil, this figure represents between 12.3% and 14.7%. Since here, we have both untreated individuals to which we assign treatment and treated individuals whom we assign to the untreated group, it is not clear in which direction this measurement error biases our estimates. In light of those issues, we treat this analysis simply as a robustness check of our main results obtained in the cross-section.

Table B.3: Effect of veiling on economic participation of adult Muslim women, retrospective panel data

Dep. variable: activity dummy	(1)	(2)	(3)	25 y.o. +
Veiling status				
Sometimes discrete	0.102***	0.002	0.006	-0.013
	(0.026)	(0.020)	(0.020)	(0.038)
Always discrete	0.077*	-0.031	-0.024	-0.050
	(0.030)	(0.021)	(0.021)	(0.039)
Sometimes conspicuous	-0.120***	-0.052*	-0.039	-0.046
	(0.035)	(0.026)	(0.026)	(0.036)
Always conspicuous	-0.365***	-0.216***	-0.176***	-0.203***
	(0.020)	(0.017)	(0.017)	(0.023)
Educational attainment				
Years of schooling in France		0.012***	0.010***	0.009***
		(0.001)	(0.001)	(0.001)
Years of schooling abroad		0.001	0.001	0.000
		(0.001)	(0.001)	(0.001)
Time-varying demographics				
Age		-0.010*	-0.008	0.020*
		(0.004)	(0.005)	(0.008)
Age squared		0.000	0.000	-0.000**
		(0.000)	(0.000)	(0.000)
Number of children		-0.007	-0.007	-0.022***
		(0.005)	(0.005)	(0.006)
Number of children below age 4		-0.089***	-0.089***	-0.066***
		(0.006)	(0.006)	(0.007)
Married		-0.147***	-0.139***	-0.068***
		(0.014)	(0.014)	(0.019)
Constant	0.629***	0.756***	0.928***	0.484*
	(0.019)	(0.074)	(0.108)	(0.234)
Time-invariant controls	N	N	Y	Y
Year fixed effects	Y	Y	Y	Y
Number of individuals	2,790	2,790	2,790	2,053
Total observations (N X Years)	37680	37680	37680	25354
R^2	0.124	0.394	0.405	0.345

This table shows the results of random-effects regression models of the economic activity dummy on the veiling status and other covariates in the retrospective panel dataset. Standard errors clustered at the individual level in parentheses. The estimation sample is restricted to adult Muslim women with no missing covariates and to time periods during which the individual was in France. In the last column, we estimate the specification in column (3) on the restricted sample of individuals aged at least 25 years old. Level of significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B.1.4 Multi-logit regressions with controls

In Table B.4 we present results similar to those of Table 3.5, but including additional controls.

Table B.4: Determinants of joint employment and veiling decision, multinomial logit.

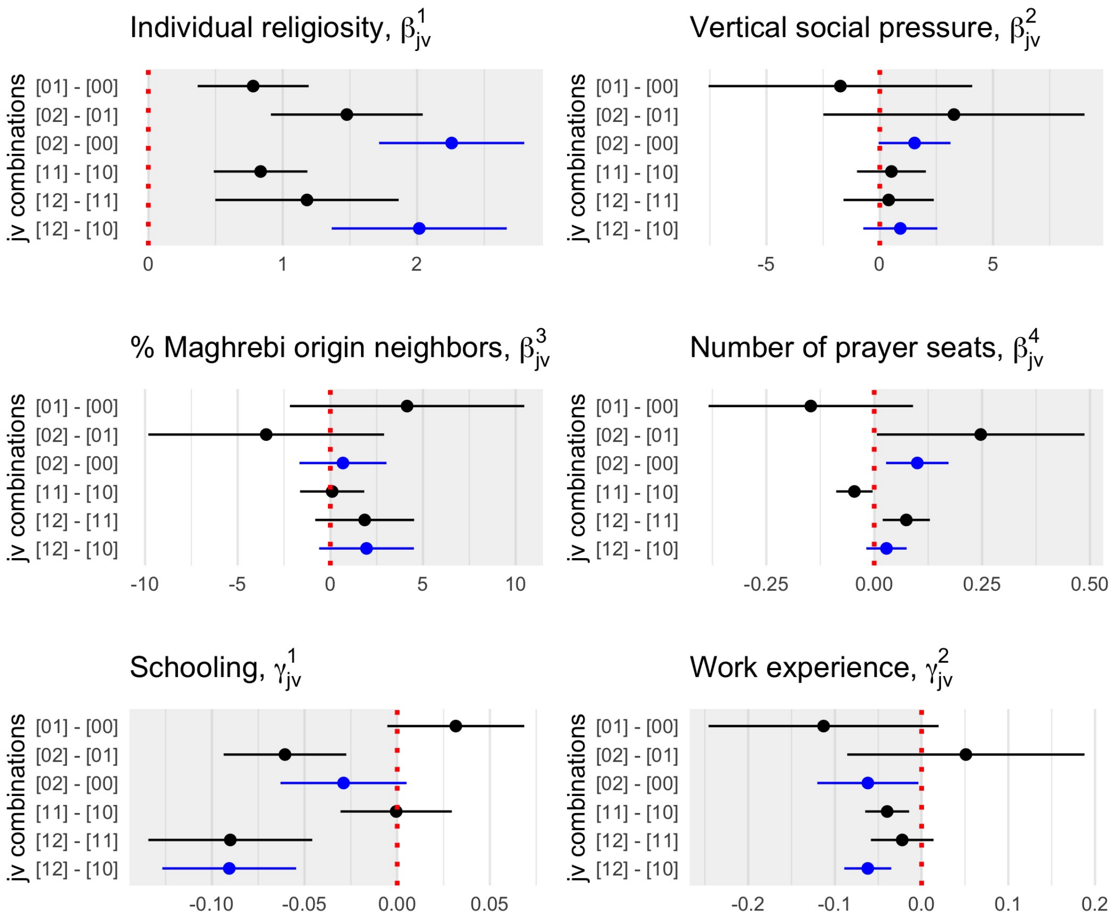
Activity choice (j)	Inactive ($j = 0$)			Active ($j = 1$)		
Veiling choice (v)	None (*baseline*)	Discreet (1)	Conspicuous (2)	None (3)	Discreet (4)	Conspicuous (5)
Indiv. religiosity (β^1_{jv})	0	0.42* (0.24)	2.13*** (0.26)	0.19 (0.18)	1.06*** (0.22)	2.19*** (0.35)
Vert. pressure (β^2_{jv})	0	-0.39 (1.44)	1.84** (0.83)	0.61 (0.75)	1.61* (0.96)	1.66* (0.97)
Horiz. pressure						
ShareMaghrebi$_i$ (β^3_{jv})	0	3.59* (2.12)	0.85 (1.13)	0.01 (0.89)	0.08 (1.04)	2.35 (1.53)
CapacityMosques$_i$ (β^4_{jv})	0	-0.12* (0.07)	0.10*** (0.03)	0.01 (0.03)	-0.05* (0.03)	0.04 (0.03)
Schooling (γ^1_{jv})	0	-0.03 (0.03)	-0.05** (0.02)	0.07*** (0.02)	0.03 (0.03)	-0.02 (0.02)
Work experience (γ^2_{jv})	0	-0.09* (0.05)	-0.04 (0.03)	0.17*** (0.02)	0.17*** (0.03)	0.11*** (0.03)
Observations	2802					
Sampling weights	✓					
Additional controls[1]	✓					
Pseudo R^2	0.216					

Note: This table reports estimates of the parameters of the econometric model (3.11). The baseline category is the choice of inactivity and not wearing any religious symbol. Individual religiosity and vertical religious pressures are measured as indices (with mean zero and variance 1) constructed from multiple proxies available in the TeO data (see Appendix B.1.1 for details). ShareMaghrebi$_i$ is the proportion of the local population that is of Maghrebi origin. CapacityMosques$_i$ is the estimated capacity in Muslim places of worship in the area of residence. Robust standard errors in parentheses. Level of statistical significance : * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
[1] Additional controls include age, age squared, marital status (a dummy for having a partner), a dummy equal to one if the partner is working, immigration status and a set of dummy variables for quintiles of the local (neighborhood-level) unemployment rate of immigrants.

B.1.5 Plots for the tests of the four implications

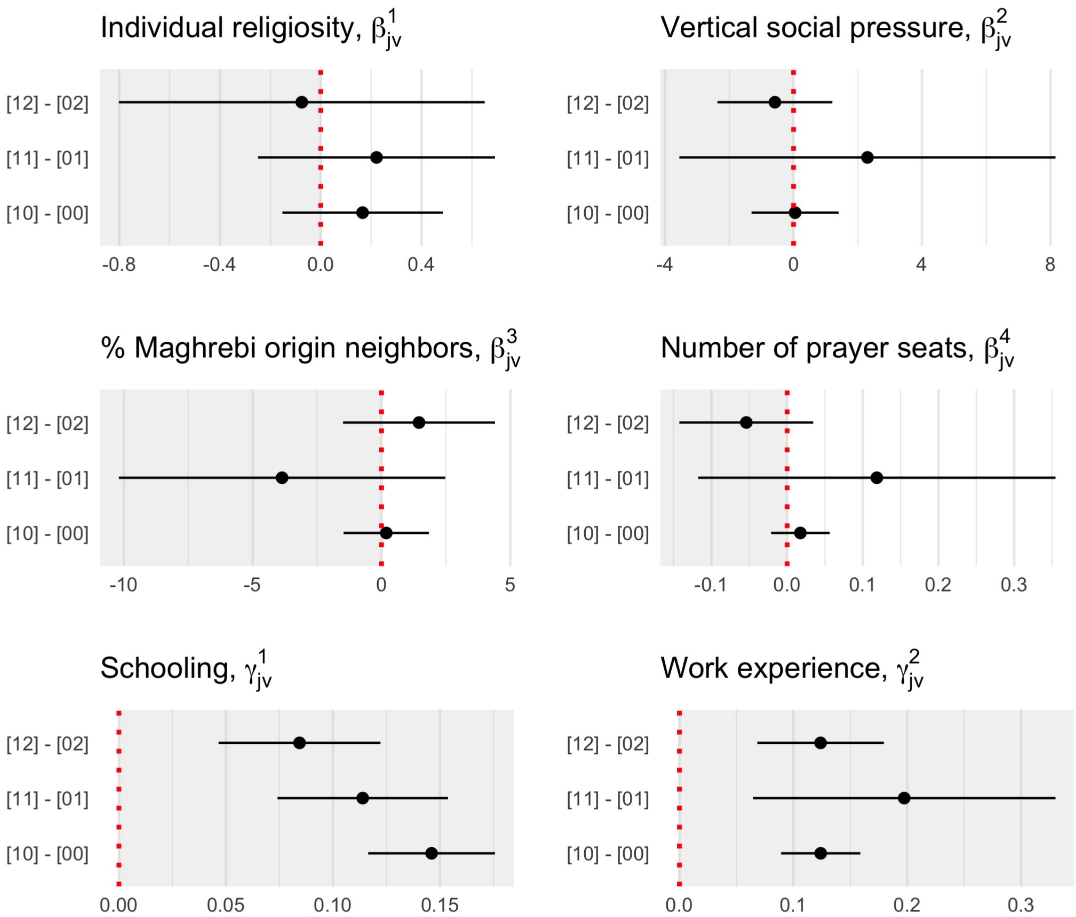
In Figures B.1 to B.4 we present the results of the tests of Implications 1–4, respectively.

Figure B.1: Hypothesis tests for Implication 1



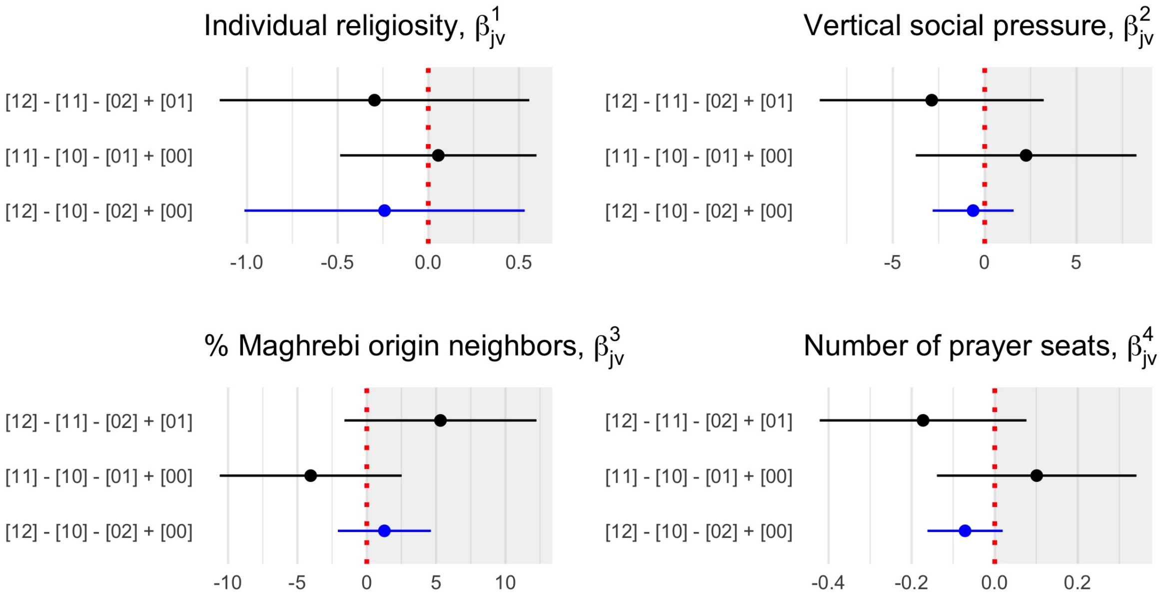
Note: Shaded areas correspond to the region where estimates are predicted to fall. Vertical axis labels correspond to the combination of (j, v) alternatives (e.g. the first line of the top-left graph plots the estimate for $\beta^1_{01} - \beta^1_{00}$). In blue: combinations which compare conspicuous symbol-wearing with no symbol-wearing. In black: combinations which include intermediate comparisons with discrete symbol-wearing. 95% confidence intervals are reported.

Figure B.2: Hypothesis tests for Implication 2



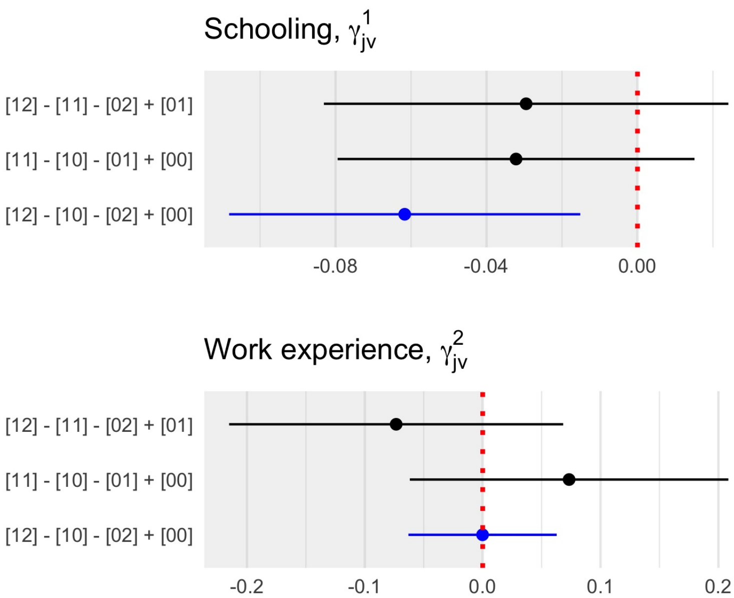
Note: Shaded areas correspond to the region where estimates are predicted to fall. Vertical axis labels correspond to the combination of (j, v) alternatives (e.g. the first line of the top-left graph plots the estimate for $\beta^1_{12} - \beta^1_{02}$). 95% confidence intervals are reported.

Figure B.3: Hypothesis tests for Implication 3: Religious motives channel



Note: Shaded areas correspond to the region where estimates are predicted to fall. Vertical axis labels correspond to the combination of (j, v) alternatives (e.g. the first line of the top-left graph plots the estimate for $\beta^1_{12} - \beta^1_{11} - \beta^1_{02} + \beta^1_{01}$). In blue: combinations which compare conspicuous symbol-wearing with no symbol-wearing. In black: combinations which include intermediate comparisons with discrete symbol-wearing. 95% confidence intervals are reported.

Figure B.4: Hypothesis tests for Implication 4: Economic discrimination channel.



Note: Shaded areas correspond to the region where estimates are predicted to fall. Vertical axis labels correspond to the combination of (j, v) alternatives (e.g. the first line of the top-left graph plots the estimate for $\beta^1_{12} - \beta^1_{11} - \beta^1_{02} + \beta^1_{01}$). In blue: combinations which compare conspicuous symbol-wearing with no symbol-wearing. In black: combinations which include intermediate comparisons with discrete symbol-wearing. 95% confidence intervals are reported.

B.2 Veiling and economic outcomes in Turkey

In this Appendix, we explore the relationship between veiling and economic outcomes in Turkey and compare it to what we obtained for France and to that found by Shofia (2020) for Indonesia. Turkey is an interesting context to study veiling patterns since "it has long been considered a unique case of successful modernization through secularization" (Platteau, 2017, , p.355). Between the proclamation of the Turkish Republic, in October 1923, and the rise of the pro-Islamic conservative Justice and Development Party (AKP) to power in the early 2000s, the country was ruled by secular governments. The founders of the Republic implemented a top-down nationalist modernization project to "Westernize" Turkey. A major aspect of the multiple reforms adopted over the following decades was their secular nature as the government wanted to build a national identity that would subordinate the religious one (Sakalli, 2019). Inspired by French State secularization, reforms ranging from the abolishment of the Caliphate to the adoption of Western dress codes profoundly changed the Turks' religious life. The series of secular legislation included veil bans in the public sphere. The 1982 Turkish constitution regulates veiling for civil servants, requiring women to uncover their head while on duty. The ban on headscarves was then extended to all universities in Turkey in 1997. Those regulations stayed in effect until they were gradually repealed by AKP: in 2010 for university campuses; in 2013 for state institutions; in 2014 for high schools; in 2016 for policewomen; and in 2017 for female army officers (Corekcioglu, 2021).

Given that, despite the secular modernization of Turkey, Islam is by far the most prominent religion in the country, we see Turkey as an intermediate case between France and Indonesia in our theoretical framework. Similar to France, women face legal disincentives to veil in public. However, like Indonesia, Turkey is a Muslim-majority country. Therefore, we would expect the correlation between veiling and economic outcomes in Turkey to mirror those differences. Specifically, we expect the correlation between veiling and economic participation to be *negative*, but lower in magnitude than what we see in France because most of the Turkish society is religious.

To study the patterns of veiling and economic participation, we use Turkish data compiled from multiple sources by Livny (2020).[3] Importantly, these data contain information on veiling practices in Turkey, which is available at the district level. We collapse the different types of veils (turban, hijab, and burka) so as to obtain a single measure of veiling rate in each district. For economic outcomes, so as to harmonize those variables with our measures of veiling that span the years 2010 to 2015, we take the average of the outcomes in the district (province for GDP per capita) over the same time period. In Figure B.1, we plot the relationship between the veiling rate and four measures of economic participation (female primary and secondary school completion, the female

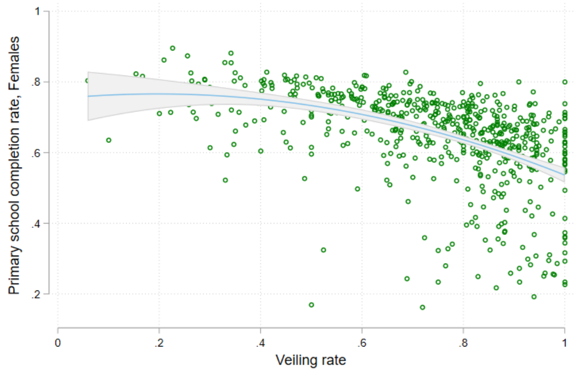
[3] The data are publicly available on Avital Livny's website (https://www.alivny.com/data).

literacy rate, and GDP per capita) along with a quadratic fit.[4] For all of the outcomes we observe a negative association, suggesting that, in Turkey as in France, the veil might not act as an integration strategy. Interestingly, these negative relationships appear to be linear as most of the (small) curvature is driven by regions of the veiling-rate distribution with low mass (i.e. districts with low veiling rates).

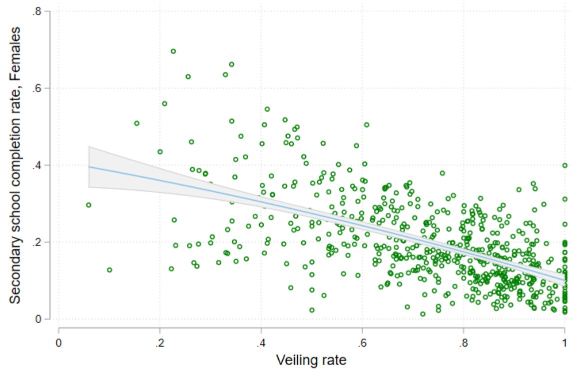
We take these results as further suggestive evidence in line with the theory. The wearing of the veil was frowned upon by the secular elite before the bans were repealed, thus imposing a high cost to women when they veil and are economically active. Actually, as Platteau (2017) argues, the rise of an Islamist party to power reinforced the laicists' attachment to the secular values. Islamic symbols, such as the veil, were sometimes also seen as manifesting a political identity in the public sphere in an increasingly polarized political context. Thus, even if Turkey is a Muslim-majority country, we find that the positive correlation documented by Shofia (2020) in Indonesia does not hold in this data. This suggests that her results regarding veiling behavior and economic participation are context-specific. Viewed through the lens of our theoretical framework, such a correlation can hold in Indonesia only because of two concomitant factors: (1) Indonesia is a Muslim-majority country, and (2) the veil is not subject to social or legal disapproval.

[4]For robustness, we also checked whether this relationship could be driven by religiosity of the district. We produced similar plots in which we control for religiosity and find very similar conclusions. Results are available upon request.

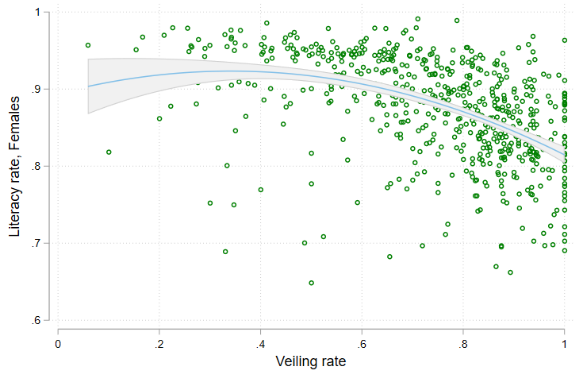
Figure B.1: Relationship between veiling and economic outcomes at district level, Turkey 2010–2015



(a) Primary school completion



(b) Secondary school completion



(c) Literacy

(d) GDP per capita

Note: The data source is Livny (2020). These figures plot the relationship between the veiling rate in a district in 2010–2015 and the average of an economic outcome in that district over the same period, along with a quadratic fit and 95% confidence bands. For GDP per capita, the dependent variable is measured at the province level.

Chapter 4

Incorrect Proxy, Incorrect Conclusions? Revisiting the Economic Impacts of the French Headscarf Ban in Schools

Sébastien Montpetit

Abstract

This paper studies the effect of prohibiting the wearing of the Islamic veil for pupils on educational attainment of Muslim women. In a difference-in-difference analysis, I find that the directive to school principals to ban the veil in French schools in 1994 induced a very large decline in high-school completion rates of Muslim women. There is further evidence that the effect on the intensive margin of education lasts in the medium-run. The data suggests that the impact of the ban operates through increased experiences of discrimination against Muslims and mistrust of the French school rather than through a change in Muslim parents' investments into their daughters' education. I show how using an inappropriate measure of the treatment group as in previous work substantially alters conclusions on the impacts of the policy. In the long-run, cohorts affected by the ban display lower levels of religiosity.

4.1 Introduction

With the rise of right-wing parties and immigration flows in Europe, the tension between Western values and religious practices of Muslims resurfaces in public debates. In particular, the wearing of the Islamic veil is often perceived as a signal of the subordination of women and a threat to State secularism. As a response, about one third of European countries have enacted policies to limit the wearing of Islamic clothing in public spaces (Abdelgadir and Fouka, 2020). While such regulations might have desirable effects on the preservation of the majority culture, their consequences on the targeted populations are a priori ambiguous. On the one hand, if women do not choose to veil willingly, bans can liberate them from a cultural norm imposed upon them. On the other hand, if veiling provides significant religious benefits, prohibitions may lead to more social exclusion and segregation (Carvalho, 2013). To shed light on this question, I focus on the case of France, the only country today where the headscarf is prohibited in public schools for both employees and pupils.[1]

In this paper, I revisit the current evidence on the impacts of the French headscarf ban in a difference-in-difference analysis comparing cohorts of women who have reached the age of veiling before and after the policy. I leverage rare rich survey data in which religious affiliation is observed along with religious practices and other detailed information about respondents, allowing to study the impact of the policy on new outcomes. The richness of the data further allows to study various sources of heterogeneity in policy impacts as well as to provide additional evidence on the mechanisms that are likely at play.

Using data in which religion, the appropriate measure of the treatment and control groups, is observed is a key improvement over previous studies of the French ban. Two recent papers reach opposite conclusions on the effects of this policy on educational attainment. On the one hand, Abdelgadir and Fouka (2020) find that the 2004 ban depressed schooling outcomes of French girls of North-African origin. Their results suggest that the negative impact of the ban operates through increased perceptions of discrimination at school. On the other hand, Maurin and Navarrete-H. (2023) find that the 1994 ministerial circular asking school principals to prohibit the wearing of the veil in schools (henceforth Bayrou circular) had a positive impact on their educational attainment. They find that the issuance of the circular is positively associated with other measures of social integration such as mixed marriages. Even if they are comparing different cohorts of adolescents and different treatments, these contradictory pieces of evidence are puzzling. One potential reason behind the fragility of these results is the

[1] To my knowledge, Turkey is the only other country where the veil had been prohibited for students. Veil bans in some public settings were part of the top-down nationalist modernization project to "Westernize" Turkey. Headscarf prohibitions for students on university campuses were put in place in 1997. The regulation stayed in effect until it was repealed by the pro-Islamic conservative Justice and Development Party (AKP) in 2010 (Corekcioglu, 2021).

fact that they do not use the same measures of treatment in their analysis, sometimes even within the same paper.[2]

I show that properly identifying the treatment group in this context yields several new insights on the effects of religious prohibitions in secular countries. I notably show that the documented positive impact of the Bayrou circular holds for individuals of African origin, but that the impact on the actual treatment group (Muslim women) is of opposite sign. I document a very large short-term negative impact of the circular issuance on Muslim women's probability to have completed high school. The point estimates suggest a decline in the high-school completion of about 25% of the pre-treatment mean. The results imply that Muslim girls lost about 5 years in the catching-up process ongoing prior to the ban. Moreover, thanks to the richness of the TeO data on respondents' education, I show that the negative impact also holds at the intensive margin of education. For this outcomes, I find that the negative impact lasts in the medium-run. The stark difference in estimated impacts when using religious affiliation or a proxy for it suggests the presence of a strong bias due to measurement error in previous studies.

Next, I leverage the richness of the data to dig into the potential mechanisms that might explain the drop in schooling. I show that, consistent with Abdelgadir and Fouka (2020), Muslim women exposed to the new directive are more likely to report experiences of discrimination due to their religion. This effect is not explained by their different origins since discrimination based on this trait decreases. The Bayrou circular is also associated with a very large increase in mistrust of the French school by affected Muslims. On the other hand, the impact could be driven by Muslim parents increasing pressures on their daughters to strictly follow their religious tradition. However, the data suggests that parents played a limited role in this depressed schooling. I find that the negative impact on schooling is not stronger in families that devoted more effort in transmitting their religion. In the same line of thought, affected Muslim girls are not less likely to had been helped by their parents and siblings with their homework. Muslim girls are also not more likely to attend a private school – where the ban did not apply – or a school not in their sector of residence. In a heterogeneity analysis, I also show that the depressed schooling is rather concentrated among girls that are more socially isolated. The impact is indeed much stronger for women whose friends are mainly Muslims and among those whose mother was not working while they were adolescents.

Last, I study the long-run impacts of the ministerial circular. I focus on religiosity outcomes and the possibility of a religious backlash. Contrary to the Turkish case (Sakalli, 2019), I find that treated Muslim women display lower levels of religiosity later in life and are less likely to use their religion to self-identify. Also consistent with the absence

[2] Abdelgadir and Fouka (2020) check the robustness of their results to using alternative data sources, including the first wave of the Trajectories and Origins (TeO) survey in which religion is observed. However, by using only the first wave, they obtain a very small working sample of less than 2,000 observations in the regression analysis.

of religious backlash, I find no significant impact on veiling behavior. I hypothesize that the increased emphasis on State secularism (*laïcité*) as "a pillar, even the identity and foundation of the community life" (Andriantsimbazovina et al., 2020, , p. 7) might induce religious minorities in France to reduce the intensity of their religious life.

This study first contributes to the literature on assimilation and integration policies. I show that the French headscarf ban depressed schooling outcomes of Muslim girls, and that the impact seems to operate through increased discrimination against this religious group. These results are in line with previous literature studying other forms of *assimilationist* regulations which show that these can backfire. Fouka (2020) shows that German language prohibitions in U.S. schools after WWI made German-Americans less likely to volunteer in World War II and increased their cultural distance with the majority. On the contrary, easier access to citizenship (often thought as an *integration* policy) for immigrants in their host country were found to improve labor-market attachment and social integration of immigrants (see Gathmann and Garbers, 2023, for a recent review of this literature). A notable exception is Dahl et al. (2022) who show that the introduction of automatic birthright citizenship in Germany had negative impacts on Muslim girls. They find that Muslim girls born soon after the policy have lower life satisfaction and self-esteem and are less socially integrated into German society. Consistent with evidence on assimilationist policies, I show that the headscarf ban might not be a tool to facilitate their economic integration either.

Second, my findings contribute to the literature on the interactions between education and identity. I show that restricting the extent to which a minority can signal its identity in an educational context can have unintended consequences. Rather than promoting Muslim girls' integration into the secular society, I find that the French headscarf ban both reduces their educational attainment and increases religious segregation. This is consistent with economic theory which suggests that a ban on veiling can increase religiosity (Carvalho, 2013) and that marginalized cultural communities may underinvest in education following a strengthening of the secular content of mainstream education (Carvalho et al., 2017, 2024). Similarly, Cantoni et al. (2017) find that a major textbook reform in China shaped attitudes towards the government, democracy, and free markets. Squicciarini (2020) shows that, as a response to the modernization of education in 19th century France, more religious areas pushed for religious education and were slower to transition to modern education. Closer to this study, Sakalli (2019) finds that the secularization of Turkish schools led to a decline of schooling and an increase in religiosity, particularly in pre-secularization pious districts.[3] Despite the conclusions being similar, my work differs from that study because I am interested in a context in which Muslims form a minority and in which the content of education itself was unchanged.

[3] Similarly, Benzer (2022) finds that the subsequent re-introduction of Islamic schools in Turkey, which do not prohibit the headscarf, had positive impacts on girls' educational attainment.

Third, this paper relates to a recent literature on reproducibility and replicability in Economics. A key finding in this literature is that results of many published articles do not replicate, thereby questioning the reliability of causal statements in empirical economics (Gertler et al., 2018; Huntington-Klein et al., 2021; Brodeur et al., 2024). Related evidence of p-hacking and the lack of transparency in economics research (e.g. Brodeur et al., 2016, 2020) has set in motion a new set of research norms which views replication of published studies as an essential diagnostic tool (Miguel, 2021). I contribute to this literature by applying these concepts to the case of a reform that has substantial policy implications for integration policy of religious minorities in Western countries. I perform a *direct replication* of the two previous studies of the French headscarf ban in schools using a different data source allowing for better identification of the treatment group, but using similar procedures.[4] In this context, I show that appropriately measuring a treatment group can substantially affect the conclusions on the effects of such policy on economic integration of religious minorities. I find that, while my findings are consistent with results in Maurin and Navarrete-H. (2023) using origin as a proxy for Muslim affiliation, the impact on the actual treatment group is of opposite sign.

The rest of the article is structured as follows. Section 4.2 describes the institutional context. Section 4.3 presents the data sources along with summary statistics. Section 4.4 evaluates the impact of the ban on Muslim girls' educational attainment and discusses the mechanism. Section 4.5 shows impacts on long-term outcomes. Finally, section 4.6 concludes.

4.2 Institutional context

4.2.1 The French headscarf ban in schools

The wearing of the Islamic veil has been a burning issue in France since at least three decades. In 1989, the "*affaire des foulards*" (headscarf affair) garnered nationwide attention when three girls were expelled from their middle school for refusing to remove their headscarves. The incident sparked heated debates and was followed by similar disputes in other schools. Eventually, the affair was settled after the highest French administrative court (the *Conseil d'État*) ruled in favor of the expelled girls (Scott, 2009). In its ruling, the Council stated that banning the wearing of signs of religious affiliation by students in public schools was against their freedom of religion.

The 1994 Ministry circular. Five years later, following the election of a right-wing government, the minister of education, François Bayrou issued a circular asking

[4] Different types of replication exercises and their definitions are suggested in Dreber and Johannesson (2023) and by the Institute for Replication: https://i4replication.org/definitions.html.

school principals to prohibit conspicuous religious symbols worn by students.[5] To justify the government's position, the document insists on the distinction between conspicuous symbols and discreet signs that was already present in the 1989 ruling. The Minister argued that conspicuous symbols are "in themselves acts of proselytizing" and should thus be prohibited in public schools. This interpretation of the Council's ruling is somewhat different from its original meaning, which stated that as long as the student's behavior was not disrupting class activities, one should not be refused admission to school for wearing a veil. Despite some opposition from the *Conseil d'État*, several school principals decided to follow Bayrou's recommendation and to adopt the ban over the following years.

To help implement the bans, Simone Veil, the minister of social affairs, appointed a woman of North-African origin, Hanina Chérifi, as mediator to handle problems on the ground. In the school year that followed the circular, around 3,000 cases required an intervention from the mediator with only 139 leading to exclusions. This former figure quickly dropped to 1,000 in 1996 and to about 150 in 2002, suggesting that most establishments implemented the suggested ban soon after the circular was issued.

The 2004 law. Partly fuelled by concerns about terrorism after the 9/11 attacks in New York City, President Jacques Chirac appointed a commission in July 2003 to explore the feasibility of enacting the ban proposed in the 1994 circular into law. Once again, the public debate was fierce. On the one hand, proponents of the ban argued that the policy could "free" Muslim girls from religious pressures and that headscarves "infringed on the liberty of conscience of other pupils and represented the triumph of communitarian pressures" (Abdelgadir and Fouka, 2020, p. 4). On the other hand, critics replied that it would rather impede the integration of Muslim girls by driving them away from public education.

The commission ultimately recommended a ban on conspicuous religious signs in public schools, which was enshrined in law in March 2004 and started being enforced in October 2004. As for the ministry circular, the Islamic veil was the main target of the law and Muslim girls were the main group affected in practice. A report from Chérifi (2004) based on fieldwork in four school academies shows that, in 2004-2005, only 639 students showed up to school wearing a conspicuous religious sign, less than half than in the previous academic year. About 200 of these 639 students switched to private schools or opted for distance learning (Mattei and Aguilar, 2016). Despite these small numbers, all the media and political attention might have changed the schooling environment for Muslim girls beyond the management of these cases (Abdelgadir and Fouka, 2020).

[5]In France, a ministry circular is a governmental document which gives a clarification or an interpretation of the law or establishes guidelines for civil servants.

4.2.2 The French educational system

The headscarf ban targets pupils at the primary and secondary levels, but does not apply to students attending college. Therefore, potential impacts of the ban should be concentrated in pupils' schooling trajectory that precedes university studies. In France, pupils enter elementary school at age 6 and this lasts for 5 years (until age 11). Then, they attend middle-school (*collège*) for four years (until age 15) and school attendance is mandatory until age 16. After middle-school, students enter high school, either to pursue a vocational degree in a professional high school or to prepare for the *baccalauréat* in a general or a technological high school. The *baccalauréat* is the diploma that allows for the possibility of continuing in higher education. Following previous studies of the French headscarf ban, completion of this degree is the primary outcome of interest.

4.3 Data and empirical strategy

4.3.1 Data and sample

My primary data source is the two *Trajectoires et Origines* surveys (henceforth TeO; Beauchemin et al. (2016, 2023)). Conducted in 2008–2009 and 2019-2020 by the French National Institutes for Demographic Studies, the TeO surveys targeted adults between 18 and 60 years old residing in metropolitan France. Purposefully oversampling immigrants and minorities, it includes 3,033 and 3,519 women who identify as Muslim in the first and second waves respectively. To my knowledge, these are the largest samples of this kind in France.[6] When including Muslim men and other religious groups, the entire surveys contain more than 21,000 observations each.

The TeO datasets are a comprehensive source of information on various aspects of respondents' lives, including living conditions (such as employment, education, housing, commune of residence, and health), social life (such as migration history, language use, family, and children), and public life (such as political views, experiences of discrimination, and social relationships). Of particularly value for this study is the religion section, which is a unique inclusion in a French survey of this scale since the collection of individual information on religion is closely monitored in France. This section includes variables such as religious affiliation, measures of religiosity, religious symbols worn, and intergenerational religious transmission. This is a key advantage over the French Labour Force Surveys (LFS), the data source used in Abdelgadir and Fouka (2020) and Maurin and Navarrete-H. (2023). Indeed, the LFS only offers a proxy for religious affiliation, namely the parents' place of (and nationality at) birth. This information is available in

[6] Two surveys conducted by private firms, namely Institut Montaigne (2016) and Institut Français d'Opinion Publique [IFOP] (2019), have much smaller sample sizes (slightly above 1,000 individuals of Muslim origin, both genders included) and do not have a similarly deep content as that of TeO.

the TeO surveys and I can thus use it for comparative purposes.

Measurement of religion. A contribution of this study is to better identify the treatment and control groups than in previous studies. The empirical analysis relies on an actual measure of religious affiliation along with measures of religiosity which I describe here. All measures are self-reported by respondents, who had the possibility to refrain from answering these sensitive questions (while most other sections of the TeO surveys are mandatory). Nevertheless, response rates to the religion section are very high (above 98% in each wave for religious affiliation).

First, for religious affiliation, respondents are asked whether they currently have a religion. If they answer "yes", then they are asked which one and the interviewer notes the exact answer given. Particular denominations within each religious family are then pooled into 12 groups in the survey data. For the main empirical analysis, I further pool denominations into five groups: Atheists (no religion), Catholics, Other Christians, Muslims, and Others. The latter category is mostly composed of Buddhists, Jews, and Hindu/Sikh. The same questions are asked about other family members, namely the father, the mother, and the partner.

Second, the TeO surveys include a set of questions about individual religiosity. My preferred measure is the frequency of attendance of religious ceremonies, a standard measure of religiosity which focuses on religious practice (Iyer, 2016). I also consider other measures of individual religiosity: the self-reported importance of religion in the respondent's life and whether she uses her religion to self-identify. To exploit these multiple measures altogether, I also build a single measure of individual religiosity. To do so, I use a measurement system to construct a latent index of individual religiosity, as is being done for different latent variables in Heckman et al. (2013) or Bolt et al. (2021). In Appendix C.1 I provide details on the procedure and on the survey questions.

Third, for veiling behavior, I use the following question from the TeO survey:

> *In your daily life, do you wear in public a piece of clothing or jewelry that might evoke your religion?* *(1) Never* *(2) Sometimes* *(3) Always*

If applicable, respondents were subsequently asked to report which religious symbols they wear. Answers were later sorted by the survey institute into four categories: jewelry, clothing, headcoverings, or others. Because they visibly signal religion and are the ones usually targeted by secular policies, I group the clothing and headcoverings categories together as *conspicuous symbols*.

Sample selection. The main empirical analysis focuses on the effects of the 1994 circular on girls reaching puberty around its issuance. Because individuals aged less than 20 years old might still be in high school, I focus on women aged at least 21 years old.

This leads me to restrict the sample to individuals born between 1971 and 1987 so as to have similar numbers of observations on both sides of the cohort threshold. Given these restrictions, I end up with a working sample of 7,758 women (and 7,261 men for placebo exercises), of whom more than 21% are Muslim.

4.3.2 Descriptive analysis

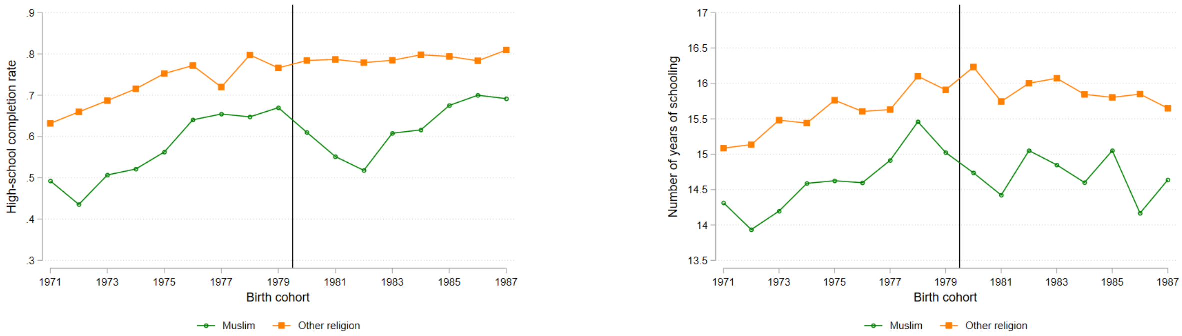
I start the empirical analysis by comparing the characteristics of women in the treatment and control groups. Table C.1 contains summary statistics, separately for Muslim and non-Muslim women. Compared with their counterparts of other religious groups, Muslim women display worse economic outcomes: they are less educated and much less likely to be employed at the time they are surveyed. In terms of educational attainment, the main outcome of interest, Muslim women also appear to be lacking behind. Not only are they twice as less likely to have completed any degree, but they are also substantially less present at higher levels. Only a third of Muslim women has completed a college degree while more than half of non-Muslim women graduated in higher education. This translates into a difference of one year of schooling. Muslim women also substantially differ from atheists and individuals with other religious affiliations on various aspects of their lives. For example, they display higher levels of religiosity and are more likely to report that their parents invested in their religious education (measured by whether religion was very important in their education).

To understand whether the headscarf ban might have played any role in Muslim women's situation today, I first plot the time series of my main outcomes of interest in Figure 4.1, separately for Muslim women and women of other religious affiliation. A striking pattern emerges: while the educational attainment of Muslim women born in the 1970s was somewhat catching-up over that of women of other religious groups, there is a significant break in this trend that coincides with the issuance of the ministerial circular. The drop is very large and increases over the first three affected cohorts before the trend bounces back to that observed in the cohorts of the 1970s. Thus, at first glance, this suggests that the issuance of the circular abruptly delayed the catching-up process of Muslim girls over other girls. At the intensive margin, a similar decrease occurs, but it is much less pronounced. However, contrary to the high-school graduation rate, the drop seems to be longer-lasting. To verify whether I can interpret this potential impact as causal, I now move to the estimation of a more parsimonious difference-in-difference model.

4.3.3 Empirical strategy

The main empirical specification is a standard difference-in-differences design, which compares cohorts of women of different religions reaching puberty (the age of veiling)

Figure 4.1: Evolution of educational attainment of Muslim and non-Muslim women across birth cohorts, 1971-1987



(a) High-school completion

(b) Years of schooling

Note: Evolution of the high-school completion rate across birth cohorts of Muslim and non-Muslim women born in France. On the right of the vertical line are cohorts who are subject to the 1994 Bayrou circular asking school principals to prohibit the headscarf in schools.

before or after the 1994 circular was issued. I define women of all Muslim denominations as the treatment group and I compare their educational outcomes to those of other religious groups that are unaffected by the ban. This approach differs from those of Abdelgadir and Fouka (2020) and Maurin and Navarrete-H. (2023) (henceforth AF and MN) in one crucial dimension, which is that I observe respondents' religious affiliation. The strategies used in those papers, summarized in Table 4.1, use proxies of religious affiliation, namely the father's country of birth or the father's nationality at birth. As also shown in this Table, those papers also differ on the age threshold chosen to define treated cohorts and the specific policy being studied. In this paper, as in MN, I focus on the 1994 circular because most schools already implemented bans in 1994 (Chérifi, 2004). As for the age threshold, I also follow that study and compare girls reaching the age of veiling before and after the circular was issued.

Formally, for individual i in birth cohort c, I estimate the following model:

$$Y_{i,c} = \alpha + \beta_1 \text{Muslim}_i \times \text{Post}_c + \beta_2 \text{Muslim}_i + \gamma_c + X'_{i,c}\delta + \varepsilon_{i,c} \tag{4.1}$$

where Post_c is an indicator variable taking the value of 1 if the individual is born after 1979, Muslim_i indicates Muslim affiliation, and γ_c is a full set of cohort dummies. In my simplest specification, the vector of controls X contains a dummy for the second survey wave, dummies for the other religious groups, and a Muslim-specific linear trend. In another specification, I also control for the respondents' living conditions when aged 15 years old. That is, I include indicators for whether the individual's father was working, the mother was working, and dummies for the *département* of residence. The main outcome variables Y are a dichotomous variable taking the value of one if individual i has completed high school (a *baccalauréat*) or the number of years of schooling.

Table 4.1: Empirical strategies used in previous studies of the French headscarf ban

Study	Policy	Age threshold	Data	Muslim proxy	Clustering level
Abdelgadir and Fouka (2020)	2004 law	19 y.o. (end of *bacc*)	LFS 05-12	Father's birthplace	Father's birthplace
Maurin and Navarrete-H. (2023)	1994 circular	15 y.o. (puberty)	LFS 05-19	Father's nationality	Father's nat. × department of birth

Note: This table presents the empirical strategies used in previous studies of the impact of the French headscarf ban on the high-school graduation probability of Muslim women. The acronym LFS refers to the French Labour Force Surveys.

Inference. The appropriate method for calculating standard errors is yet another source of disagreement between the two previous studies. While AF cluster standard errors at the father's birthplace level, MN do not use this approach, arguing that it yields only 14 clusters. However, their clustering level might not be more appropriate either since this is not the level at which treatment varies (Abadie et al., 2023). Here, treatment varies at the religion (or father's origin) and birth cohort levels, but not across the individual's birthplace within France. Therefore, I rather cluster standard errors at the religion × born-post-1979 level. This accounts for the within-religious group temporal correlation of errors while allowing for a structural break in this correlation for the treated cohorts. Since this leaves me with few clusters, I also report p-values calculated using the wild bootstrap procedure of Cameron et al. (2008) accounting for the small number of clusters.

4.4 Impact on educational attainment

As shown in Section 4.3.2, the time series of high-school graduation rates suggests the 1994 circular induced a strong decline in the educational attainment of Muslim girls. To confirm the graphical evidence, I estimate equation (4.1) and report the results in column (1) of Table 4.2. Consistent with the observed trends, results from the regression analysis suggest that the circular had a large negative impact on the treated cohorts.

I report the results in columns (1) and (2) of Table 4.2. In Panel A, I first focus on the effect of the circular on the high-school completion rate, the main outcome studied in previous papers. I find that the ministerial circular causes an average decline in the probability of Muslim girls to complete high school of 15 percentage points. This effect is economically large in magnitude as it represents 25% of the pre-ban mean for this group. I remain cautious in the interpretation of this estimate because, while it is statistically significant at conventional levels when clustering, the p-value I obtain

To better gauge the size of the impact on educational attainment, in Panel B, I exploit the richness of the TeO data to evaluate the impact at the intensive margin of

Table 4.2: Impact of the 1994 ministerial circular on educational attainment

			Proxies			
Measure of treatment:	Religious affiliation		Father's nationality		Father's birthplace	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: High school completion						
$\text{Muslim}_i \times \text{Post}_c$	-0.150***	-0.156***	0.054**	0.045**	0.082**	0.072**
	(0.036)	(0.023)	(0.020)	(0.020)	(0.016)	(0.018)
	[0.212]	[0.132]	[0.186]	[0.146]	[0.003]	[0.017]
Muslim_i	-0.171***	-0.154***	-0.242***	-0.204***	-0.089**	-0.077***
	(0.024)	(0.016)	(0.043)	(0.035)	(0.032)	(0.018)
	[0.243]	[0.014]	[0.072]	[0.054]	[0.126]	[0.208]
15 y.o. controls		✓		✓		✓
Mean dep. var.	0.599	0.611	0.615	0.625	0.631	0.641
N	7,046	6,905	6,960	6,818	7,014	6,872
R^2	0.04	0.04	0.047	0.08	0.043	0.076
Panel B: Years of schooling						
$\text{Muslim}_i \times \text{Post}_c$	-0.564***	-0.705***	0.352***	0.178	0.420***	0.225
	(0.244)	(0.196)	(0.115)	(0.163)	(0.169)	(0.160)
	[0.280]	[0.280]	[0.054]	[0.269]	[0.145]	[0.369]
Muslim_i	-0.970***	-0.635***	-1.707***	-1.501***	-0.515**	-0.346
	(0.150)	(0.167)	(0.235)	(0.151)	(0.207)	(0.231)
	[0.121]	[0.264]	[0.062]	[0.008]	[0.271]	[0.437]
15 y.o. controls		✓		✓		✓
Mean dep. var.	14.72	15.04	14.82		15.03	
N	6,300	6,174	6,233		6,273	
R^2	0.040	0.071	0.050		0.047	

Note: This Table reports regression estimates of the impact of the issuance of the headscarf ban ministerial circular on educational attainment of Muslim girls. It compares results when using religious affiliation to measure treatment versus using a proxy as suggested in the previous literature. Means of the dependent variable in the treatment group over pre-1980 cohorts are reported. Control variables are full sets of birthyear, survey waves, and religion (or father's origins) dummies and a Muslim-specific linear trend. Standard errors clustered at the religion (or father's origin) × born-post-1979 level reported in parentheses. *p*-values computed using the wild bootstrap procedure are reported in brackets. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

schooling. Precisely measuring years of completed education is not possible with the Labour Force Surveys and thus these results are new in the literature on this reform. The TeO surveys report the specific grade level at which the individual left school up to the Master's (*BAC+5*) level. I can thus reconstruct the number of years of schooling (from elementary school), capped at twenty years.[7]

In columns (3) to (6) of Table 4.2, I assess how my results compare with previous studies. To do so, I estimate the impact of the circular using proxies for religious affiliation used in previous papers (see Table 4.1). Interestingly, I find that I can replicate the main result of Maurin and Navarrete-H. (2023) in the TeO data. They find that the issuance of the circular had a positive impact on girls of African origin, specifically of women whose father's has an African nationality. I obtain an estimate that is slightly lower, but similar in magnitude to that found in that study (5.4 versus 7.8 percentage points). This result is in sharp contrast with the estimated impact

[7]Even if completing graduate studies is somewhat more frequent in France than in North America, only about 11% of the sample has completed at least a Master's degree.

One important point to note, however, is that the statistical significance of the results is quite sensitive to the method used to obtain them. When accounting for the small number of clusters using the wild bootstrap of Cameron et al. (2008), most estimates have high p-values and thus lose statistical significance. However, this adjustment to standard errors for the small number of clusters was not made in the two previous papers (see Table 4.1) so my results using the standard clustering approach are more comparable to the current evidence. I still report the bootstrapped p-values here for research transparency purposes.

Threats to identification. The identification assumption for a causal interpretation is the standard parallel-trends assumption in that Muslim and non-Muslim girls' educational attainment would have evolved the same absent the circular. As shown in Figure 4.1, the two groups were not evolving the same prior to treatment. Therefore, to support the parallel-trends assumption, I estimate an event-study model.[8] Results are reported in Figure 4.2. For high-school completion, except one aberrant cohort (that of 1975), the treatment and control groups appear to evolve similarly in the pre-treatment periods. Indeed, Muslims girls catch-up on the control group in the early 1970s, but the two groups evolve similarly from 1976 up to the first treated cohort. This Figure also reveals that the average impact in Table 4.2 masks substantial dynamics in that the initially large negative impact vanishes for subsequent cohorts. For the number of years of schooling, there is small pre-trend in that I find a positive coefficient for the 1978 cohort, but it is quite small an unlikely to explain the large drop from the first treated cohort. For this extensive margin of schooling, the dynamics point to a rather persistent negative impact, rather than simply a short-term phenomenon.

Another potential threat to identification arises if another shock differentially affecting Muslim girls occurred simultaneously. I believe that this is unlikely for two reasons. For one, the main other episode that spurred discrimination against Muslims in this period is the September 2001 attacks in New York City, which occurred several years after the circular. Second, the one change in the educational system that coincides with the issuance of the circular concerns a reform that aimed at making vocational high school more attractive. This reform reduced the number of years required to complete the vocational degree by one year for some tracks in 2008 and then for all occupational tracks in 2009. It also introduced catch-up exams for vocational high school students in their final year. While ethnic (and religious) minorities tend to be over-represented

[8]Formally, this event-study model writes:

$$Y_{i,c} = \alpha + \sum_{\substack{c=1971 \\ c\neq 1979}}^{1987} \beta_c \,\text{Muslim}_i \times D_c + \gamma_r \text{Muslim}_i + \gamma_c + X'_{i,c}\delta + \varepsilon_{i,c} \tag{4.2}$$

where D_c is an indicator variable taking the value of 1 if the individual is in birth cohort c.

in the vocational tracks (Belzil and Poinas, 2010), as Maurin and Navarrete-H. (2023) argue, there is no strong reason to believe that this policy change would differentially affect male and female students.

4.4.1 Robustness

Results of the main regression analysis suggest that, contrary to previous findings, banning the veil in French schools worsens educational outcomes of Muslim girls. To validate this result, I perform a series of robustness checks and report them in Table 4.3. In column (2), instead of restricting the sample to individuals born in France (as in previous papers), I instead consider individuals who have completed their education in France. This increases the sample size because many immigrants had studied in France even if born abroad. The estimated impacts are essentially the same when I use this different sample restriction. In the third column, I change the composition of the control group by excluding atheists. The point estimate on high-school completion decreases a little in magnitude. This is expected given that atheists have a null religiosity by definition and thus should be totally unaffected by the Bayrou circular. This estimated impact is also statistically significant even when accounting for the small number of clusters. On years of schooling, the effect is imprecisely estimated, but still negative.

Next, in the last two columns, I estimate the difference-in-difference model using placebo groups that are unlikely to be affected by the ban. In column (3), I use the sample of men and consider Muslim men as the treatment group. As expected, since few Muslim men wear conspicuous religious symbols, I find no evidence of negative effect of the ban on this group. If anything, there is a small positive impact, but, as shown in Table C.2, this result is not robust to including additional control variables. In addition, there is no statistically significant impact on the intensive margin of education.

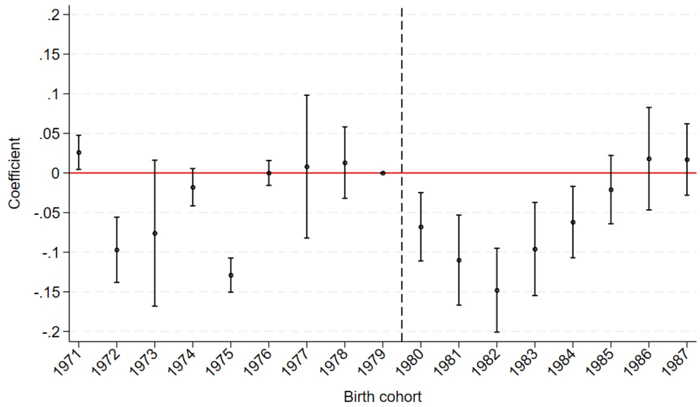
4.4.2 Discussion

In this section, I discuss the main result of the paper along two dimensions. I first explore potential channels through which the ban depressed educational outcomes of Muslim girls. Then, I relate my results to previous studies of the French headscarf ban and discuss the reasons that might explain the large differences in estimated impacts.

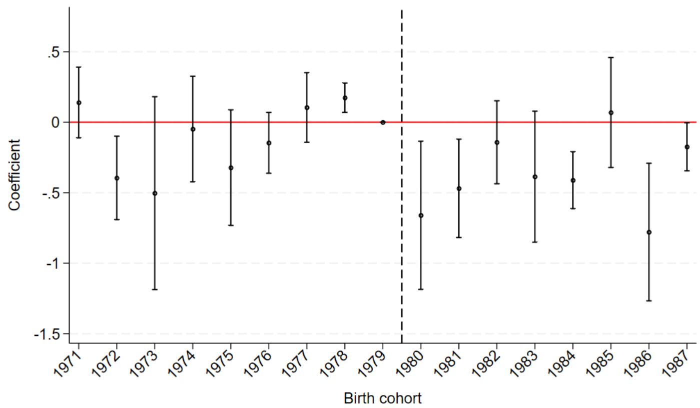
Mechanism

The richness of the TeO surveys offers a unique opportunity to investigate the mechanisms through which the ban induces a decline of schooling of Muslim girls. Given the limited content of the Labour Force Surveys (LFS) to explore such mechanisms, Abdelgadir and Fouka (2020) turn to the first wave of the TeO surveys in their analysis of the

Figure 4.2: The dynamic impact of the 1994 Bayrou circular on Muslim girls' educational attainment



(a) High-school completion



(b) Years of schooling

Note: These figures plot the coefficients of event-study regressions along with 95% confidence intervals. Outcomes variables are the high-school completion rate and the number of years of schooling. Control variables are full sets of birthyear, survey waves, and religion (or father's origins) dummies and a Muslim-specific linear trend. On the right of the vertical line are cohorts who are subject to the 1994 Bayrou circular asking school principals to prohibit the headscarf in schools.

Table 4.3: Robustness checks

				Placebo groups	
	(1) Baseline	(2) Schooling in France	(3) Excluding atheists	(4) Muslim Men	(5) Other religions
Panel A: High school completion					
$\text{Muslim}_i \times \text{Post}_c$	-0.150***	-0.148***	-0.123***	0.038**	
	(0.036)	(0.027)	(0.018)	(0.012)	
	[0.212]	[0.202]	[0.012]	[0.180]	
Muslim_i	-0.171***	-0.223***	-0.235***	-0.035***	
	(0.024)	(0.017)	(0.011)	(0.009)	
	[0.243]	[0.033]	[0.000]	[0.219]	
$\text{Has religion}_i \times \text{Post}_c$					-0.028
					(0.036)
					[0.523]
Has religion_i					0.061**
					(0.021)
					[0.281]
Mean dep. var.	0.599	0.568	0.599	0.514	0.746
N	7,046	8,704	4,254	6,613	5,431
R^2	0.042	0.043	0.053	0.033	0.025
Panel B: Years of schooling					
$\text{Muslim}_i \times \text{Post}_c$	-0.564***	-0.581***	-0.153	-0.470**	
	(0.244)	(0.117)	(0.252)	(0.180)	
	[0.280]	[0.187]	[0.727]	[0.327]	
Muslim_i	-0.970***	-1.226***	-1.105***	-0.892***	
	(0.150)	(0.061)	(0.163)	(0.098)	
	[0.121]	[0.007]	[0.070]	[0.202]	
$\text{Has religion}_i \times \text{Post}_c$					-0.658*
					(0.289)
					[0.094]
Has religion_i					-0.043
					(0.262)
					[0.930]
Mean dep. var.	14.72	14.82	14.72	14.38	15.67
N	6,300	7,696	3,791	5,996	4,860
R^2	0.040	0.053	0.049	0.028	0.022

Note: This Table reports results of robustness checks, namely of restricting the sample to individuals who have studied in France and to using placebo groups. Means of the dependent variable in the treatment group over pre-1980 cohorts are reported. Control variables are full sets of birthyear, survey waves, and religion (or father's origins) dummies and a Muslim-specific linear trend. Standard errors clustered at the religion × born-post-1979 level reported in parentheses. p-values computed using the wild bootstrap procedure are reported in brackets. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

2004 law. Despite their small sample and the fragility of their results (as shown, for example, in Maurin and Navarrete-H. (2022)), their analysis highlights potential channels. They argue that the intense public debates might have spurred discrimination against Muslims given that much of the public debate adopted an anti-veiling and anti-Muslim tone (Scott, 2009). In turn, increased perceptions of discrimination might have reduce school performance through a feeling of alienation.

In Table 4.4, I explore the discrimination channel by using self-reported experiences of discrimination as the outcome variable. In column (1) of Table 4.4, I show that treated cohorts are about 8.6 percentage points (24% of the pre-ban mean) more likely to report any experience of discrimination. However, since Muslim affiliation is correlated with being of African (especially Maghrebi) origin, these discriminatory treatments might not be solely related to religion. Interestingly, the TeO survey also asked respondents about what was the source and the context of the alleged discrimination. In columns (2) to (4), I leverage this information and I find that treated cohorts are more likely to perceive being discriminated against because of their religion and less so because of their origins. The impact on religion-based discrimination is very large as it corresponds to an increase of 66% of the mean. While there is no evidence that these additional experiences of discrimination occurred at school, I find that treated girls are more likely not to trust the French school. Overall, this evidence suggests that the impact of the ban operates through an alienation of Muslim girls at school.

Another potential channel which may explain the negative impact I document could be a reaction to the ban from pupils' families. For example, Dahl et al. (2022) find that, in reaction to an integration policy in Germany, Muslim parents were less likely to help their daughters with their homework and learning. In theory, a similar disinvestment in girls' education could have occurred in the French context as a reaction to the perceived secularization of education (Carvalho et al., 2024). To help rule out this mechanism, I first evaluate the impact of the circular on whether girls received help at school from their family and peers. In columns (1) to (4) of Appendix Table C.3, I show the absence of impact on whether the girl was helped by her father, her mother, her siblings, and her friends respectively. Second, the circular only applied to public schools and, therefore, private school were forbidden to apply the ban. I check whether parents were more likely to send their young daughters to private schools so as to avoid the veil ban, a possibility suggested by De Giorgi et al. (2023) in their panel discussion of Maurin and Navarrete-H. (2023). Results in columns (5) and (6) suggest that this is not the case. The estimated DinD coefficients on whether the respondent went to a private school or to a school in another sector than that of residence are very small and statistically insignificant. Thus, Muslim girls' disengagement from formal schooling does not appear to be driven by their families fearing they expose themselves without the veil.

Relation to previous literature

Maurin and Navarrete-H. (2023) find negative effects of the Bayrou circular on high-school completion rates of Muslim women of African origin. To justify this positive impact, they state that their results are consistent with the existence of a "silent majority" of Muslim girls who do not wish to veil and are under pressure to do so by their family. This argument is often pushed by proponents of veil bans who believe that such policies might liberate Muslim women from oppression. I see two main reasons why this story is unlikely. First, as De Giorgi (2023) argues in a discussion of that paper, this point is very

Table 4.4: Impact of the 1994 ministerial circular on experiences of discrimination

	(1) Any experience	(2) Due to her religion	(3) Due to her origins	(4) At school	(5) Does not trust French school
$\text{Muslim}_i \times \text{Post}_c$	0.086**	0.066***	-0.057**	-0.001	0.075***
	(0.037)	(0.012)	(0.023)	(0.038)	(0.011)
	[0.332]	[0.209]	[0.094]	[0.973]	[0.098]
Muslim_i	0.197***	0.124***	0.229*	0.049*	0.016
	(0.023)	(0.009)	(0.019)	(0.026)	(0.009)
	[0.170]	[0.064]	[0.001]	[0.143]	[0.390]
Mean dep. var.	0.364	0.099	0.791	0.171	0.118
N	7,074	3,623	4,085	7,074	7,008
R^2	0.038	0.089	0.037	0.014	0.010

Note: This Table reports regression estimates of the impact of the issuance of the headscarf ban ministerial circular on self-reported experiences of discrimination of Muslim girls. Means of the dependent variable in the treatment group over pre-1980 cohorts are reported. Control variables are full sets of birthyear, survey waves, and religion dummies and a Muslim-specific linear trend. Standard errors clustered at the religion × born-post-1979 level reported in parentheses. p-values computed using the wild bootstrap procedure are reported in brackets. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

difficult to establish with the Labour Force Surveys (LFS). Indeed, not only is religious affiliation not observed, but the LFS also do not contain much information about the respondents' family or social circles. Furthermore, qualitative evidence from interviews with Muslim women do not support this assertion. The vast majority of Muslim women claim that wearing the veil is their personal choice, some of them even doing so *against* their parents' will (Gaspard and Khosrokhavar, 1995; Institut Montaigne, 2016; Institut Français d'Opinion Publique [IFOP], 2019).

Second, the apparent absence of impact of the ministerial circular on parents' choices discussed above might also suggest that parental pressures do not appear to be the main driver. To further assess to role of parental religious influence, I verify whether the estimated impact differs for women whose parents invested more in transmitting their religion. I have two measures of parental transmission of religion: a question on the self-reported importance of religion in the respondent's education and religious first names.[9] I describe these variables in more detail in Appendix C.1. I report results of this heterogeneity analysis in columns (1) and (2) of Table 4.5. I find that stronger parental religious influence, if anything, slightly reduces the negative effect on schooling. Parental pressures thus do not appear to be the main mediator variable.

While I show that their result of positive effects replicates in the TeO data, I show that when using a better measure of treatment, the impact is of opposite sign. What could explain this discrepancy? The TeO data suggests that individuals of African origin are more religiously mixed than the authors claim. Only 72% of women whose father has

[9]Name-giving has been recognized as an important cultural transmission channel (Fryer and Levitt, 2004; Abramitzky et al., 2020; Algan et al., 2022).

a nationality from an African country are indeed Muslims.[10] One fifth of these second-generation immigrants are atheists and therefore unaffected by the Bayrou circular. It is thus likely that the measurement error induced by using a proxy for religion explains the difference in results. I further test for the possibility of measurement error driving their results by interacting their measure of treatment (having a father who is an African national) with a Muslim dummy. Results are reported in Appendix Table C.4. While the estimates are imprecise in this specification, I find that the coefficient on the interaction is negative. This thereby suggests that using father's nationality at birth as a proxy likely captures positive impacts on other religious groups and not on Muslims.

As discussed in a panel discussion of this study, Barbara Petrongolo mentions that there are probably important group dynamics going on and that negative effects on Muslim girls might still be present (De Giorgi et al., 2023). I am reluctant here to use veiled women as the treatment group because this variable is measured much later in life and veiling should thus be thought as an outcome. I study the long-term impact on veiling in the next section.

Table 4.5: Heterogeneous impacts on high-school completion rates

	(1)	(2)	(3)	(4)	(5)
	Parental religious transmission		Social circle	Parents' labor-force status	
Panel A: High school completion					
$\text{Muslim}_i \times \text{Post}_c$	-0.155***	-0.154***	-0.071*	-0.167***	-0.137***
	(0.036)	(0.034)	(0.036)	(0.035)	(0.036)
	[0.179]	[0.185]	[0.743]	[0.162]	[0.286]
$\text{Muslim}_i \times \text{Post}_c$ $\times\text{Religious education}_i$	0.009**				
	(0.004)				
	[0.359]				
$\text{Muslim}_i \times \text{Post}_c$ $\times\text{Religious first name}_i$		0.091***			
		(0.002)			
		[0.938]			
$\text{Muslim}_i \times \text{Post}_c$ $\times\text{Most friends Muslims}_i$			-0.111***		
			(0.001)		
			[0.544]		
$\text{Muslim}_i \times \text{Post}_c$ $\times\text{Mother was working}_i$				0.062***	
				(0.005)	
				[0.786]	
$\text{Muslim}_i \times \text{Post}_c$ $\times\text{Father was working}_i$					-0.016***
					(0.001)
					[0.990]
Mean dep. var.	0.608	0.599	0.599	0.599	0.599
N	6,986	7,046	7,046	7,046	7,046
R^2	0.040	0.040	0.042	0.040	0.040

Note: This Table reports regression estimates the heterogeneous impact of the issuance of the headscarf ban ministerial circular on educational attainment of Muslim girls. Means of the dependent variable in the treatment group over pre-1980 cohorts are reported. Control variables are full sets of birthyear, survey waves, and religion (or father's origins) dummies and a Muslim-specific linear trend. Standard errors clustered at the religion × born-post-1979 level reported in parentheses. p-values computed using the wild bootstrap procedure are reported in brackets. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In Table 4.5, I also evaluate other sources of heterogeneity. I find, in particular, that

[10] This figure is even lower for Muslim women whose father is born in an African country, the proxy used in Abdelgadir and Fouka (2020). Less than 60% of this population is indeed Muslim.

the negative impact is even stronger for women whose mother was not working when they were 15 years of age. Moreover, the negative effect is concentrated among Muslim women whose friends are mainly Muslims. This could suggest that some form of peer effect might play a role. Muslim women whose (Muslim) friends are struggling to manage religious prescriptions along with their education in a secular context might get more discouraged and disengage from school.

4.5 Long-term outcomes

Having established that the 1994 ministerial circular depressed schooling outcomes of Muslim girls, I now explore potential effects of this policy on social and economic integration of this group in the long-run. I focus on two groups of outcomes, namely economic conditions and religiosity. While it can be expected that the decreased schooling translates into worse economic outcomes in the long-run, the expected impact on social integration is ambiguous. On the one hand, there could be an identity backlash to the assimilationist policy (Sakalli, 2019; Fouka, 2020). On the other hand, the increased emphasis on State secularism (*laïcité*) as “a pillar, even the identity and foundation of the community life” (Andriantsimbazovina et al., 2020, , p. 7) might induce religious minorities to reduce the intensity of their religious life.

Results in Table 4.6 are consistent with the latter interpretation. In columns (4) to (6), I estimate the impact of the Bayrou circular on religiosity outcomes, namely the importance of religion in the respondent’s life, the use of religion to self-define, and a religiosity index (see Section 4.3.1). I find that treated Muslim women display lower levels of religiosity later in life. Also consistent with the absence of religious backlash, I find no significant impact on veiling behavior (on the wearing of conspicuous religious symbols).

In evaluate potential impacts on economic outcomes in columns (1) to (3). I find weak evidence of a decrease in the probability to be employed at the time of the survey. However, this might be due to the fact that treated women are more likely to have children. These coefficients are also far from being statistically significant and I thus refrain from over-interpreting these findings.

4.6 Conclusion

In this paper, I revisit previous contradictory evidence on the French headscarf ban using rare rich observational data on religion in France. In 1994, the Minister of Education François Bayrou issued a circular asking school principals to prohibit the wearing of conspicuous religious symbols in their establishment. Many schools readily implement

Table 4.6: Impact of the 1994 ministerial circular on long-term outcomes

	Economic situation			Religiosity			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dep. var:	Employed	Lives in a couple	Has children	Religion is very imp.	Religion to self-define	Religiosity index	Veiling†
$\text{Muslim}_i \times \text{Post}_c$	-0.042	0.025	0.056	-0.055	-0.060***	-0.073***	0.025
	(0.031)	(0.039)	(0.044)	(0.033)	(0.009)	(0.017)	(0.035)
	[0.643]	[0.657]	[0.649]	[0.633]	[0.009]	[0.136]	[0.741]
Muslim_i	-0.191***	-0.025	0.073*	0.342***	0.173***	0.282***	0.135***
	(0.020)	(0.044)	(0.038)	(0.024)	(0.007)	(0.016)	(0.024)
	[0.247]	[0.751]	[0.369]	[0.054]	[0.002]	[0.018]	[0.248]
Mean dep. var.	0.642	0.640	0.788	0.431	0.213	0.252	0.150
N	7,074	7,074	7,074	7,065	6,967	7,074	7,064
R^2	0.075	0.061	0.225	0.273	0.102	0.173	0.147

Note: This Table reports regression estimates of the impact of the issuance of the headscarf ban ministerial circular on long-term outcomes of Muslim girls. Means of the dependent variable in the treatment group over pre-1980 cohorts are reported. Control variables are full sets of birthyear, survey waves, and religion dummies and a Muslim-specific linear trend. Standard errors clustered at the religion × born-post-1979 level reported in parentheses. p-values computed using the wild bootstrap procedure are reported in brackets. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
† Veiling is a dichotomous variable taking the value of one if the individual wears a conspicuous religious symbol in public spaces.

prohibitions as shown in the large number of reported cases of conflict, in large part due to the Islamic veil. I find that this disruption in the schooling environment for Muslim girls is associated with a large decline in the educational attainment of exposed cohorts. The average impact of 25% of the mean delayed the catching-up process of Muslim girls over their counterparts by about five cohorts. The impact on the number of years of schooling persists over the medium-run. I provide suggestive evidence that the negative impact of the ban operates through heightened discrimination against Muslims and increased mistrust of the French school rather than via parents' reactions to the prohibition.

While using an actual measure of the treatment group improves upon previous studies of the reform, it comes at the cost of much smaller sample sizes and thus of more statistical uncertainty. Most of the estimates are statistically significant under different clustering strategies for standard errors, but do not remain so when employing the wild bootstrap. Therefore, interpreting these results requires some caution. Nevertheless, this study highlights that improperly measuring the group targeted by an assimilationist policy might lead to incorrect conclusions. I find that the positive effects on educational attainment documented in Maurin and Navarrete-H. (2023) appear to be attributable to individuals of African origin who are not Muslim. It is therefore unlikely that they capture any impact of the religious prohibition.[11]

My results suggest that forced assimilation policies such as headscarf bans are not a successful tool to foster integration of minorities and immigrants. This result is consistent

[11] In the Labour Force Surveys, there seems to be a strong pre-trend in the treatment group. When looking at Figure 1 in Maurin and Navarrete-H. (2023), the results on the high-school completion rate appear to be entirely driven by a stagnation in the control group rather than any change in the underlying trend in the "Muslim group".

with Fouka (2020) and Sakalli (2019) who similarly find that these types of policies might backfire. However, at the same time, some well-intentioned integration policies might also hamper assimilation in contexts in which minorities are strongly attached to their traditional norms (Dahl et al., 2022). Therefore, in a context of increased global migrations from countries with non-Western cultures, more work is needed to better understand which policies can foster integration. If both the carrot and the stick are insufficient or might backfire, governments may have to enact innovative policies and carefully study their economic impacts.

Appendix C

Supplementary Material for Chapter 4

C.1 Measurement of individual religiosity and communitarian pressures

The TeO datasets contains rich information on respondents' religious life. I first describe the variables I use to proxy for individual religiosity. I then detail how I combine those multiple measures into meaningful indices through a measurement system.

Individual religiosity. I measure individual religiosity using survey questions on the frequency of attendance of religious ceremonies, the self-reported importance of religion in the respondent's life, whether she uses her religion to self-identify, the respect of religious dietary restrictions, and religious marriage. I list details of these variables below:

Variable name	Values	Question	Type
attendance of religious ceremonies	never; for familial ceremonies only; for religious feasts only; one or twice a month; weekly	"How often do you attend religious ceremonies?"	ordinal
importance of religion in respondent's life	no importance; a little; quite important; very important	"What importance do you give to religion in your life today?"	ordinal
uses religion to self-identify	yes; no	"Among the following characteristics, which ones define you best? [...] Your religion?"	indicator
respect of dietary restrictions	never; sometimes; always; none (coded as a dummy if "always")	"In your daily life, do you respect your religion's dietary restrictions?"	indicator
religious marriage	yes; no	"Did you and your husband do a religious wedding?"	indicator

Measurement system. Since there is no natural way to combine the ordinal and indicator variables described above into meaningful indices, I formulate a measurement system. I am interested in a latent variable, *individual religiosity*, which I assume loads into the proxies listed above. I interpret those proxies as noisy measures of the associated unobserved, underlying concept. Denote by Z the vectors of proxies for individual religiosity. I assume ordinal relationships between measures $\{Z\}$ and the underlying factor $\text{IndivReligiosity}_i$ such that:

$$Z_{i,j} = \mu_{1,j} + \lambda_j \text{ IndivReligiosity}_i + \varepsilon_{i,j} \tag{C.1}$$

where ε is a measurement error assumed to be i.i.d. and to follow an ordinal logistic distribution. As the latent factor does not have a natural scale or location, to simplify interpretations, I normalize the means of $\text{IndivReligiosity}_i$ to zero, and its variance to one. I then predict the latent factor for each individual by calculating its empirical Bayes mean (Skrondal and Rabe-Hesketh, 2009).

Parental religious transmission. I measure vertical religious pressures using two variables, namely the self-reported importance of religion in the respondent's education and religious name-giving.

Variable name	Values	Question	Type
importance of religion in education	no importance; a little important; quite important; very important	"What importance did religion have in the education you received in your family?"	ordinal
religious first name	yes; no	constructed by the author using respondent's first name	indicator

I classify as religious the names of the Islamic prophet's wives, Khadija, Sawda, Aicha, Hafsa, Zainab, Hind, Juwairiya, Safiya, Ramla, and Maimuna (Morsy, 1989); and of his daughter Fatima. Variations in spelling are permitted. For male first names, I follow Sakalli (2019) by considering a name as religious if it is a variation of the prophet's name (Mohamed in French) or if it begins with "Abd-" ("servant of…" in Arabic).

C.2 Additional Tables and Figures

Table C.1: Summary statistics

Statistic:	Mean	SD	N
Muslim women			
Demographics			
Age	32.933	7.539	1,860
Born in metropolitan France	0.853	0.354	1,860
Married	0.507	0.500	1,860
Highest degree completed			
No degree	0.13	0.34	1,619
CAP/BEP	0.25	0.43	1,619
High school (*bacc*)	0.28	0.45	1,619
Higher education	0.33	0.43	1,619
Economic outcomes			
Employed	0.585	0.493	1,860
Unemployed	0.176	0.381	1,860
Inactive	0.187	0.390	1,860
Years of schooling	14.71	3.14	1,440
Religious outcomes			
Religiosity index	0.343	0.791	1,860
Religion is very important in life	0.469	0.499	1,854
Attends religious ceremonies regularly	0.065	0.247	1,856
Had conflict over religion with parents	0.177	0.382	1,860
Most friends are of the same religion	0.700	0.458	1,860
Religion was very important in education	0.432	0.495	1,852
Partner of same religion	0.553	0.497	1,860
Wears a religious symbol	0.286	0.452	1,854
Wears a conspicuous religious symbol	0.192	0.394	1,854
Non-Muslim women			
Demographics			
Age	35.039	7.364	5,898
Born in metropolitan France	0.919	0.273	5,898
Married	0.364	0.481	5,898
Highest degree completed			
No degree	0.06	0.23	5,453
CAP/BEP†	0.19	0.39	5,453
High school (*bacc*)	0.22	0.41	5,453
Higher education	0.53	0.39	5,453
Economic outcomes			
Employed	0.804	0.397	5,898
Unemployed	0.096	0.295	5,898
Inactive	0.064	0.244	5,898
Years of schooling	15.73	3.01	4,879
Religious outcomes			
Religiosity index	-0.25	0.61	5,472
Religion is very important in life	0.135	0.342	2,886
Attends religious ceremonies regularly	0.140	0.347	2,886
Had conflict over religion with parents	0.148	0.355	5,898
Most friends are of the same religion	0.805	0.396	5,898
Religion was very important in education	0.139	0.346	5,831
Partner of same religion	0.226	0.418	5,898
Wears a religious symbol	0.234	0.424	2,888
Wears a conspicuous religious symbol	0.002	0.049	2,888

Note: The data source is the Trajectories and Origins (TeO) surveys of 2008-2009 and 2019-2020.

† The CAP (*Certificat d'Aptitude Professionnelle*) and the BEP (*Brevet d'Études Professionnelles*) are vocational high-school degrees aimed at acquiring skills specific to a chosen occupation (such as plumbing, butchery, or bakery).

Figure C.1: Evolution of educational attainment of Muslim and non-Muslim men across birth cohorts, 1971-1987

(a) High-school completion (b) Years of schooling

Note: Evolution of the high-school completion rate across birth cohorts of Muslim and non-Muslim men born in France. On the right of the vertical line are cohorts who are subject to the 1994 Bayrou circular asking school principals to prohibit the headscarf in schools.

Table C.2: Impact of the 1994 ministerial circular on educational attainment, additional controls

Measure of treatment:	(1) Muslim men
Panel A: High school completion	
$\text{Muslim}_i \times \text{Post}_c$	0.021
	(0.019)
	[0.477]
Muslim_i	-0.042**
	(0.018)
	[0.410]
Mean dep. var.	0.516
N	6,485
R^2	0.066
Panel B: Years of post-elementary schooling	
$\text{Muslim}_i \times \text{Post}_c$	-0.054
	(0.164)
	[0.875]
Muslim_i	-0.079
	(0.102)
	[0.534]
Mean dep. var.	10.94
N	5,784
R^2	0.063

Note: This Table reports regression estimates of the impact of the issuance of the headscarf ban ministerial circular on educational attainment of Muslim men using additional control variables. Means of the dependent variable in the treatment group over pre-1980 cohorts are reported. Control variables are full sets of birthyear, survey waves, and religion (or father's origins) dummies, a Muslim-specific linear trend, dummies for whether the mother and the father worked when the girl was 15 years old, and a full set of *départements* of residence when 15 years old dummies. Standard errors clustered at the religion × born-post-1979 level reported in parentheses. p-values computed using the wild bootstrap procedure are reported in brackets. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: Impact of the 1994 ministerial circular on Muslim girls' relatives' investments in their education

	Was helped at school by				Type of school	
	(1) Father	(2) Mother	(3) Siblings	(4) Friends	(5) Private	(6) Not in sector
$\text{Muslim}_i \times \text{Post}_c$	-0.014	0.057	0.066	-0.086	0.009	-0.026
	(0.067)	(0.058)	(0.044)	(0.060)	(0.018)	(0.055)
	[0.881]	[0.483]	[0.345]	[0.528]	[0.725]	[0.744]
Muslim_i	-0.173***	-0.436***	0.343***	0.098**	-0.084**	-0.081*
	(0.041)	(0.037)	(0.025)	(0.043)	(0.013)	(0.038)
	[0.250]	[0.054]	[0.037]	[0.179]	[0.113]	[0.470]
Mean dep. var.						
N	6,842	6,951	6,596	7,011	7,074	7,063
R^2	0.032	0.093	0.058	0.018	0.192	0.014

Note: This Table reports regression estimates of the impact of the issuance of the headscarf ban ministerial circular on Muslim girls' relatives' investments in their education. Means of the dependent variable in the treatment group over pre-1980 cohorts are reported. Control variables are full sets of birthyear, survey waves, and religion dummies and a Muslim-specific linear trend. Standard errors clustered at the religion $\times$ born-post-1979 level reported in parentheses. p-values computed using the wild bootstrap procedure are reported in brackets. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.4: Heterogeneity analysis of the impact on women of African origin

Measure of treatment:	(1) Father's nationality
Panel A: High school completion	
African origin$_i$ × Post$_c$	0.072*
	(0.036)
	[0.139]
African origin$_i$ × Post$_c$ ×Muslim$_i$	-0.034
	(0.056)
	[0.559]
Mean dep. var.	0.631
N	6,960
R^2	0.047
Panel B: Years of post-elementary schooling	
African origin$_i$ × Post$_c$	0.546
	(0.341)
	[0.147]
African origin$_i$ × Post$_c$ ×Muslim$_i$	-0.366
	(0.539)
	[0.551]
Mean dep. var.	15.03
N	6,223
R^2	0.050

Note: Means of the dependent variable in the treatment group over pre-1980 cohorts are reported. Control variables are full sets of birthyear, survey waves, and religion (or father's origins) dummies, a Muslim-specific linear trend. Standard errors clustered at the father's origin × born-post-1979 level reported in parentheses. p-values computed using the wild bootstrap procedure are reported in brackets. Level of statistical significance: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Chapter 5

Conclusion

This thesis presents three essays analyzing female labor-force participation and economic integration of women in developed countries. We focus on how different policies interact with key trade-offs faced by women when choosing their labor supply. In the first essay, we focus on public childcare provision. We leverage variation in price and quantity supplied from the introduction of universal low-fee childcare in the Canadian province of Québec in 1997. We evaluate the impact of this policy on social welfare while taking into account two overlooked channels through which such policy can impact welfare. That is, we measure non-pecuniary benefits and we account for the fact that this reform is a non-marginal change in the economic environment. In the second and third essays, we study the joint decision of veiling and economic integration for Muslim women in France. The second essay combines descriptive evidence and economic theory to understand the main motivations driving these joint choices. In the third essay, I use the French headscarf ban in schools to evaluate the causal impact of policies restricting the wearing of religious symbols on economic integration of this population.

The first essay provides several lessons for empirical welfare analysis of universal preschool programs. The Québec childcare reform is often seen as problematic due to evidence of negative short-term impacts on eligible children by Baker et al. (2008, 2019) and because its cost exceeds tax returns for the government (Haeck et al., 2015). However, our results provide some nuance on both arguments because we document that (i) key sources of gains have been overlooked in previous cost-benefit analyses of the policy and (ii) impacts on children might not be as serious as feared. We find that universal childcare provision can provide substantial returns on initial public investment, in particular in the form of non-pecuniary benefits for mothers. In addition, the results suggest the reform induced little harm on children in the long-run despite the fast expansion of the market. A second lesson of this paper is that sufficient-statistic estimators of beneficiaries' willingness-to-pay applied to non-marginal reforms may overlook key welfare gains of universal programs. We find that this strategy underestimates the policy's benefits by more than half. Last, our study highlights the crucial role of increasing availability of childcare at the local level. While the previous literature focused on the affordability channel, we find that the government could have achieved higher welfare gains by channeling more resources towards opening spots rather than lowering childcare fees as much. This is because, in our estimated model, families have a larger willingness-to-pay for increased coverage than for a price reduction.

In the second and third essays, we find that policies restricting the wearing of religious symbols in the public sphere impede the economic integration of Muslim women in France. In the second essay, we find that veiling imposes a strong economic cost on Muslim women. The negative association between veiling and economic participation contrasts with the existing economic theory of veiling in Muslim-majority countries. We show that an extension of the theoretical framework of Carvalho (2013) which includes labor-market discrimination against veiled women can rationalize the Muslim-minority context. The data are consistent with this theoretical extension: a structural estimation of the model suggests that the main channel explaining that veiled women work less is the economic penalty when veiling. The results thus suggest that secular regulations limiting work opportunities for veiled women likely amplify economic segregation of the population.

To obtain a more causal assessment of this possibility, the third essay investigates the impacts of the French headscarf ban in schools. The results are in line with what the second essay might suggest. That is, we find that the French headscarf ban induced a strong decline in educational attainment of Muslim girls who reached the age of veiling after the ban. Digging into the mechanisms through which the ban reduced educational success, I find suggestive evidence that it operates through increased discrimination against Muslim girls rather than through reduced parental investments or increased religious pressures. This paper also shows how using a proxy for religion based on parents' origins as in previous studies substantially alters the estimated impacts.

Taken together, the findings in this thesis advise caution in the design of policies aimed at favoring female labor-force participation. In the case of childcare policy, the public debate has typically focused on affordability of the service. The Canadian experience, where both the Québec governments of the late 1990s and the current federal government insisted on charging low fees, is particularly telling. The first essay rather suggests that more efforts and resources should be invested in ensuring that the supply of care is sufficient to meet the growing demand. While affordability remains important, policymakers should also devote considerable attention to daycare supply and include measures to facilitate entry into the market in their policy mix. In the case of integration policy, the second and third essays suggest that "assimilationist" policies further impede the already-difficult economic integration of minorities in developed countries. Further restricting religious expression in public is therefore likely to reduce social welfare rather than fostering integration of the Muslim population. Our theoretical results rather suggest that policymakers who wish to improve the economic conditions of these groups should try to remove the existing barriers they face on the labor-market such as discriminatory treatments.

These three essays have some limitations that could serve as a basis for future research and extensions. In the first essay, the analysis focuses on short-term impacts of the policy, both because of data constraints and the difficulty to incorporate changes in policy features a few years after implementation. A natural extension would be to consider the welfare impact of the childcare reform over a longer horizon in a dynamic setting. Analyzing the types of jobs mothers could obtain thanks to the reform would also be of particular interest. A second research avenue could be to investigate how the policy affected behavior of the household, beyond the labor-supply responses, and how these changes matter for welfare. In particular, other documented impacts of the policy concern parents' health, in particular mental health outcomes. A model that addresses these concerns could be useful since parental mental health matters for child development and is a source of fiscal externality. In the second chapter, one key limit is surely that we are not able to identify a causal relationship. Therefore, despite our unusually large set of controls, it is possible that selection into the veiling status drives our results. While exogenous variation in the decision to veil might be difficult to obtain, a more robust assessment of our findings could be made by using events that spurred discrimination against Muslims but not the religious environment such as terrorist attacks. Last, in the last chapter, the key limitation is certainly the small sample size that the data offers. Despite using the most ambitious survey of immigrant populations and their descendants in France, we cannot match the sample size of previous papers. Thus, appropriately observing the treatment and control groups in this setting comes at the cost of potentially increased statistical noise. For the study of long-term impacts on labor-market outcomes, matching the survey data with administrative sources could be of great value. Another possibility would be to use Turkish data – at the cost of studying a Muslim-majority country –, which imposed a veil ban in the past and for which data on religion is more readily available.

Bibliography

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35.

Abdelgadir, A. and Fouka, V. (2020). Political secularism and Muslim integration in the West: Assessing the effects of the French Headscarf Ban. *American Political Science Review*, 114(3):707–723.

Abdelhadi, E. (2019). The hijab and muslim women's employment in the united states. *Research in Social Stratification and Mobility*, 61:26–37.

Abramitzky, R., Boustan, L., and Eriksson, K. (2020). Do immigrants assimilate more slowly today than in the past? *American Economic Review: Insights*, 2(1):125–41.

Adida, C. L., Laitin, D. D., and Valfort, M.-A. (2010). Identifying barriers to muslim integration in france. *Proceedings of the National Academy of Sciences*, 107(52):22384–22390.

Adida, C. L., Laitin, D. D., and Valfort, M.-A. (2016). *Why Muslim integration fails in Christian-heritage societies*. Harvard University Press.

Ajbli, F. (2011). *Les Françaises musulmanes face à l'emploi: le cas des pratiquantes "voilées" dans la métropole lilloise*. PhD thesis, Paris, EHESS.

Akee, R. and Clark, L. (2023). Universal preschool lottery admissions and its effects on long-run earnings and outcomes. Working papers, U.S. Census Bureau, Center for Economic Studies.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The quarterly journal of economics*, 115(3):715–753.

Aksoy, O. (2017). Motherhood, sex of the offspring, and religious signaling. *Sociological Science*, 4:511–527.

Aksoy, O. and Gambetta, D. (2016). Behind the veil: the strategic use of religious garb. *European Sociological Review*, 32(6):792–806.

Aksoy, O. and Gambetta, D. (2021). The politics behind the veil. *European Sociological Review*, 37(1):67–88.

Algan, Y., Malgouyres, C., Mayer, T., and Thoenig, M. (2022). The economic incentives of cultural transmission: Spatial evidence from naming patterns across france. *The Economic Journal*, 132(642):437–470.

Almond, D., Currie, J., and Duque, V. (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature*, 56(4):1360–1446.

Andresen, M. E. (2019). Child care for all? treatment effects on test scores under essential heterogeneity. Technical report, Working Paper, Statistics Norway.

Andresen, M. E. and Havnes, T. (2019). Child care, parental labor supply and tax revenue. *Labour Economics*, 61:101762.

Andriantsimbazovina, J., Kabou, P., et al. (2020). *Laïcité et défense de l'État de droit*. Presses de l'Université Toulouse 1 Capitole.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 24(2):3–30.

Apps, P., Kabátek, J., Rees, R., and van Soest, A. (2016). Labor supply heterogeneity and demand for child care of mothers with young children. *Empirical Economics*, 51(4):1641–

1677.

Atasoy, Y. (2006). Governing women's morality: a study of islamic veiling in Canada. *European Journal of Cultural Studies*, 9(2):203–221.

Atkin, D., Colson-Sihra, E., and Shayo, M. (2021). How do we choose our identity? a revealed preference approach using food consumption. *Journal of Political Economy*, 129(4):1193–1251.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020a). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review*, 110(1):48–85.

Attanasio, O., Meghir, C., and Nix, E. (2020b). Human capital development and parental investment in india. *The Review of Economic Studies*, 87(6):2511–2541.

Bailey, M. J., Hoynes, H. W., Rossin-Slater, M., and Walker, R. (forthcoming). Is the social safety net a long-term investment? large-scale evidence from the food stamps program. *Review of Economic Studies*.

Bailey, M. J., Sun, S., and Timpe, B. (2021). Prep school for poor kids: The long-run impacts of head start on human capital and economic self-sufficiency. *American Economic Review*, 111(12):3963–4001.

Baird, S., Hicks, J. H., Kremer, M., and Miguel, E. (2016). Worms at work: Long-run impacts of a child health investment. *The quarterly journal of economics*, 131(4):1637–1680.

Baker, M. (2011). Innis lecture: Universal early childhood interventions: what is the evidence base? *Canadian Journal of Economics/Revue canadienne d'économique*, 44(4):1069–1105.

Baker, M., Gruber, J., and Milligan, K. (2005). Universal childcare, maternal labor supply, and family well-being. Working Paper 11832, National Bureau of Economic Research.

Baker, M., Gruber, J., and Milligan, K. (2008). Universal child care, maternal labor supply, and family well-being. *Journal of Political Economy*, 116(4):709–745.

Baker, M., Gruber, J., and Milligan, K. (2019). The long-run impacts of a universal child care program. *American Economic Journal: Economic Policy*, 11(3):1–26.

Baril, R., Lefebvre, P., and Merrigan, P. (2000). Quebec family policy: impact and options. *Choices: Family Policy*, 6(1).

Barr, A. and Gibbs, C. R. (2022). Breaking the cycle? intergenerational effects of an antipoverty program in early childhood. *Journal of Political Economy*, 130(12).

Bastian, J. E. and Jones, M. R. (2021). Do eitc expansions pay for themselves? effects on tax revenue and government transfers. *Journal of public economics*, 196:104355.

Beauchemin, C., Hamel, C., and Simon, P. (2016). *Trajectoires et origines: enquête sur la diversité des populations en France*. INED éditions.

Beauchemin, C., Ichou, M., and Simon, P. (2023). Trajectoires et origines 2019-2020 (teo2): présentation d'une enquête sur la diversité des populations en france. *Population (édition française)*, 78(1):11–28.

Becker, G. S. (1965). A theory of the allocation of time. *The economic journal*, 75(299):493–517.

Becker, G. S. (1991). *A treatise on the family: Enlarged edition*. Harvard university press.

Bellemare, M. F., Novak, L., and Steinmetz, T. L. (2015). All in the family: Explaining the persistence of female genital cutting in west africa. *Journal of Development Economics*, 116:252–265.

Belzil, C. and Poinas, F. (2010). Education and early career outcomes of second-generation immigrants in france. *Labour Economics*, 17(1):101–110.

Benzer, T. (2022). Removing cultural barriers to education: State-run islamic schools and girls' education in turkey. Unpublished manuscript.

Bergolo, M. and Cruces, G. (2021). The anatomy of behavioral responses to social assistance when informal employment is high. *Journal of Public Economics*, 193:104313.

Berlinski, S., Ferreyra, M. M., Flabbi, L., and Martin, J. D. (2020). Child care markets, parental labor supply, and child development. Working Paper 12904, IZA Institute of Labor

Economics.
Bernal, R. (2008). The effect of maternal employment and child care on children's cognitive development. *International Economic Review*, 49(4):1173–1209.
Bernal, R. and Keane, M. P. (2010). Quasi-structural estimation of a model of childcare choices and child cognitive ability production. *Journal of Econometrics*, 156(1):164–189.
Biavaschi, C., Giulietti, C., and Siddique, Z. (2017). The economic payoff of name americanization. *Journal of Labor Economics*, 35(4):1089–1116.
Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2011). Ethnic identity and labour market outcomes of immigrants in europe. *Economic Policy*, 26(65):57–92.
Bisin, A., Patacchini, E., Verdier, T., and Zenou, Y. (2016). Bend it like beckham: Ethnic identity and integration. *European Economic Review*, 90:146–164.
Bisin, A. and Verdier, T. (2000). "beyond the melting pot": cultural transmission, marriage, and the evolution of ethnic and religious traits. *The Quarterly Journal of Economics*, 115(3):955–988.
Blau, D. M. (2021). The effects of universal preschool on child and adult outcomes: A review of recent evidence from europe with implications for the united states. *Early Childhood Research Quarterly*, 55:52–63.
Bolt, U., French, E., Maccuish, J. H., and O'Dea, C. (2021). The intergenerational elasticity of earnings: exploring the mechanisms. Cepr discussion paper no. dp15975.
Bouzar, D. and Bouzar, L. (2009). *Allah a-t-il sa place dans l'entreprise?* Albin Michel.
Bravo, D., Hojman, A., and Rodríguez (2022). Welfare effects of childcare policies. Working paper.
Brenøe, A. A., Canaan, S. P., Harmon, N. A., and Royer, H. N. (2020). Is parental leave costly for firms and coworkers? Working Paper 26622, National Bureau of Economic Research.
Brink, A., Nordblom, K., and Wahlberg, R. (2007). Maximum fee versus child benefit: a welfare analysis of swedish child-care fee reform. *International Tax and Public Finance*, 14:457–480.
Brodeur, A., Cook, N., and Heyes, A. (2020). Methods matter: P-hacking and publication bias in causal analysis in economics. *American Economic Review*, 110(11):3634–3660.
Brodeur, A., Lé, M., Sangnier, M., and Zylberberg, Y. (2016). Star wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1):1–32.
Brodeur, A., Mikola, D., Cook, N., et al. (2024). Mass reproducibility and replicability: A new hope. I4R Discussion Paper Series 107.
Buera, F. J., Kaboski, J. P., and Townsend, R. M. (2023). From micro to macro development. *Journal of Economic Literature*, 61(2):471–503.
Cabral, M. and Dillender, M. (2021). The impact of benefit generosity on workers' compensation claims: Evidence and implications. Working paper.
Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The review of economics and statistics*, 90(3):414–427.
Cantoni, D., Chen, Y., Yang, D. Y., Yuchtman, N., and Zhang, Y. J. (2017). Curriculum and ideology. *Journal of political economy*, 125(2):338–392.
Carta, F. and Rizzica, L. (2018). Early kindergarten, maternal labor supply and children's outcomes: evidence from italy. *Journal of Public Economics*, 158:79–102.
Carvalho, J.-P. (2013). Veiling. *The Quarterly Journal of Economics*, 128(1):337–370.
Carvalho, J.-P., Koyama, M., and Sacks, M. (2017). Education, identity, and community: lessons from jewish emancipation. *Public Choice*, 171:119–143.
Carvalho, J.-P., Koyama, M., and Williams, C. (2024). Resisting education. *Journal of the European Economic Association*, page jvae008.
Cascio, E. U. (2009). Maternal labor supply and the introduction of kindergartens into american public schools. *The Journal of Human Resources*, 44(1):140–170.
Cascio, E. U. (2023). Does universal preschool hit the target? program access and preschool impacts. *Journal of Human Resources*, 58(1):1–42.

Chan, M. K. and Liu, K. (2018). Life-cycle and intergenerational effects of child care reforms. *Quantitative Economics*, 9(2):659–706.

Chaparro, J., Sojourner, A., and Wiswall, M. J. (2020). Early childhood care and cognitive development. Working Paper 26813, National Bureau of Economic Research.

Chérifi, H. (2004). Application de la loi du 15 mars 2004 sur le port des signes religieux ostensibles dans les établissements d'enseignement publics.

Cintas, C., Gosse, B., and Vatteville, É. (2012). Quand l'identité religieuse devient une préoccupation du management des ressources humaines. In *Management et religion: Décryptage d'un lien indéfectible*, pages 83–98. EMS Editions.

Cohen, M. A. (2020). *The costs of crime and justice*. Routledge.

Cohodes, S. R. and Goodman, J. S. (2014). Merit aid, college quality, and college completion: Massachusetts' adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6(4):251–85.

Corekcioglu, G. (2021). Unveiling the effects of a headscarf ban: Evidence from municipal jobs in turkey. *Journal of Comparative Economics*, 49(2):382–404.

Cornelissen, T., Dustmann, C., Raute, A., and Schönberg, U. (2018). Who benefits from universal child care? estimating marginal returns to early child care attendance. *Journal of Political Economy*, 126(6):2356–2409.

Cornwell, C., Mustard, D. B., and Sridhar, D. J. (2006). The enrollment effects of merit-based financial aid: Evidence from georgia's hope program. *Journal of Labor Economics*, 24(4):761–786.

Cunha, F. and Heckman, J. (2007). The technology of skill formation. *American economic review*, 97(2):31–47.

Cunha, F. and Heckman, J. J. (2008). Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation. *Journal of human resources*, 43(4):738–782.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Currie, J. and Gruber, J. (1996). Health Insurance Eligibility, Utilization of Medical Care, and Child Health*. *The Quarterly Journal of Economics*, 111(2):431–466.

Cutler, D. M. and Gruber, J. (1996). Does Public Insurance Crowd out Private Insurance?*. *The Quarterly Journal of Economics*, 111(2):391–430.

Dahl, G. B., Felfe, C., Frijters, P., and Rainer, H. (2022). Caught between cultures: Unintended consequences of improving opportunity for immigrant girls. *The Review of Economic Studies*, 89(5):2491–2528.

Daruich, D. (2022). The macroeconomic consequences of early childhood development policies. Working paper.

Dave, D., Decker, S. L., Kaestner, R., and Simon, K. I. (2015). The effect of medicaid expansions in the late 1980s and early 1990s on the labor supply of pregnant women. *American Journal of Health Economics*, 1(2):165–193.

De Chaisemartin, C. and d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9):2964–96.

De Giorgi, G. (2023). Giacomo De Giorgi discussion of: Religious symbols in schools. *Economic Policy*, 38(113):99–100.

De Giorgi, G., Fack, G., Galbiati, R., Lim, C., Limodio, N., and Mejean, I. (2023). Panel discussion of: Religious symbols in schools. *Economic Policy*, 38(113):103–105.

De Groote, O. and Rho, M. (2023). Matching disadvantaged children to daycare: Evidence from a centralized platform. Unpublished manuscript, Toulouse School of Economics.

Del Boca, D., Flinn, C., and Wiswall, M. (2014). Household choices and child development. *Review of Economic Studies*, 81(1):137–185.

Deshpande, M. (2016). Does welfare inhibit success? the long-term effects of removing low-income youth from the disability rolls. *American Economic Review*, 106(11):3300–3330.

Deshpande, M., Gross, T., and Su, Y. (2021). Disability and distress: The effect of disability programs on financial outcomes. *American Economic Journal: Applied Economics*, 13(2):151–78.

Deshpande, M. and Mueller-Smith, M. (2022). Does welfare prevent crime? the criminal justice outcomes of youth removed from ssi. *The Quarterly Journal of Economics*, 137(4):2263–2307.

Ding, W., Kottelenberg, M. J., and Lehrer, S. F. (2020). Anticipating the (un) expected: Evidence from introducing a universal childcare policy with a shortage of spaces. Technical report, Working Paper.

Dreber, A. and Johannesson, M. (2023). A framework for evaluating reproducibility and replicability in economics. *Available at SSRN 4458153*.

Droogsma, R. A. (2007). Redefining hijab: American muslim women's standpoints on veiling. *Journal of Applied Communication Research*, 35(3):294–319.

Drouin, C., Bigras, N., Fournier, C., Desrosiers, H., and Bernard, S. (2004). Grandir en qualité 2003. enquête québécoise sur la qualité des services de garde éducatifs. Technical report, Institut de la statistique du Québec.

Drydakis, N. (2013). The effect of ethnic identity on the employment of immigrants. *Review of Economics of the Household*, 11(2):285–308.

Duchini, E. and Van Effenterre, C. (2022). School schedule and the gender pay gap. *Journal of Human Resources*, forthcoming.

Duncan, G., Kalil, A., Mogstad, M., and Rege, M. (2023). Investing in early childhood development in preschool and at home. In *Handbook of the Economics of Education, Volume 6*, pages 235–297. Elsevier.

Evans, D. K., Jakiela, P., and Knauer, H. A. (2021). The impact of early childhood interventions on mothers. *Science*, 372(6544):794–796.

Fan, X. and Wu, L. (2022). The shaping of a gender norm: Marriage, labor, and foot-binding in historical china. Working paper.

Felfe, C. and Lalive, R. (2018). Does early child care affect children's development? *Journal of Public Economics*, 159:33–53.

Fernández-Reino, M., Di Stasio, V., and Veit, S. (2022). Discrimination unveiled: A field experiment on the barriers faced by muslim women in germany, the netherlands, and spain. *European Sociological Review*.

Finkelstein, A. and Hendren, N. (2020). Welfare analysis meets causal inference. *Journal of Economic Perspectives*, 34(4):146–67.

Firpo, S., Fortin, N. M., and Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica*, 77(3):953–973.

Fitzpatrick, M. (2010). Preschoolers enrolled and mothers at work? the effects of universal prekindergarten. *Journal of Labor Economics*, 28(1):51–85.

Fort, M., Ichino, A., and Zanella, G. (2020). Cognitive and noncognitive costs of day care at age 0–2 for children in advantaged families. *The Journal of Political Economy*, 128(1):158–205.

Fortin, P., Godbout, L., and St-Cerny, S. (2013). L'impact des services de garde à contribution réduite du québec sur le taux d'activité féminin, le revenu intérieur et les budgets gouvernementaux. *Revue Interventions économiques. Papers in Political Economy*, 47.

Fouka, V. (2020). Backlash: The unintended effects of language prohibition in U.S. schools after World War I. *The Review of Economic Studies*, 87(1):204–239.

Fryer, R. G. and Levitt, S. D. (2004). The causes and consequences of distinctively black names. *The Quarterly Journal of Economics*, 119(3):767–805.

Ganimian, A. J., Muralidharan, K., and Walters, C. R. (2021). Augmenting state capacity for child development: Experimental evidence from india. Working Paper 28780, National Bureau of Economic Research.

García, J. L. and Heckman, J. J. (2022a). On criteria for evaluating social programs. Working Paper 30005, National Bureau of Economic Research.

García, J. L. and Heckman, J. J. (2022b). Three criteria for evaluating social programs. *Journal of Benefit-Cost Analysis*, 13(3):281–286.

García, J. L., Heckman, J. J., Leaf, D. E., and Prados, M. J. (2020). Quantifying the life-cycle benefits of an influential early-childhood program. *Journal of Political Economy*, 128(7):2502–2541.

García, J. L., Heckman, J. J., and Ronda, V. (2023). The lasting effects of early-childhood education on promoting the skills and social mobility of disadvantaged african americans and their children. *Journal of Political Economy*, 131(6):1477–1506.

Gaspard, F. and Khosrokhavar, F. (1995). *Le foulard et la République*. FeniXX.

Gathmann, C. and Garbers, J. (2023). Citizenship and integration. *Labour Economics*, 82:102343.

Gauthier, A. H., Smeeding, T. M., and Furstenberg Jr, F. F. (2004). Are parents investing less time in children? trends in selected industrialized countries. *Population and development review*, 30(4):647–672.

Gelbach, J. B. (2002). Public schooling for young children and maternal labor supply. *American Economic Review*, 92(1):307–322.

Gertler, P., Galiani, S., and Romero, M. (2018). How to make replication the norm. *Nature*, 554(7693):417–419.

Giesecke, M. and Jäger, P. (2021). Pension incentives and labor supply: Evidence from the introduction of universal old-age assistance in the uk. *Journal of Public Economics*, 203:104516.

Goldin, C. (2006). The quiet revolution that transformed women's employment, education, and family. *American economic review*, 96(2):1–21.

Gray, C., Leive, A., Prager, E., Pukelis, K. B., and Zaki, M. (forthcoming). Employed in a snap? the impact of work requirements on program participation and labor supply. Technical report.

Gresh, A. (2020). Islam: A one-sided debate.

Griffen, A. S. (2019). Evaluating the effects of childcare policies on children's cognitive development and maternal labor supply. *Journal of Human Resources*, 54(3):604–655.

Gulesci, S., Jindani, S., La Ferrara, E., Smerdon, D., Sulaiman, M., and Young, H. (2021). A stepping stone approach to understanding harmful norms. CEPR Discussion Paper No. DP15776.

Guner, N., Kaygusuz, R., and Ventura, G. (2020). Child-related transfers, household labour supply, and welfare. *The Review of Economic Studies*, 87(5):2290–2321.

Haeck, C., Lebihan, L., Lefebvre, P., and Merrigan, P. (2022). Childcare, parental behavior and well-being during childhood. *The BE Journal of Economic Analysis & Policy*, 22(2):311–360.

Haeck, C., Lebihan, L., and Merrigan, P. (2018). Universal child care and long-term effects on child well-being: Evidence from canada. *Journal of Human Capital*, 12(1):38–98.

Haeck, C., Lefebvre, P., and Merrigan, P. (2015). Canadian evidence on ten years of universal preschool policies: The good and the bad. *Labour Economics*, 36:137–157.

Haeck, C., Lefebvre, P., and Merrigan, P. (2016). La politique des services de garde à contribution réduite (sgcr) : retombées probantes, faiblesses et pistes d'améliorations. Unpublished manuscript, Groupe de recherche sur le capital humain.

Harrison, K. A. (2016). Hiding under the veil of" dress policy": Muslim women, hijab, and employment discrimination in the united states. *Geo. J. Gender & L.*, 17:831.

Havnes, T. and Mogstad, M. (2011a). Money for nothing? universal child care and maternal employment. *Journal of Public Economics*, 95(11):1455–1465.

Havnes, T. and Mogstad, M. (2011b). No child left behind: Subsidized child care and children's long-run outcomes. *American Economic Journal: Economic Policy*, 3(2):97–129.

Havnes, T. and Mogstad, M. (2015). Is universal child care leveling the playing field? *Journal of Public Economics*, 127:100–114.

Heckman, J., Pinto, R., and Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6):2052–86.

Heckman, J. J. (2010). Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic literature*, 48(2):356–98.

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. (2010). The rate of return to the highscope perry preschool program. *Journal of Public Economics*, 94(1):114–128.

Hendren, N. (2016). The policy elasticity. *Tax Policy and the Economy*, 30(1):51–89.

Hendren, N. and Sprung-Keyser, B. (2020). A Unified Welfare Analysis of Government Policies*. *The Quarterly Journal of Economics*, 135(3):1209–1318.

Hendren, N. and Sprung-Keyser, B. (2022). The case for using the mvpf in empirical welfare analysis. Working Paper 30029, National Bureau of Economic Research.

Hendren, N., Sprung-Keyser, B., Stuart, E., Kelly, J., Aaronson, S., Odayappan, S., Creeron, A., Emery, J., Hutchinson, C., and Regier, J. (2023). Policy impacts library. `https://policyimpacts.org/policy-impacts-library`. Last accessed: 2023-01-16.

Herbst, C. M. (2017). Universal Child Care, Maternal Employment, and Children's Long-Run Outcomes: Evidence from the US Lanham Act of 1940. *Journal of Labor Economics*, 35(2):519–564.

Hojman, A. and Lopez Boo, F. (2022). Public childcare benefits children and mothers: Evidence from a nationwide experiment in a developing country. *Journal of Public Economics*, 212:104686.

Hotz, V. J. and Wiswall, M. (2019). Child care and child care policy: existing policies, their effects, and reforms. *The ANNALS of the American Academy of Political and Social Science*, 686(1):310–338.

Hoynes, H., Schanzenbach, D. W., and Almond, D. (2016). Long-run impacts of childhood access to the safety net. *American Economic Review*, 106(4):903–34.

Hoynes, H. W. and Schanzenbach, D. W. (2009). Consumption responses to in-kind transfers: Evidence from the introduction of the food stamp program. *American Economic Journal: Applied Economics*, 1(4):109–39.

Hoynes, H. W. and Schanzenbach, D. W. (2012). Work incentives and the food stamp program. *Journal of Public Economics*, 96(1):151–162.

Hu, Z. and Wang, Z. (2021). Nutrition, labor supply, and productivity: Evidence from ramadan in indonesia. Working paper.

Huntington-Klein, N., Arenas, A., Beam, E., Bertoni, M., Bloem, J. R., Burli, P., Chen, N., Grieco, P., Ekpe, G., Pugatch, T., et al. (2021). The influence of hidden researcher decisions in applied microeconomics. *Economic Inquiry*, 59(3):944–960.

Hyman, B. (2018). Can displaced labor be retrained? evidence from quasi-random assignment to trade adjustment assistance. Working Paper 3155386.

Iannaccone, L. R. (1992). Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy*, 100(2):271–291.

Institut de la Statistique du Québec (2022). Estimations de la population selon l'âge et le sexe, québec, 1er juillet 1971 à 2022. Last accessed 1 November 2023.

Institut Français d'Opinion Publique [IFOP] (2019). Etude auprès de la population musulmane en france, 30 ans après l'affaire des foulards de creil. Technical report.

Institut Montaigne (2014–2021). Religion au travail: croire au dialogue. baromètre du fait religieux en entreprise 2014–2021. Technical report.

Institut Montaigne (2016). A french islam is possible. Technical report.

Iyer, S. (2016). The new economics of religion. *Journal of Economic Literature*, 54(2):395–441.

Jácome, E. (2022). Mental health and criminal involvement: Evidence from losing medicaid eligibility. Working papaer.

Japel, C., Tremblay, R. E., and Côté, S. (2005). Quality counts! assessing the quality of daycare services based on the quebec longitudinal study of child development. *Choices: Investing in*

Our Children, 11(5).

Jia, R. and Persson, T. (2021). Choosing ethnicity: The interplay between individual and social motives. *Journal of the European Economic Association*, 19(2):1203–1248.

Jouili, J. S. (2020). *Pious Practice and Secular Constraints*. Stanford University Press.

Kaboski, J. P. and Townsend, R. M. (2011). A structural evaluation of a large-scale quasi-experimental microfinance initiative. *Econometrica*, 79(5):1357–1406.

Kang, Z. Y. and Vasserman, S. (2022). Robust bounds for welfare analysis. Working Paper 29656, National Bureau of Economic Research.

Karademir, S., Laliberté, J.-W., and Staubli, S. (2023). The multigenerational impact of children and childcare policies. IZA Discussion Paper 15894.

Khosrokhavar, F. (2004). L'islam des jeunes filles en france. In *Le foulard islamique en questions*, pages 89–94. Éditions Amsterdam.

Kleven, H., Landais, C., Posch, J., Steinhauer, A., and Zweimüller, J. (2021). Do family policies reduce gender inequality? evidence from 60 years of policy experimentation. Working Paper 28082, National Bureau of Economic Research.

Kleven, H. J. (2021). Sufficient statistics revisited. *Annual Review of Economics*, 13(1):515–538.

Kline, P. and Walters, C. R. (2016). Evaluating Public Programs with Close Substitutes: The Case of Head Start*. *The Quarterly Journal of Economics*, 131(4):1795–1848.

Kottelenberg, M. J. and Lehrer, S. F. (2013). New evidence on the impacts of access to and attending universal child-care in canada. *Canadian Public Policy*, 39(2):263–286.

Kottelenberg, M. J. and Lehrer, S. F. (2014). Do the perils of universal childcare depend on the child's age? *CESifo Economic Studies*, 60(2):338–365.

Kottelenberg, M. J. and Lehrer, S. F. (2017). Targeted or universal coverage? assessing heterogeneity in the effects of universal child care. *Journal of Labor Economics*, 35(3):609–653.

Kottelenberg, M. J. and Lehrer, S. F. (2018). Does quebec's subsidized child care policy give boys and girls an equal start? *Canadian Journal of Economics/Revue canadienne d'économique*, 51(2):627–659.

Krinsky, I. and Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The review of economics and statistics*, pages 715–719.

Kuka, E. and Shenhav, N. (2020). Long-run effects of incentivizing work after childbirth. Working Paper 27444, National Bureau of Economic Research.

La Boussole (2004). *Annuaire des mosquées de France*.

Lefebvre, P. and Merrigan, P. (2008). Child-care policy and the labor supply of mothers with young children: A natural experiment from canada. *Journal of Labor Economics*, 26(3):519–548.

Lefebvre, P., Merrigan, P., and Verstraete, M. (2009). Dynamic labour supply effects of childcare subsidies: Evidence from a canadian natural experiment on low-fee universal child care. *Labour Economics*, 16(5):490–502.

List, J. A. (2022). *The voltage effect: How to make good ideas great and great ideas scale.* Currency.

List, J. A., Suskind, D., and Supplee, L. H. (2021). *The Scale-Up Effect in Early Childhood and Public Policy: Why Interventions Lose Impact at Scale and What We Can Do About It.* Routledge.

Livny, A. (2020). *Trust and the Islamic advantage: religious-based movements in Turkey and the Muslim world.* Cambridge University Press.

Maillard, D. (2017). *Quand la religion s'invite dans l'entreprise.* Fayard.

Mattei, P. and Aguilar, A. (2016). *Secular institutions, Islam and education policy: France and the US in comparative perspective.* Springer.

Maurin, É. and Navarrete-H., N. (2022). Behind the veil: the effect of banning the islamic veil in schools. Working paper, Paris School of Economics.

Maurin, É. and Navarrete-H., N. (2023). Behind the veil: the effect of banning the islamic veil in schools. *Economic Policy*, 38(113):63–98.
McFadden, D. L. (1974). Conditional logit analysis of qualitative choice behavior. In P, Z., editor, *Frontiers in Econometrics*, pages 105–142. Academic Press.
Miguel, E. (2021). Evidence on research transparency in economics. *Journal of Economic Perspectives*, 35(3):193–214.
Miller, S. and Wherry, L. R. (2019). The long-term effects of early life medicaid coverage. *Journal of Human Resources*, 54(3):785–824.
Milligan, K. (2019). Canadian tax and credit simulator. Database, software and documentation, Version 2019-1.
Ministère de la Famille (2013–2019). Rapport annuel. Technical report, Gouvernement du Québec.
Ministère de la Famille, des Aînés et de la Condition Féminine (2005–2006). Rapport annuel de gestion. Technical report, Gouvernement du Québec.
Ministère de la Famille et de l'Enfance (1997–2003). Rapport annuel de gestion. Technical report, Gouvernement du Québec.
Ministère de la Famille et de l'Enfance (1998–2001a). Situation des centres de la petite enfance et des garderies au québec. Technical report, Gouvernement du Québec.
Ministère de la Famille et de l'Enfance (2001b). Situation des centres de la petite enfance et des garderies au québec en 2001. Technical report, Gouvernement du Québec.
Ministère de la Famille et des Aînés (2007–2012). Rapport annuel de gestion. Technical report, Gouvernement du Québec.
Ministère de l'Emploi, de la Solidarité sociale et de la Famille (2004). Rapport annuel de gestion 2003-2004. Technical report, Gouvernement du Québec.
Molnár, T. L. (2023). Costs of daycare, complementarities, and heterogeneous productivity of parenting time in child skill formation. Working paper.
Morrissey, T. W. (2017). Child care and parent labor force participation: a review of the research literature. *Review of Economics of the Household*, 15(1):1–24.
Morsy, M. (1989). *Les femmes du Prophète*. FeniXX.
Nollenberger, N. and Rodriguez-Planas, N. r. (2015). Full-time universal childcare in a context of low maternal employment: Quasi-experimental evidence from spain. *Labour Economics*, 36(C):124–136.
Nordmann, C. (2004). *Le foulard islamique en questions*. Éditions Amsterdam.
Novak, L. (2020). Persistent norms and tipping points: The case of female genital cutting. *Journal of Economic Behavior & Organization*, 177:433–474.
Office des services de garde à l'enfance (1994–1997a). Estimation régionale des besoins en services de garde régis par la loi. Technical report, Gouvernement du Québec.
Office des services de garde à l'enfance (1994–1997b). Situation des centres de la petite enfance et des garderies au québec. Technical report, Gouvernement du Québec.
Office des services de garde à l'enfance (1997a). Estimation régionale des besoins en services de garde régis par la loi. Technical report, Gouvernement du Québec.
Office des services de garde à l'enfance et Secrétariat à la famille (1997). Rapport annuel 1996-1997. Technical report, Gouvernement du Québec.
Paradisi, M. (2021). Firms and policy incidence. Mimeo.
Patel, D. S. (2012). Concealing to reveal: The informational role of islamic dress. *Rationality and Society*, 24(3):295–323.
Patnaik, A. (2019). Reserving time for daddy: The consequences of fathers' quotas. *Journal of Labor Economics*, 37(4):1009–1059.
Platteau, J.-P. (2017). *Islam instrumentalized*. Cambridge University Press.
Read, J. G. and Bartkowski, J. P. (2000). To veil or not to veil? a case study of identity negotiation among muslim women in austin, texas. *Gender & society*, 14(3):395–417.

Sakalli, S. O. (2019). Secularization and religious backlash: Evidence from turkey. Working paper.

Scott, J. W. (2009). *The politics of the veil*, volume 7. Princeton University Press.

Sénat (2021). Compte rendu intégral: Séance du mardi 30 mars 2021. *Journal Officiel de la République française.*

Seward, B., Dhuey, E., and Pan, A. (2023). The big short: Expansion of early childhood education in post-pandemic canada. *Canadian Public Policy*, 49(3):306–329.

Shofia, N. M. (2020). Why veil? religious headscarves and the public role of women. Working paper.

Silhouette-Dercourt, V., Sy, O. S., and Desjeux, D. (2019). Cosmopolitan veiling in paris: Young french muslim women in transition. *Youth and Globalization*, 1(1):65–87.

Skrondal, A. and Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 172(3):659–687.

Squicciarini, M. P. (2020). Devotion and development: Religiosity, education, and economic progress in nineteenth-century france. *American Economic Review*, 110(11):3454–3491.

Statistics Canada (2023). Table 39-10-0041-01 census families with children by age of children and children by age groups. Last accessed 6 September 2023.

The White House (2021). Fact sheet: Biden-harris administration announces american rescue plan funding to rescue the child care industry so the economy can recover. `https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/15/`. Last accessed: 2023-04-15.

Todd, P. E. and Wolpin, K. I. (2006). Assessing the impact of a school subsidy program in mexico: Using a social experiment to validate a dynamic behavioral model of child schooling and fertility. *American economic review*, 96(5):1384–1417.

Todd, P. E. and Wolpin, K. I. (2023). The best of both worlds: Combining randomized controlled trials with structural modeling. *Journal of Economic Literature*, 61(1):41–85.

Valfort, M.-A. (2020). Anti-muslim discrimination in france: Evidence from a field experiment. *World Development*, 135:105022.

van der Hasselt, G. (2019). The muslim veil in france: Why so controversial? three questions to hakim el kaouri.

Weichselbaumer, D. (2020). Multiple discrimination against female immigrants wearing headscarves. *ILR Review*, 73(3):600–627.

Wettstein, G. (2020). Retirement lock and prescription drug insurance: Evidence from medicare part d. *American Economic Journal: Economic Policy*, 12(1):389–417.

Yamaguchi, S., Asai, Y., and Kambayashi, R. (2018a). Effects of subsidized childcare on mothers' labor supply under a rationing mechanism. *Labour Economics*, 55:1–17.

Yamaguchi, S., Asai, Y., and Kambayashi, R. (2018b). How does early childcare enrollment affect children, parents, and their interactions? *Labour Economics*, 55:56–71.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.