Essays in Applied Microeconomics

by

Tímea Laura Molnár

B.A. and M.A.,Corvinus University of Budapest, 2010
M.A., Central European University, 2011

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate and Postdoctoral Studies

(Economics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2017

Abstract

The first chapter studies parents' intra-household time and resource allocation, focusing on parental quality time, and the implications for early childhood development. I develop a model that explains the "parental time-education gradient" puzzle, and I confirm its predictions, exploiting exogenous, drastic daycare price decrease in Quebec. I find that as daycare becomes cheaper, parents increase time devoted to their children, at the expense of home production and leisure, while consuming more of home production market goods (eating out, domestic help), and child market goods (daycare, games and toys); the time reallocation is larger for higher-educated parents. The estimated structural parameters uncover the pivotal role of complementarity (substitutability) between time and market goods in child human capital (home) production, and suggest a time efficiency advantage in non-market activities for higher-educated parents. I use them to assess how universal daycare shapes skill gaps in early childhood. My findings point to differential parental investment and time efficiency as important mechanisms behind widening skill gaps in early childhood.

The second chapter measures the causal impact of academic redshirting—the practice of postponing school entry of an age-eligible child—on student achievement and mental health. I use Hungarian administrative testscore data for 2008-2014, and an instrumental variable framework. The institutional feature I exploit is a school-readiness evaluation, compulsory for potentially redshirted children born before January 1st. I compare children born around this cutoff and find that (1) although there are large student achievement gains for all, disadvantaged boys benefit the most from redshirting—and only they benefit in terms of mental health; (2) the positive effects of higher school-starting age are driven by absolute, rather than within-class relative age advantage.

The third chapter studies how closely private insurers' payment schedules follow Medicare's, exploiting institutional changes in Medicare's payments and dramatic bunching in markups over Medicare rates. We find that, although Medicare's rates are influential, 25 percent of physician services, representing 45 percent of spending, deviate from this benchmark. Heterogeneity in the pervasiveness and direction of deviations reveals that the private market coordinates around Medicare's pricing for simplicity but innovates when sufficient value is at stake.

Lay Summary

My dissertation consists of three chapters in family economics and child development, economics of education and health economics. First, I investigate how parents choose the amount of quality time spent with their children, and why higher-educated parents spend more. I propose and confirm the mechanism of higher-educated parents being more efficient in translating time with their children into child human capital. Second, I measure the causal impact of academic redshirting on student achievement and mental health. Redshirting is the practice of postponing school entry of an age-eligible, but potentially not school-ready child. I show that disadvantaged boys are the main beneficiaries of this practice. Third, we study to what extent US private insurers benchmark to the corresponding Medicare price, when negotiating on how much to pay to doctors for a given medical service. In our data, 75% of physician services, representing 55% of spending, are directly linked to Medicare.

Preface

Chapters 2 and 3 are original, unpublished and independent work by the author, Tímea Laura Molnár. Chapter 4, forthcoming in the Journal of Health Economics, is a joint work with Professor Jeffrey Clemens (UCSD and NBER) and Professor Joshua Gottlieb (UBC and NBER). Chapter 3 was approved by the UBC Human Ethics Research Board under the name "Insurer-Physician Pricing Relationships" and project number H1302021. I have been involved throughout each stage of the research: preparing data, designing the empirical method, carrying out estimation, developing the economic model, and organizing and presenting results.

Table of Contents

Appendices

List of Tables

List of Figures

Acknowledgements

I would like to express my deepest gratitude to Kevin Milligan, Nicole Fortin, and Joshua Gottlieb for their invaluable support, excellent guidance, and infinite patience throughout my years at UBC. They were extremely generous with their time and encouragement, and taught me how to be a better economist and researcher. I also want to thank David Green and Vadim Marmer, who, even though not formally on my committee, still devoted a lot of time towards helping me with modeling and econometric theory. Gábor Kézdi (CEU) taught me the foundations of empirical research, provided essential conceptual suggestions to improve this thesis, and has had a life-long impact on me and my career. I am also indebted to Matilde Bombardini, Patrick Francois, Florian Hoffmann, Hiro Kasahara, Thomas Lemieux, Jamie McCasland and Marit Rehavi (UBC), Irene Botosaru (SFU), Hedvig Horváth (UCL), and Péter Hudomiet (RAND) for very useful suggestions and discussions at different stages of my doctoral studies. Finally, I have had many extremely talented and encouraging fellow students at UBC, such as Oscar Becerra, Anujit Chakraborty, Thomas Cornwall, Alix Duhaime-Ross, Alastair Fraser, Coral Gonzales, Brad Hackinen, Nouri Najjar, Adlai Newson, Jutong Pan, Pierluca Pannella, Rogerio Santarrosa, Daniel Shack, Iain Snoddy, Lori Timmins, Ruoying Wang, and Tzu-Ting Yang. All aforementioned helped me more than they can imagine.

Most of the analysis in the first chapter was completed at the UBC Research Data Centre, where Cheryl Fu, Lee Grenon and Wendy Kei provided invaluable help with the data.

Regarding data access to the second chapter, I would like to thank members of the Economics of Education Research Unit at the Hungarian Academy of Sciences, Institute of Economics, especially Gábor Kertesi, Gábor Kézdi, Zoltán Hermann and Melinda Tir, for providing me the datasets of the Hungarian National Assessment of Basic Competences, the Hungarian Life Course Survey, KIR-Stat and the Local Government Treasury Dataset, and for helping with the data. My thanks go to Gábor Kertesi for providing me the municipality-level demography dataset for Hungary.

Regarding the third chapter, I and co-authors Jeffrey Clemens and Joshua Gottlieb, are grateful to Luisa Franzini, Cecilia Ganduglia-Cazaban, Osama Mikhail, and the UTSPH/BCBSTX Payment Systems and Policies Research Program at the University of Texas School of Public Health for data access, and for their extensive assistance in navigating the BCBSTX claims data. We thank SSHRC, Jon Skinner and the Dartmouth Institute for support, and Victor Saldarriaga and Andrew Vogt for excellent research assistance. We are grateful to Leila Agha, Liran Einav, Je Emerson, Amy Finkelstein, John Friedman, Matt Gentzkow, R.B. Harris, David Laibson, Neale Mahoney, Matt Panhans, Maria Polyakova, Jon Skinner, and Ashley Swanson for helpful comments.

Finally, I thank my parents, my sister, and best friends Zsuzsi, Adlai and Miklós for their unconditional love, guidance, and encouragement throughout my life.

Chapter 1

Introduction

This thesis consists of three chapters in applied microeconomics. These chapters touch upon areas of family economics and early childhood development, economics of education and health economics. They all exploit some unique institutional feature or policy change of major economic sectors—the labor market, the educational and health systems—shaping incentives, to identify the causal impact on families' or firms' behavior, primarily focusing on disadvantaged members of the population.

The *first chapter* asks: what drives parents' child time allocation choices, and why do these differ across education groups? I focus on parental quality time, and the implications for early childhood development and skill gaps. From a Beckerian point of view it is puzzling why higher-educated parents spend more time with their children, despite their less time spent on home production and leisure, and their higher opportunity cost of time in non-market activities. This cross-sectional observation, called the "parental time-education gradient" puzzle—documented by Guryan, Hurst and Kearney (2008)—is the starting point of this chapter.[1] To resolve it, I examine parents' time and market good responses to a shock to the opportunity cost of their time, by exploiting a drastic and exogenous universal daycare price decrease in Quebec in 1997. I ask whether different substitutability possibilities between time and market goods in child human capital and home production can explain parents' time choices both cross-sectionally and in response to the shock. I also ask whether heterogeneous time efficiency in non-market activities by education can explain differential responses across education groups. Looking at parents' responses to a price change helps to rule out competing explanations for the cross-sectional observation, since alternative models predict responses the data rejects.

Examining responses to a daycare price change also has immediate policy implications: governments in many countries subsidize the labor market reintegration of mothers with young children, in part responding to concerns about gender inequality within the household and on the labor market. However, a question arises whether these policies—such as providing direct daycare subsidies, tax credits or tax deductions to eligible daycare expenses—crowd out or reinforce parental investments in child human capital formation, other than daycare, such as parental quality time. Differential responses by education have important implications for whether policies should target low-education families to "level the playing field"[2], or rather should be universal. To the extent that parental quality time

[1] Early references to this puzzle are Hill and Stafford (1974) or Leibowitz (1974).

[2] Evidence on the impact of universal or large-scale daycare programs is inconclusive on the mean impact on child outcomes. The results of Havnes and Mogstad (2015) suggest that the effects of subsidized daycare vary systematically across the outcome distribution, and that disadvantaged children of are the primary beneficiaries. An example of targeted daycare policy is the one implemented exactly in Quebec: the contribution for a subsidized daycare place recently changed from being universal to a system of a basic contribution plus an additional contribution, the latter adjusted by family income. As of January 1, 2016, the basic contribution is $7.55 per day, per child, and the additional contribution is 0 below a family income of $50,000, then it increases gradually up to $13.15 per day, corresponding to $158,820. The additional contribution is reduced by 50% for second, and by 100% for third and additional children. Additional details can be found at http://www.budget.finances.gouv.qc.ca/Budget/outils/garde_en.asp.

and skill-enhancing resources contribute to child skill formation, differential time allocation choices shape skill gaps in early childhood between children of higher- and lower-educated parents. Parental time investment behavior and the extent policies might influence them have important implications for equality of opportunity, given the high correlation between early childhood human capital and adult outcomes on the one hand, and the increasing consequences of accident of birth on the other.[3]

I have three contributions. First, the theoretical contribution is extending the classic Beckerian framework of child skill formation with time efficiency differences across education-groups in non-market activities, and allowing for differential substitutability between time and market goods in child human capital and home production. Second, motivated by the predictions of the model, I offer a new set of reduced-form results on parental time and household expenditures using exogenous daycare price variation. Using the same moments in the data, I estimate the model's structural parameters on the substitutability between time and market goods in child human capital and home production, and on the between-education-group heterogeneity in non-market time efficiency. Third, I test the behavioral skill gap between children of higher- and lower-educated parents in the first five years of life; using the structural estimates and parents' observed choices, I assess how universal daycare shapes the gap.

I find that as daycare becomes cheaper, mothers are more likely to work and they supply more hours on the market. At the same time, both mothers and fathers increase time devoted to their children, at the expense of home production time, while consuming more of home production market goods (eating out in a restaurant and hiring domestic help), and child market goods (daycare, games, books and toys). I find the time reallocation to be larger for higher-educated parents, having completed some post-secondary education. I use the model to estimate structural parameters, which uncover the pivotal role of complementarity between time and market goods in child skill formation, and substitutability in home production. In addition, they suggest a time efficiency advantage in non-market activities for higher-educated parents, outweighing their higher opportunity cost of time.

My findings on differential parental time responses link to the literature on human capital gaps in early childhood. There is evidence on significant cognitive, noncognitive (also called behavioral or developmental)[4] and health gaps between high- and low-educated parents' children, opening up early and widening over ages.[5] Although the literature on these gaps often uses the label 'early childhood',

[3]Chetty et al. (2011) document that kindergarten test scores are associated with better adult outcomes on a wide range, such as earnings at age 27, college attendance, home ownership, and retirement savings. Chetty et al. (2014) show evidence that although rank-based mobility measures remained stable for children born between 1971 and 1993, since income inequality increased, the "birth lottery" has larger consequences today.

[4]Noncognitive or behavioral skills, such as temperament, persistence, self-discipline, adaptability, reliability, etc., have only recently attracted economists' interest, in explaining general educational/ labor market outcomes (e.g. Rubinstein and Heckman (2001), Osborne et al. (2001), Heckman et al. (2006), Borghans et al. (2008), Deming (2015)), labor market returns to particular personality traits (e.g. Osborne (2005), Mueller and Plug (2006), Heineck and Anger (2010)) or the probability of dropping out of school (Coneus et al., 2011). This research indicates that noncognitive skills play a significant, increasing role in the labor outcome process; e.g. Deming (2015) shows the link between jobs' social skill requirements and wage growth since 1980. Additionally, there is a growing literature on the technology of human capital formation, including noncognitive skills (e.g. Cunha et al. (2006), Heckman and Cunha (2007)). The policy implication argued by this research area is to invest in young disadvantaged children, as there is no efficiency-equity trade-off of such investments and consequences of the accident of birth could be alleviated (Heckman and Masterov (2007)).

[5]Fryer and Levitt document that the Black-White achievement gap increases on average by 10 percent of a standard deviation per school year in the first three grades (Fryer and Levitt (2004), Fryer and Levitt (2006)), and rule out genetic differences across races to account for the gap (Fryer and Levitt, 2013). Cunha et al. (2006) and Carneiro et al. (2005) show an 8 and a 6 percentage points increase in the Peabody Individual Achievement Test (PIAT) mathematics score

the majority of this evidence relies on children aged five or older. I test the noncognitive gap in the first 3-4 years of life, often labeled as the 'critical period' in child development. I find the largest widening between children aged 0-2 and 3-4 years old. I find that bedtime reading, maternal (mental) health and positive parenting practices are more important transmission mechanisms from socio-economic background to behavioral scores than are daycare time or maternal work. These results motivate the chosen age range and justify the focus on parental time in this chapter.

My findings point to parental investment as one important mechanism behind widening skill gaps in early childhood. In this area of research there is more evidence on health gaps[6], than there is on cognitive/non-cognitive skill gaps. According to Heckman and Cunha (2007), binding family income constraints in early childhood lead to underinvestment in skills relative to the case of perfect credit markets, and variation in parental environment and initial endowment are the main reasons behind widening skill gaps. Buttressing the importance of parental environment, I find that higher-educated mothers most likely to be drawn (back) into the labor market increase time spent with their children the most, and decrease home production and leisure time the most. Also, they invest more in reading, parenting and—if depression score is an adequate measure of health investments,—also in their own mental health. Consistent with these larger investments, the behavioral outcomes of these mothers' children deteriorated less, or even improved after the introduction of the Quebec Daycare Policy.

There is a literature in early childhood development on whether parents invest in a reinforcing or a compensatory manner; however, with respect to the child's initial endowment. Rosenzweig and Wolpin (1988) find that children with better health endowments are more likely to be breastfed, Datar et al. (2010) find that breastfeeding, nursery school enrollment and maternal time increase with birth weight and Aizer and Cunha (2012) find that the degree of reinforcement increases with family size. Less educated mothers are found to invest in a reinforcing, while more educated ones in a compensatory manner (Hsin, 2012), although the literature is inconclusive about the mechanisms at play. My results in the first chapter indicate that high-educated parents invest in a compensatory manner with respect to the mother's labor supply, with their time efficiency advantage being the underlying mechanism.

The *second chapter* focuses on the impact of academic redshirting on student achievement and mental health, and its motivation stems from the literature on school entry delay and school starting age. There are two main practices to delay school entry by a year: one is compliance with the school enrollment cutoff date, the other is the practice of postponing an age-eligible, but potentially non-school-ready child's school entry – called academic redshirting. In this chapter I measure the causal impact of starting school a year older, using Hungarian data for years 2008-2014, and exploiting two discontinuity points in month of birth in an instrumental variable framework.

gap between ages 5 and 13, between children from families in the lowest and highest income quartile and between Blacks and Whites, respectively. Case et al. (2002) and Currie and Stabile (2003) provide evidence on the steepening health - status gradient by age for the United States and Canada, respectively. Cunha et al. (2006) document a 4 percentage point increase in the anti-social behavioral score gap between both poor and rich children and between Black and White children aged 4-12. These gaps disappear when controlling for family structure, maternal education or maternal ability, measured by the AFQT-score; although there are no formal tests shown on the significance or the shape of the gap.

[6]Case et al. (2002) document that poor children with a chronic condition have worse health than rich children with a chronic condition. Additionally, by ruling out mechanisms operating through health insurance, health at birth and genetics, they find that income buffers children from adverse effects of chronic conditions and this buffering effect is cumulative. According to Currie and Stabile (2003), the worsening health-status gradient is more likely due to the higher arrival rate of shocks, rather than the lower recovery rate for low-status children.

To the best of my knowledge, mine is the first attempt to measure the causal impact of redshirting on child outcomes, using a natural experimental design. The main institutional feature I exploit is a school-readiness evaluation in the Hungarian educational system, compulsory for potentially redshirted children born before January 1st. By comparing children born around January 1st, I measure the combined impact of age and boosted human capital due to redshirting, for complier children who might or might not have struggled with school-readiness problems. By comparing children born around the June 1st school enrollment cutoff date, I measure the sole age impact of starting school a year older for the complier children who, by definition, did not struggle with any school-readiness problems.

I provide three further contributions. First, I show the impact on broader cognitive and different noncognitive child outcomes, than shown previously. Using Hungarian administrative testscore data for 2008-2014, I show the impact on mathematics and reading testscores at grades 6, 8 and 10, grade repetition by grades 6, 8 and 10, and secondary school track choice. Using Hungarian survey data for 2008, I show the impact on mental stability measures at grade 8, measured by anxiety and exhaustion. I present the impact by the interaction between gender and parental education; an angle that has been ignored by the majority of the reviewed literature. Second, I contribute to the literature on academic redshirting also by relating the propensity of being redshirted to birth shocks, family shocks, and health/developmental obstacles in early childhood. Third, exploiting natural variation in the fraction of summer-born children in class and month of birth, I disentangle absolute and relative age effects of starting school a year older due to compliance with the school enrollment cutoff date.

Studies in the existing literature identify the impact of school starting age primarily using school enrollment cutoff dates. For instance, Elder and Lubotsky (2009) exploit variation in kindergarten entrance age stemming from two sources: the distribution of birth dates throughout the calendar year among those who comply with their state's school enrollment cutoff date, and differences across states in cutoff dates among children born on the same day but living in different states. Puhani and Weber (2007) use a similar identification strategy for Germany. I measure both the causal impact of starting school later using the school enrollment cutoff date, and the causal impact of academic redshirting.

The relevant literature focuses primarily on the effect of school starting age on student achievement test results in kindergarten and at grades 4, 6 and 8. Additionally, there is some evidence on grade repetition of first grade, secondary school track choice after grade 6, IQ measured at the age of 18 and earnings at the beginning of the labor market history. Noncognitive and prosocial outcomes are also considered, as persistence, stability and hyperactivity at ages 8, and 11 and teenage pregnancy.

The literature on testscores shows consistently across countries that children who entered school at a higher age–due to the school enrollment cutoff date–perform better on achievement tests. Puhani and Weber (2007) provide evidence for Germany, Kollo and Hamori (2011) for Hungary, Fredriksson and Ockert (2006) for Sweden, McEwan and Shapiro (2008) for Chile, Elder and Lubotsky (2009) for the United States and Crawford et al. (2007) for England. The range of the impact on standardized testscores is found to be between 0.2 and 0.5 standard deviations, with generally decreasing impacts across grades 4 to 8 and larger effects for boys and disadvantaged children.[7] The similarly positive

[7] Puhani and Weber (2007) present evidence that entering the German primary school system at the age of 7 instead of 6 increases 6th-grade reading testscores by 40 percent of a standard deviation. Kollo and Hamori (2011) show that the effects are larger for disadvantaged children. They find very large effects for student achievement in grade 4 (80 percent

effects on grade repetition and secondary school track choice are documented by McEwan and Shapiro (2008) and Puhani and Weber (2007); entering school a year older is found to decrease the chance of repeating first grade by 2 percentage points and to increase the probability of attending the most advanced school track by 12 percentage points.[8] In contrast, Black et al. (2011) find a small positive effect of starting school younger on IQ scores at the age of 18. Regarding the noncognitive outcomes, Evans et al. (2010) show that older children have a significantly lower incidence of Attention Deficit Hyperactivity Disorder (ADHD) diagnosis and treatment, Muhlenweg et al. (2012) present German evidence that school starting age has a stable positive effect on persistence, a short-run negative impact on hyperactivity, and a long-run effect on being more adaptable to change, while Fortin et al. (2015) find that being young in class aggravates an underlying propensity toward inattentive or hyperactive behaviour, that is more prevalent for boys. Black et al. (2011) present Norwegian evidence that children starting school at a younger age have a larger probability of teenage pregnancy and a short-run advantage in earnings (although disappearing by age the age of 30). Using extensive Danish data on mental health outcomes, Dee and Sievertsen (2015) find that starting school a year older substantially reduces inattention/hyperactivity at the age of 7, and the effect persists to the age of 11. I show the impact on broader cognitive (up to grade 10) and different noncognitive child outcomes (as anxiety and exhaustion), and by the interaction between gender and parental education.

Similarly to school starting age, relative age effects (separately) are found to be positively related to various dimensions of educational success.[9] However, this literature on relative age effects does not answer the question whether relative age effects remain important beyond school starting age, that can be of interest of both parents and policy makers. In the last part of the chapter, exploiting natural variation in the fraction of summer-born children in class and month of birth, I find the positive effects of higher school entry age to be driven by absolute, rather than within-class relative age effects. The results suggest that entering primary school a year later matters because it makes the child older in absolute terms, rather than making the child older relative to her classmates at the time of the test.

The *third chapter*, joint work with Jeffrey Clemens and Joshua Gottlieb, studies how closely private insurers' payment schedules follow that of Medicare's. One of private health insurers' unique roles in the United States is to negotiate physician payment rates on their customers' behalf. Do private insurers' payment schedules differ from that of Medicare, their public sector counterpart, or is the ostensibly prominent private sector a mirage? We investigate the frequency with which privately negotiated payments deviate from the public sector benchmark using two empirical approaches. The

of a standard deviation), while 25-40 percent for student achievement in grade 8 for children whose mother finished at most primary school. The same effects for children whose mother received a tertiary degree are 0.3 and 0.2 standard deviations, respectively. They find substantially higher effects for reading, than for mathematics testscore. McEwan and Shapiro (2008) find that starting school one year later leads to a more than 30 percent increase in 4th-grade and 8th-grade standardized testscores. They show that entering primary school one year older increases 4th-grade testscores by 0.29 standard deviations in mathematics and 0.38 standard deviations in language. Elder and Lubotsky (2009) find evidence that entering kindergarten one year older leads to a 53 (83) percentage point increase in reading (mathematics) testscores during the fall of kindergarten. However, these effects fade away, faster for disadvantaged children.

[8] In a similar study, Muhlenweg and Puhani (2010) find that early school entrants are only 2/3 as likely to enter Gymnasium (the academic school track in Germany) as older entrants.

[9] Dhuey and Lipscomb (2010) present evidence from the United States that an additional month of relative age decreases the probability of receiving special education by 2-5 percent; Dhuey and Lipscomb (2008) find that the oldest students are 4-11 percent more likely to be high school leaders (and presumably enjoy the wage premium attached to high school leadership); Bedard and Dhuey (2006) show the effect of relative age on testscore across the OECD-countries.

first exploits changes in Medicare's payment rates and the second exploits dramatic bunching in markups over Medicare rates. Although Medicare's rates are influential, we find that prices for 25 percent of physician services, representing 45 percent of spending, deviate from this benchmark.

To understand private insurers' objectives, we examine heterogeneity in the pervasiveness and direction of deviations they make from the Medicare benchmark. We show that the Medicare-benchmarked share is high for services provided by small physician groups. It is low for capital-intensive care, for which Medicare's average-cost reimbursements deviate most from marginal cost. When relative prices deviate from Medicare's, they adjust towards the marginal costs of treatment.

One plausible interpretation of these findings emphasizes the complexity of the insurer-physician contracting environment. To manage the tension between gains from fine-tuning payments and costs from making contracts complex, insurers may draw on Medicare's relative value scale for the purpose of contract simplification, while strategically adapting their contracts where the value is highest. This view is consistent with the heterogeneity we observe: the benefits of fine-tuning payments will tend to be largest within contracts with large physician groups and for the capital-intensive services for which Medicare's average cost payments deviate most from marginal cost. Complementary interpretations may highlight the relevance of large firms' bargaining power. The information content of the relative value scale on which Medicare's payments rely can also be interpreted as a knowledge standard or, more generally, as a public good.

Our results are of potential interest to analyses of two broader contexts. Learning how prices are set in health care—a sector comprising 18 percent of the economy—is essential for understanding macroeconomic price-setting dynamics. The service sector in general (Nakamura and Steinsson, 2008), and medical care in particular (Bils and Klenow, 2004), have especially sticky prices. We provide evidence on how this stickiness arises.[10] Consistent with the evidence provided by Anderson et al., (forthcoming) from retail, the complexity of physician contracting may explain both the long duration of these prices and the public-private linkages we identify.

Public polices' residual influence on private firms is relevant in a wide range of contexts. Outside of the health care context, labor contracts sometimes benchmark wage rates to the statutory minimum.[11] Within the health sector, Medicare has been found to shape aspects of private players' behavior in the pharmaceutical, hospital, and physician marketplaces (Duggan and Scott Morton, 2006; Alpert et al., 2013; White, 2013; Clemens and Gottlieb, 2017). The forces we investigate here differ conceptually from those that our prior work uncovered, including the analysis of Clemens and Gottlieb (2017) on physician payments. Clemens and Gottlieb (2017) assess how incentives and outside options lead to Medicare's influence on physicians' bargaining positions. Our research emphasizes the Medicare payment model's role as an "industry standard." This role stems in part from the information contained in the Medicare payment model's estimates of services' relative input costs. Our analysis provides insights into the overall pervasiveness of benchmarking against Medicare's relative cost schedule and into the types of contracts in which customization is most prevalent.

[10] In particular, our empirical evidence supports price-setting mechanisms with the flavour of Christiano et al. (2005) or Smets and Wouters (2003, 2007).

[11] A publicly posted contract template of the United Food and Commercial Workers Union (2002), for example, includes the requirement that "At no time during the life of this Agreement will any of the bagger/carry-out rates be less than twenty-five ($0.25) cents an hour above the Federal minimum wage."

Chapter 2

How Do Mothers Manage? Universal Daycare, Child Skill Formation, and the Parental Time-Education Puzzle

2.1 Introduction

A mother working full-time today spends more time with her children than non-working mothers did in 1960, and today's non-working mother spends less time on housework than working mothers spent in 1960.[12] Highly-educated working mothers today devote on average 2 hours more per week to work and 1.7 hours more to their children than lower-educated working mothers do, at the expense of their home production and leisure time.[13] Parental time investments in early childhood are crucial; for instance, reading to the child an extra day per week in the first ten years of life was shown to increase reading testscores by more than 40 percent at the age of 11.[14] In this chapter I ask: what drives parents' child time allocation choices, and why do these allocations differ across education groups?

The classic Beckerian framework suggests that higher-educated parents, similarly to their less time spent on home production and leisure, should spend less time with their children, since their non-market activities have higher opportunity costs. However, previous research finds the opposite, leading to a "parental time-education gradient" puzzle, documented by Guryan, Hurst and Kearney (2008). In this chapter I develop a model of intra-household time and resource allocation that explains this puzzle. It also offers a mechanism based on heterogeneous time efficiency in non-market activities, that leads higher-educated parents to respond differently to price shocks.

In my model setup parents derive utility from three commodities: child human capital, home production goods and leisure goods. These commodities are produced from time and a market good,

[12] Based on time use diary data compiled on 16 industrialized countries by Gauthier et al. (2004); see Table 1 for married or co-habitating women between ages 20 and 49, with at least one child under the age of 5.

[13] Based on data from the Canadian General Social Survey - Time Use Diary (1998), for working women below 50 in two-parent households, having at least one child aged 0-4; low-educated denotes having at most a high school degree.

[14] Using data from the NLSY'79, Price (2012) accounts for parents' endogenous reading behavior by using birth order as an instrument, exploiting the empirical observation that first-born children are read to more often. In an earlier paper (Price, 2008), based on the American Time Use Survey, he provides evidence that first-born children receive more quality time from both the father and the mother—by approximately 20 and 25 minutes—per day at each age, than do second-born children at the same age. Using both the birth order instrument and a propensity score approach on data from the Longitudinal Study of Australian Children, Kalb and van Ours (2014) find that being read to at the age of 4 and 5 regularly has significant positive effects on the reading and cognitive skills of children up to an age of 10 or 11: reading 3-5 (6-7) days per week increases a cognitive skill index by half a (almost one) standard deviation. Using data from the Fragile Families and Child Wellbeing Study and a propensity score approach, Hale et al. (2011) find a positive relationship between language-based bedtime routines and nighttime sleep duration, general health and verbal test scores, and a negative relationship with behavior problems (anxious, withdrawn, and aggressive behaviors).

with constant elasticity of substitution (CES) production functions; and they are assumed to differ in how easily time can be substituted for the market good in their production processes. By assuming that parents use daycare while the mother is working, I link daycare price to the opportunity cost of the mother's time in non-market activities. This framework maps structural parameters into optimal allocations, gives an explanation to the parental time-education puzzle and generates a rich set of predictions on responses to a shock to the opportunity cost of time. To account for differential responses by education, I allow for heterogeneity in the efficiency of parents' time in child human capital and home production. The model predicts that less expensive daycare induces mothers to work more, to use more daycare and to purchase more market goods, with their time response in parenting and home production depending on the complementarity or substitutability between time and market goods in child human capital and home production. Higher-educated parents are predicted to change their time allocation more, if they have a time efficiency advantage in non-market activities that outweighs their larger opportunity cost of time in non-market activities.

I confirm the model's predictions, by exploiting exogenous daycare price drop in Quebec (1997). I use data from the National Longitudinal Survey of Children and Youth (NLSCY), the Labor Force Survey (LFS), the Survey of Household Spending (SHS) and the Canadian Time Use Diary and Census datasets, on parents with children aged 0-4. I present new evidence on the total effect of an uncompensated daycare price change on parents' time allocation at home, on their home production- and child-related market good expenditures, and on their reading practices. I also exploit the Thuringian daycare policy change (2006) and data from the German Socio-Economic Panel (GSOEP) to estimate the impact of a compensated daycare price change and a pure income shock on parents' time allocation, with the aim of disentangling income and substitution effects. I use the model to estimate structural parameters, with which I am able to assess how universal daycare shapes skill gaps in early childhood.

I confirm existing evidence on the Quebec daycare policy change on large maternal labor supply and daycare use responses, and find that these are driven primarily by higher-educated mothers unlikely to work in the absence of the policy. I also find that as daycare becomes cheaper, (1) parents increase time devoted to their children, at the expense of their home production and leisure time, while consuming more of home production market goods (eating out, domestic help), and child market goods (daycare, child games and toys); (2) the time reallocation is larger for higher-educated parents.

The first set of reduced-form findings (1) on time and market good expenditures generate particular implications for the signs of the substitution parameters, and uncover the pivotal role of substitutability and complementarity between time and market goods: parents substitute their time away from activities where time is substitutable with market goods (home production) to activities where time is complementary to market goods (child human capital production). These findings are consistent with available evidence: Aguiar and Hurst (2007), using scanner and time use data, find that time and money for home production and shopping are highly substitutable. Cunha et al. (2010) find that lagged human capital and investment inputs in child human capital production are complementary.

The second set of reduced-form findings (2) on differential time responses imply, through the lens of the model, that higher-educated parents' time is more efficient—has a higher marginal return—in child human capital and home production, compared to lower-educated parents' time. These findings are also consistent with empirical evidence. Kalil et al. (2012) show that higher-educated mothers choose activities that most fit their child's current developmental needs[15]; Hoff (2003) shows that hearing more advanced speech at home is one reason behind higher-educated parents' children having more advanced language skills[16]; Weisleder and Fernald (2013) find large variation in child-directed speech by parental education, and argue that children exposed to it become more efficient in processing familiar words in real time, and have larger expressive vocabularies by the age of two.

The rest of the chapter is organized as follows. Section 2.2 develops the model, presents its testable implications and discusses alternative models (alternative to differences in time efficiency in non-market activities by education). Section 2.3 provides the institutional details of the Quebec Daycare Policy (1997). Section 2.4 presents the estimation approach and section 2.5 describes the data. Section 2.6 presents and discusses the results on the total impact of a daycare price change, the structural parameter estimation and the policy simulation on early childhood skill gaps. Section 2.7 discusses alternative models–alternative to time efficiency differences across education groups–, along with their predictions, that are not supported by the data. Section 2.8 disentangles income and substitution effects directly, by estimating the impact of a compensated daycare price change using the Thuringian Daycare Policy (2006-2010). Section 2.9 concludes.

2.2 Model

The model is inspired by the setup of time allocation outlined in words by Guryan et al. (2008). The following elements are taken from their setup: households derive utility from a home-produced good, a leisure good, and well-cared-for children. As in Becker (1965), these goods are produced from time and a general market good. Following Aguiar and Hurst (2007), the goods can be classified based on the elasticity of substitution between time and the market good in their production processes. Guryan et al. (2008), without formalizing this setup, discuss how optimal cross-sectional choices might change with wages, the opportunity cost of time. They suggest that such a framework can be extended by heterogeneity of preferences or time productivity according to some measure of earnings potential. However, they did not investigate parents' time reallocation responses to a shock to the opportunity cost of their time, neither did they distinguish between competing explanations for the parental time-education puzzle. My theoretical contribution is to formalize this setup, and to derive the model's explanation both for the puzzle—the *cross-sectional* observation on *choices*—, and parental time reallocation *responses* to daycare price changes (linked to the opportunity cost of mothers' time).

Parents' time reallocation in response to daycare price changes is crucial not only for identification purposes, but it also allows us to separately test the explanations Guryan et al. (2008) provide for the cross-sectional observation. They propose the following explanations of why higher-educated parents spend more time with their child(ren): (1) children's human capital is a luxury good (with higher income elasticity than home production or leisure); (2) the substitution of time and market goods in

[15] For instance, they spend most of their times reading and problem-solving while their child is in preschool, but shift to management of their children's lives outside the home while their child is in middle school.

[16] Speech at home is measured by number/length of utterances, word tokens/types, and topic-continuing replies.

producing children's human capital is perceived to be lower for higher-educated parents; (3) higher-educated parents have a higher preference for their children's human capital; (4) returns to investment in children's human capital is higher for higher educated parents. Using my model I am able to generate testable predictions for each of these four competing mechanisms, and I show that only (4) provides predictions on the parental time and daycare use responses that are consistent with the data.

In this section first I present the model and the Marshallian demands. Then, I describe the estimation approach of the structural parameters, using the demands and exploiting exogenous daycare price variation from the Quebec policy change. Lastly, I discuss the model's reduced-form predictions.

2.2.1 The Setup of the Model

Consider a model of the household that abstracts from bargaining and fertility decisions. Suppose a household has one child and derives utility from three commodities: child human capital K, home production goods H and leisure goods L, according to the following Cobb-Douglas utility function:

$$U = \beta_K \log K + \beta_H \log H + \beta_L \log L. \tag{2.1}$$

For simplicity, I impose the following normalization: $\beta_K + \beta_H + \beta_L = 1$.

Commodities K, H and L are produced from time T and market good X with CES production functions, where ρ_j is the substitution parameter between T and X, for $j = K, H, L$:

$$K = \left[(\gamma_K T_K)^{\rho_K} + X_K^{\rho_K}\right]^{\frac{1}{\rho_K}};\ H = \left[(\gamma_H T_H)^{\rho_H} + X_H^{\rho_H}\right]^{\frac{1}{\rho_H}};\ L = \left[\gamma_L T_L^{\rho_L} + X_L^{\rho_L}\right]^{\frac{1}{\rho_L}}. \tag{2.2}$$

Negative values of ρ imply complementarity between inputs, values of ρ between 0 and 1 imply substitutability between inputs, with the Leontief and perfect substitution being the limiting cases. T_K is parents' time investments in child human capital, while X_K includes daycare time D and all other child goods B (including child books, games and toys), in a nested CES with substitution parameter ρ_X: $X_K = [B^{\rho_X} + D^{\rho_X}]^{\frac{1}{\rho_X}}$; D has a price of m, while B, X_H and X_L have unit price. T_H is time spent on home production, while X_H includes, e.g., domestic help, laundry/cleaning services, pre-prepared meals and food inputs. L can be thought of as leisure experience, produced from X_L (say, movie tickets or books) and leisure time. The total household time for category j is produced from time of the mother (M) and the father (F), in a nested CES with substitution parameter ρ_{Tj} :

$$T_j = \left[\left(T_j^M\right)^{\rho_{Tj}} + \left(T_j^F\right)^{\rho_{Tj}}\right]^{\frac{1}{\rho_{Tj}}} \text{ for } j = K, H, L.$$

Mothers and fathers work to be able to buy the market goods and earn wage w^M and w^F, respectively.

Higher- and lower-educated households differ by their time efficiency in child human capital, home production, and leisure production. The time efficiency parameters γ_K,γ_H and γ_L are related to the marginal product of time investment in non-market activities, and conceptually are closely related to the marginal productivity of work at home as in Gronau (1977). These parameters are allowed to depend on years of schooling S and a good-specific error ε_j: $\gamma_j = \Upsilon(S, \varepsilon_j)$, for $j = K, H, L$, where, after the parametrization of Υ, δ_j will represent the dependence (the coefficient) of γ_j on S. I allow for the efficiency in transforming time into child human capital, home production and leisure goods to differ by education. Specifically, I allow for the possibility that a higher-educated parent's unit of time, on average, produces a different amount of commodity output, than does a lower-educated parent's unit of time. Note that these parameters are linked to the marginal product of time input in the non-market production of K, H and L, but are not linked to market productivity, captured by the wage w. I discuss and show formally in Section 2.7 that alternative models only with differing preferences or substitution parameters by education, or allowing daycare quality to decrease contemporaneously with the price decrease, all generate predictions that are not supported by the data.

The structural parameters to be estimated are the substitution parameters between time and market goods in producing child human capital (ρ_K) and producing the home production good (ρ_H), and δ_K and δ_H, that show the dependency of time efficiency on schooling.

The constraints households face incorporate the assumptions that (i) the parents use daycare while the mother is working (thus, $D = T_W^M$), (ii) parents hire a nanny while enjoying leisure time together, and (iii) there is no overlap between any two time-use categories: e.g. during home production time the child might be around the parents, but is not the primary focus of their attention. Formally, the household maximizes utility in (2.1) with respect to B, X_H, X_L bought by the household and time use allocations T_W, T_K, T_H, T_L of both the mother (M) and the father (F), subject to the time constraints $T_W^i + T_K^i + T_H^i + T_L^i = \bar{T}$ for $i = M, F$ and the household budget constraint

$$B + X_H + X_L + mT_W^M + nT_L^M = I^M + I^F. \tag{2.3}$$

In (2.3) m denotes hourly daycare price, T_W^M denotes the mother's time spent on work, n denotes the hourly nanny cost, $I^M = w^M T_W^M$ and $I^F = w^F T_W^F$ denote the mother's and the father's labor income, respectively, and $\bar{T}$ denotes total time available to each member of the household.

I make three simplifying assumptions, justified by the data, to keep the model tractable and simple. First, by forcing households to use daycare while the mother is working, the model makes daycare cost m part of the mother's, but not the father's opportunity cost of time[17]; in addition, I assume that fathers supply labor full-time inelastically, thus $I^F = \tau^F \bar{T} w^F$, where τ^F is the fraction of $\bar{T}$ spent on work by fathers. As will be shown later, fathers' labor supply is unresponsive to a daycare price change. Second, I abstract from any bargaining and time reallocation between the mother and the father after a daycare price change, and assume that the mother spends a constant θ_K and θ_H fraction of total household time on the children and home production, respectively.[18] As will be shown later, mothers' and fathers' time use responses move not only in parallel after a change in the price

[17]This can be seen by rearranging the budget constraint as $B + X_H + X_L + \left(m^M - w\right)\left(T_K^M + T_H^M + T_L^M\right) + m^F\left(T_K^F + T_H^F + T_L^F\right) + nT_L^M = \left(w^M - m\right)\bar{T} + w^F\bar{T}$.

[18] Note that since at the solution $\frac{\theta_j}{1-\theta_j} = \left(\frac{w^M - m}{w^F}\right)^{\frac{1}{\rho_{Tj}-1}}$, this assumption implies that $\left(\frac{w^M - m}{w^F}\right)^{\frac{1}{\rho_{Tj}-1}}$ is constant. A slightly stricter assumption is that $\frac{\tau^F w^F}{w - m^M}$ is constant, that will be only used for recovering structural parameters.

of daycare, but their time weights in child human capital and home production are approximately constant. Since any further extension complicates the solution of the model without adding important insights, θ_K and θ_H is considered as constant. Finally, I assume positive assortative matching on wages: $\frac{\partial I^F}{\partial w^M} > 0$, despite that the father's labor income I^F is assumed to be unrelated to m.

Let α denote the fraction of X_K of other child market goods (than daycare), so that $\alpha X_K = B$. Then the household budget constraint (2.3) can be re-arranged in terms of the mother's variables:

$$\alpha X_K + X_H + X_L + \left(w^M - m\right)\theta_K T_K + \left(w^M - m\right)\theta_H T_H + \left(w^M - m + n\right)\frac{1}{2}T_L = \left(w^M - m\right)\bar{T} + I^F, \tag{2.4}$$

where $\left(w^M - m\right)\bar{T}$ denotes the household's total income (the income the household would receive in the hypothetical situation of the mother spending all her time on working). From (2.4) the opportunity costs (denoted by o_j) of the mother's different time use categories can be defined as: $o_K = o_H = w^M - m$ and $o_L = w^M - m + n$. To ease notation, from now on the mother's wage w^M will be denoted by w.

2.2.2 The Solution of the Model

As derived in the Appendix, the Marshallian demand for child goods and child time are:

$$X_K^* = \frac{\frac{1}{\alpha}\beta_K I}{1 + \left(\frac{(w-m)\theta_K}{\alpha\gamma_K}\right)^{\frac{\rho_K}{\rho_K - 1}}} \quad \text{and} \quad T_K^* = \frac{\beta_K I\left(\left(w - m\right)\theta_K\right)^{\frac{1}{\rho_K - 1}}\left(\alpha\gamma_K\right)^{\frac{\rho_K}{1-\rho_K}}}{1 + \left(\frac{(w-m)\theta_K}{\alpha\gamma_K}\right)^{\frac{\rho_K}{\rho_K - 1}}}, \tag{2.5}$$

while the Marshallian demands for home production H and leisure L are:

$$X_j^* = \frac{\beta_j I}{1 + \left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j - 1}}} \quad \text{and} \quad T_j^* = \frac{\beta_j I\left(o_j\theta_j\right)^{\frac{1}{\rho_j - 1}}\gamma_j^{\frac{\rho_j}{1-\rho_j}}}{1 + \left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j - 1}}}, \tag{2.6}$$

where $I = \bar{T}\left(w - m\right) + I^F$ is the potential household income, and the o_j-s are defined above.

The Marshallian demands depend on schooling through the wage w and time efficiency γ_j. This delivers an important contribution of this model: total differentiation of optimal cross-sectional choices and optimal responses with respect to schooling leads to a decomposition into two key channels: a wage channel and a time efficiency channel; generally, these two channels move into the opposite direction.

Some other models of daycare and labor supply incorporate daycare expenditures as a fixed cost paid by the household if the mother is working, independent of the work time T_W^M; thus daycare price m is subtracted from household income I as a lump sum part. Contrary to these other models, my model has the ability to predict different responses for child and home production time, depending on the signs of ρ_K and ρ_H; observe that this is due to the link between the net wage and the substitution parameters in the power in (2.5) and (2.6), and by making m part of the mother's opportunity cost.

2.2.3 Estimating Structural Parameters from the Model's Optimal Demands

In order to recover the structural parameters from the Marshallian demands (2.5) and (2.6), I need to impose a functional relationship between the time efficiency parameter γ and schooling. For individual i and good j $(j = K, H, L)$, I assume a multiplicative structure[19], that is $\gamma_{ji} = S_i^{\delta_j}\varepsilon_{ji}$. Then, (2.5) implies the following regression equation for $j = K$ between a non-linear function of the Marshallian demand, the logarithm of schooling S_i, and the logarithm of the net wage $w_i - m_i$:

$$\log\left(\frac{\frac{\beta_K}{\theta_K}\bar{T}\left(1+c\right)}{T^*_{Ki}} - 1\right) = \frac{\rho_K}{1-\rho_K}\log\theta_K + \frac{\rho_K}{\rho_K - 1}\delta_K \log S_i - \frac{\rho_K}{\rho_K-1}\log\left[\left(w_i - m_i\right)\left(1 + m_i^{\frac{1}{\rho_X - 1}}\right)\right] + \frac{\rho_K}{\rho_K - 1}\log\varepsilon_{Ki}, \tag{2.7}$$

where $c = \frac{\tau^F w^F}{w-m}$ is a constant, by the stricter assumption that θ_K is a constant. Then, by estimating the regression model in (2.7), the coefficient on the log function of the net wage can be estimated and ρ_K recovered. From ρ_K and the coefficient on $\log S_i$, δ_K can be recovered. In (2.7) the log of net wage $w_i - m_i$ is likely endogenous, but, as discussed in more detail in the identification section, can be instrumented by an exogenous policy change. The corresponding regression equation for $j = H$ is:

$$\log\left(\frac{\frac{\beta_H}{\theta_H}\bar{T}\left(1+c\right)}{T^*_{Hi}} - 1\right) = \frac{\rho_H}{1-\rho_H}\log\theta_H + \frac{\rho_H}{\rho_H - 1}\delta_H \log S_i - \frac{\rho_H}{\rho_H-1}\log\left(w_i - m_i\right) + \frac{\rho_H}{\rho_H - 1}\log\varepsilon_{Hi}.$$

2.2.4 Reduced Form Predictions of the Model

This section aims to answer three questions that get at the heart of the parental time-education puzzle. First, how do the optimal cross-sectional choices of market goods and time depend on schooling. Second, how do optimal choices change in response to a daycare price decrease. Third, how do these responses depend on schooling. The answers need to account for the following two channels operating simultaneously: first, holding the mother's wage w constant, higher-educated parents' time is allowed to be differentially efficient in child human capital production, home production and leisure production ("time efficiency channel"). Second, holding the time efficiency γ_j constant, higher-educated mothers' wage (thus, the opportunity cost of her home production, child and leisure time) is higher ("wage channel"). For simplicity, m is assumed to be identical across schooling.

The Dependence of Optimal Cross-sectional Choices on Schooling

To see how optimal choices of market goods depend on schooling, I totally differentiate the optimal choices X_j^* with respect to schooling S, and decompose $\frac{dX_j^*}{dS}$ into a time efficiency channel $\frac{\partial X_j^*}{\partial\gamma_j}\frac{d\gamma_j}{dS}$ and

[19] As an alternative functional specification, I assume an additive structure on the time efficiency: $\gamma_{Ki} = \delta_0 + \delta_K S_i + \varepsilon_{Ki}$; then, the three structural parameters δ_0, δ_K and ρ_K can be estimated with instruments S_i, policy_{tp} and $S_i \times \text{policy}_{tp}$, in a General Method of Moments (GMM) framework, where ε_{Ki} can be expressed as

$$\varepsilon_{Ki} = \left(\frac{\frac{\beta_K}{\theta_K}\bar{T}}{T^*_{Ki}}\left(1+c\right) - 1\right)^{\frac{\rho_K - 1}{\rho_K}}\left(w_i - m_i\right)\left(1 + m_i^{\frac{1}{\rho_X - 1}}\right)\theta_K - \delta_K S_i.$$

a wage channel $\frac{\partial X_j^*}{\partial w}\frac{dw}{dS}$ (for $j = K, H, L$, and similarly for T_j^*). The time efficiency channel shows how the optimal choices change after a marginal change in γ, holding the wage w constant. The wage channel shows how the optimal choices change after a marginal change in w, holding time efficiency γ constant. This total differentiation is helpful to think about the puzzle; the empirical estimates indicate which of the channels needs to dominate, given the education gradient observed in the data.

In Appendix A.1 I show that these derivatives can be decomposed into three terms: a base term, an assortative matching term, and a remainder term. In the main text, I focus on the sign of the base term, as the assortative matching term is always positive, and the remainder term is relatively small.

Proposition 1: *The base wage channel, unambiguously and independently of the sign of the elasticity of substitution parameters, is positive for all goods.*

Proposition 2: *The base wage channel is positive for child time if $\rho_K < 0$ (complementarity), and negative for home production time if $\rho_H > 0$ (substitutability).*

The intuition is that, *ceteris paribus*, higher-wage parents are more able to buy home production goods, and if inputs are substitutable, they spend less time on home production than lower-wage parents do. At the same time, higher-wage parents are also more able to buy child goods, and if inputs are complementary, they spend more time with their children than lower-wage parents do.

Proposition 3: *The base time efficiency channel for the child good is positive if $\rho_K < 0$ and $\delta_K > 0$; for the home production good it is negative if $\rho_H > 0$ and $\delta_H > 0$.*

Proposition 4: *The base time efficiency channel for child time is negative if $\rho_K < 0$ and $\delta_K > 0$; for home production time it is positive if $\rho_H > 0$ and $\delta_H > 0$.*

The intuition behind these propositions lies in the production technologies: if there is complementarity between inputs in child human capital production, then if time efficiency increases, the same level of child human capital can be attained by slightly increasing the level of child goods and decreasing child time more, thereby achieving cost-savings. Similarly, suppose there is substitutability in home production; then, if time efficiency increases, the same level of home production can be attained by slightly increasing home production time and decreasing more the level of home production goods.

The model's explanation for the parental time-education puzzle is the wage channel dominating the time efficiency one: keeping time efficiency constant, higher-wage parents are more able to afford all goods, and if there is complementarity in child skill formation, they also spend more time with their children, despite their time efficiency advantage in child human capital production. Thus, the puzzle can be explained without introducing heterogeneity in time efficiency across education groups.

Response to a Daycare Price Decrease

How do optimal choices respond to decreasing daycare prices? First, a daycare price decrease generates a dominating income effect, so demand for all market good increases—these results do not hinge on the sign of the substitution parameters. The effect of a marginal decrease in daycare price on time spent with the child or on home production almost exactly mirrors the effect of a marginal increase in wage; the two effects would be identical but for the fact that fathers' income is not dependent on m and $1-\alpha$ (α is a function of the relative price of daycare and other child goods). Therefore, when remainder terms are sufficiently small, in response to falling daycare prices, home production time decreases and parental time increases—note that these results do hinge on the signs of the ρ's.

Proposition 5: *A decrease in the price of daycare induces parents to increase their demand for all (child, home production and leisure) goods.*

Proposition 6: *If $\rho_K < 0$ (complementarity), parents increase their child time when daycare price falls. If $\rho_H > 0$ (substitutability), parents decrease their home production time when daycare price falls.*

The Dependence of Optimal Responses on Schooling

How do responses to a daycare price decrease depend on schooling? To answer, let us totally differentiate the good responses $\frac{\partial X_j^*}{\partial(-m)}$ and time responses $\frac{\partial T_j^*}{\partial(-m)}$ with respect to schooling, and decompose the change into a wage channel and a time efficiency channel.

Proposition 7: *Suppose that $\delta_K > 0, \rho_K < 0$ (complementarity) and $\delta_H > 0, \rho_H > 0$ (substitutability). Then the time efficiency channel is positive for the child human capital good, and is negative for the home production good if $\rho_H < \frac{1}{2}$.*

Proposition 8: *If $\rho_K < 0$ and $\rho_H > 0$, the wage channel is negative for the child human capital good, and is positive for the home production good if $\rho_H < \frac{1}{2}$.*

Proposition 9: *Suppose that $\delta_K > 0, \rho_K < 0$ (complementarity),$\gamma_K < 1$ and $\delta_H > 0, \rho_H > 0$ (substitutability) and we exclude extremely large values of γ_H. Then the time efficiency channel is positive for child time, and is negative for home production time.*

Proposition 10: *If $\rho_K < 0$ and $\rho_H > 0$, the wage channel is negative for child time, and is positive for home production time if $\rho_H < \frac{1}{2}$.*

This model reveals and emphasizes the critical role played by complements and substitutes in household choices about children. Consider the case of complementarity between inputs in the production of K; then, for child good responses, the time efficiency channel is positive, while the wage channel is negative. Similarly for child time responses, the time efficiency channel is positive: keeping wage constant, parents with more efficient parental time will increase their child time more. At the same time, the wage channel is negative: *ceteris paribus*, parents with higher wage increase their parental time less. Thus, higher-educated parents increase time spent with their children more if the former channel dominates, i.e. if their time efficiency advantage in child human capital production is large enough to outweigh their higher opportunity cost of time in non-market activities.

For home production goods, for analytical convenience consider the case when $\rho_H < \frac{1}{2}$; then the time efficiency channel is negative, while the wage channel is positive. Regarding time responses, the time efficiency channel is negative: keeping wage constant, parents with more productive home production time will decrease their home production time more. On the other hand, *ceteris paribus*, parents with higher wages (thus higher opportunity cost of time) will decrease their home production time less. Higher-educated parents decrease more their home production time if their time efficiency advantage is large enough to outweigh their higher wage.

2.3 Institutional Background and Available Evidence

In this section I describe the policy change that provides identifying variation for my empirical strategy. I also review existing empirical evidence, emphasizing the gaps my work seeks to fill.

2.3.1 The Quebec Daycare Policy Change (1997)

To enhance mothers' labor force participation, child development, and equality of opportunity, in 1997 the government of Quebec granted children aged 0-4 universal access to centre-based or home-based government-provided, regulated, institutional daycare[20] at an out-of-pocket price of $5 per day. The access was universal, irrespective of the parents' labor market status, and without entry requirements or means-testing. The phase-in was gradual by age: in 1997 all 4-year olds were exposed to the policy, and in the three consecutive years exposure was extended to all 3-, 2-, and less than 2-year olds.

There were quantity and quality changes, as well as operational reforms associated with the policy change. First, approximately 65,000 extra regulated daycare spaces were opened between 1998 and 2001, then an additional 90,000 until 2007 (Lefebvre and Merrigan, 2008).[21] Second, in 2000 the educational requirements for the regulated daycare institutions' staff were substantially increased and their wages were scheduled to increase by 35-40 percent over a four-year period. Finally, the maximum facility size was increased by holding staff-to-child ratios fixed (with the exceptions of 4-5 year old children) and parental involvement in the board of directors increased (Baker et al., 2008). The quality audit study by Japel et al. (2005) shows that more (less) centre-based and home-based daycare providers provided good-quality (inadequate-quality) services than did for-profit daycares and unregulated for-profit daycares, by primarily excelling in the quality of the interactions between staff and children, and less in the quality of the educational activities and the personal care routines.

[20] Institutional care incorporates daycare centers, nursery school/preschool and before/after schools.

[21] According to Baker et al. (2008), the transition to the new system around 1997 happened with frictions, and the local government responded to the excess demand by creating new subsidized places; they refer to media mentions suggesting a queue of 35,000 children initially. However, there is limited information available on the characteristics of the queuing: in the NLSCY only in waves 7 and 8—thus only in the post-policy period—there is information on Quebec-residents attending the subsidized daycare program, and the reasons for not attending, including not enough space.

As Baker et al. (2008) document, this policy changed the financial incentives primarily for richer families, as direct daycare subsidies for poorer families and a refundable tax credit, depending on family income, were already available before the policy change. Parents who received a subsidized daycare spot were not eligible for any further direct subsidy and provincial tax credit for daycare expenses, but remained eligible for a federal deduction. Prior 1997 low-income families were eligible for direct daycare subsidies, and single women typically qualified for substantial subsidies. In their Appendix A, Baker et al. (2008) review the family tax credits in Quebec and Canada, including, for instance, the two types of income-dependent refundable tax credits (the Canada Child Tax Benefit until 1999, and the National Child Benefit Supplement, introduced in 1998), and the Quebec Family Allowance, that changed in 1997, moving from universal to income-tested and targeted allowances. They also graph the effective subsidy of daycare prices by province over the 1990s, separately for parents in two-parent families with at least one child below the age of 5, and for single mothers; they show that the impact of the policy is by 50% smaller for the latter. These two reasons–substantial subsidies for singles prior, and contemporaneous policy changes of the Quebec Family Allowance–led them to exclude single parents from their analysis, and I follow their sample selection.

At the same time, Lefebvre and Merrigan (2008) highlight that possibly liquidity-constrained low-income families, who might have had problems in accessing reliable daycare services before, may have made use of the new daycare regime. They also emphasize the features of the new institutional daycare places, which many parents might have preferred over the relative-provided, non-licensed daycare, as being available for longer hours, being licensed and regulated. Japel et al. (2005) find that in centre-based daycare children received services of similar average quality irrespective of family background, while in regulated and unregulated home-based daycare and for-profit daycares attended by disadvantaged children were of lower quality.

2.3.2 Available Evidence on the Quebec Daycare Policy Change (1997)

The first set of empirical evidence is clear: the policy increased maternal labor supply and daycare use.[22] Baker et al. (2008) show for two-parent families that being eligible for this program increased daycare usage, primarily in institutional care and care in other's home (provided by a licensed non-relative) and increased maternal labor supply, with the fraction of working mothers having their child in daycare increasing the most. Lefebvre and Merrigan (2008) use data from the Survey of Labor and Income Dynamics (SLID) for 1993-2002, and find that the policy had a positive effect on the short-term labor supply of mothers (on both margins and earnings), who had at least one child aged 1-5; the effects are significant for mothers with more than a high school diploma. Lefebvre et al. (2009) find that the policy had long-term labor supply effects for mothers, who benefited from the program when their child was less than 6 years old; the results are driven by less educated mothers.

The second set of empirical findings is controversial: Baker et al. (2008) show for two-parent families that, on average, program eligibility led to worse child outcomes and parenting practices,

[22] Empirical evidence on maternal labor supply from other countries include, e.g., Blau and Tekin (2007) and Tekin (2007) for daycare subsidies on single mothers, Fitzpatrick (2010) and Havnes and Mogstad (2011) for universal childcare, and Bauernschuster and Schlotter (2015) for public child care. Baker (2011) and Cascio (2015) provide a summary on the evidence base of universal childcare.

increased maternal depression and family dysfunctioning.[23] Kottelenberg and Lehrer (2013) substantiate these results by including more treated cohorts receiving the treatment when daycare centers were better established, and supply constraints were less binding. Baker et al. (2015) show that these non-cognitive deficits persisted to teen ages, and that cohorts with increased daycare access had higher crime rates, worse self-reported health, and lower life satisfaction. Kottelenberg and Lehrer (2016) find that parents in two-parent households of 4-year-old children, who were induced to increase daycare use by the policy change, increased their propensity of reading to the child daily, while parents of 0-3-year old children decreased their daily reading propensity. Brodeur and Connolly (2013) find that the policy decreased the level of happiness of married women, with no impact on their life satisfaction.

I show evidence on the impact of program eligibility in Quebec on a new set of parental time allocation outcomes, such as total time spent with the child, home production time and leisure time. In addition, I complement the parental quality time outcomes used by Kottelenberg and Lehrer (2016); I confirm that that program eligibility decreased parents' propensity to read their child daily, but show that it increases reading time at the lower end of the reading distribution, by decreasing the propensity to never read to child, and increasing the propensity to read once or several times per week.

Regarding empirical evidence in other policy contexts, my set of time use outcomes are most comparable with the outcomes studied by Cascio and Schanzenbach (2013). They find that high-quality universal pre-school availability in Georgia and Oklahoma induced low-educated mothers (having at most a high school degree) to decrease total time spent with their children present, and to increase time spent caring for or helping to children, with no detectable impacts for high-educated mothers. However, they do not find robust labor supply impacts for low-educated mothers, and do not detect any impacts for higher-educated mothers; as shown in Section 2.6.2., the time responses I detect for high-educated mothers are triggered by an increase in their labor supply. In order to look at the impacts of increasing incentives to work, Gelber and Mitchell (2012) exploit income tax changes between 1975 and 2004 using data from from the Panel Study of Income Dynamics, the American Time Use Survey and Consumers Expenditure Survey. They find that changes in the Average Net-of-Tax Rate substantially increases single mothers' hours worked, at the expense of their home production and leisure time—the same substitution pattern I show in Section 6. However, contrary to my findings, they report an insignificant positive point estimates for child time (with a value of t-statistics 1.4).

To the best of my knowledge, this is the first paper on the impact of universal daycare availability on household expenditures on child goods, food, and on goods that help parents substitute their time out from home production, such as eating out or hiring domestic help. My results align with the results of Gelber and Mitchell (2012), who find that, in response to increasing returns to work, expenditures on food prepared away from home and at home significantly increase and decrease, respectively, with no significant impact on hiring domestic services (although their point estimate is positive).

[23]Empirical evidence on child development include, for instance, Carneiro and Ginja (2014) from Head Start, Berlinski and Galiani (2007) from Argentina, Havnes and Mogstad (2015) from Norway, Datta Gupta and Simonsen (2010) from Denmark and Dustmann et al. (2013b) from Germany. Empirical evidence on maternal mental health and child-parent interactions include Herbst and Tekin (2014).

2.4 Empirical Approach

In this section I describe my empirical framework. To keep it as simple as possible, I start with a standard Difference-in-Differences (DiD) framework. I then present the estimation equations to recover structural parameters of the model. Next, I test the widening behavioral skill gap by age; to see through which channel maternal education is most related to child development, I assess how the inclusion of parental investment measures explain the gap. Finally, I discuss identification issues.

2.4.1 Differential Reduced-Form Impacts of the Quebec Daycare Policy

First, I assess the overall and the differential impacts of the "$5/day" Quebec daycare policy by a DiD identification strategy, where I estimate Intent-to-Treat (ITT) effects—the effects of program availability—in a repeated cross-section structure.[24] I test whether the policy impact differs for higher-educated families: I differentiate families where the mother obtained—in addition to a high-school degree—some post-secondary, college-level studies (either finished or unfinished), from families where the mother obtained a high-school degree but had not studied at the post-secondary level. Using children aged 0-4 in two-parent families[25] in the NLSCY, the Census and the LFS, and two-parent families with at least one child aged 0-4 in the GSS and the SHS, I estimate the following two models:

$$Y_i = \alpha_0 + \alpha_1 \text{policy}_{tp} + \alpha_2 C_i + \alpha_3 U_i + \alpha'_{4t} + \alpha'_{5p} + \alpha'_{6a} + \alpha'_7 X_i + \nu_i, \tag{2.8}$$

$$\begin{aligned} Y_i \quad &= \beta_0 + \beta_1 \text{policy}_{tp} + \beta_2 \text{policy}_{tp} \times \text{high-educ}_i + \beta_3 U_i \\ &+ \beta'_{4t} + \beta'_{5p} + \beta'_{6a} + \beta'_{7te} + \beta'_{8pe} + \beta'_9 X_i + \varepsilon_i. \end{aligned} \tag{2.9}$$

where i indexes household, t time, p province, and a indexes the child's age; Y is either a daycare use, a parental labor supply, a parental time use, a household expenditure, a home-environment, a child developmental or a parental health outcome; policy_{tp} is an interaction between the child being born in an exposed cohort (alternatively, the family observed in the post-policy period) and residing in Quebec; C indicates that the mother obtained a high school degree and had some post-secondary, college-level studies that is at lower level than a Bachelor's degree (including CEGEP, community college and trade/technical degrees); U indicates that the mother obtained at least a Bachelor's degree. In what follows, I will refer to high-educated families (high-educ = 1), where either $C = 1$ or $U = 1$. α_{4t} and β_{5t} correspond to a full set of year (or wave) dummies, α_{5p} and β_{6p} correspond to a full set of province dummies, α_{6a} and β_{7a} correspond to a full set of age dummies, β_{8te} corresponds to a vector of education-specific dummies, indicating the post-policy period, β_{9pe} corresponds to a vector of education-specific indicator-variables, indicating residence in Quebec and X includes the age and gender composition of the children, household size, and the parents' age. The household's education is determined solely by the mother's education, and I do not control for the father's education.[26]

[24] The standard DiD specification I implement is the same as implemented by Baker et al. (2008), Kottelenberg and Lehrer (2013) and Lefebvre and Merrigan (2008).

[25] Based on the discussion in Section 2.3.1 and following Baker et al. (2008), I exclude single parents from the analysis.

[26] Results are essentially unchanged when I do control for the father's education, or when I determine the household's education based on the interaction between the mother's and the father's education.

Standard errors are clustered at the (province×post)-level, to account for within-province correlation of errors over time both in the pre-policy and post-policy period, allowing for a structural break in the temporal correlation at the time of policy implementation.[27]

Since the available data structure is not panel, but repeated cross-section, I do not observe pre-policy work history of mothers. To see whether the responses differ based on the mother's propensity to work in the absence of the policy, I estimate additionally the following model:

$$\begin{aligned} Y_i \; &= \gamma_0 + \gamma_1 \text{policy}_{tp} + \gamma_2 \text{policy}_{tp} \times \text{propensity}_i + \gamma_3 \text{policy}_{tp} \times \text{propensity}_i^2 \\ &+\gamma_4 \text{propensity}_i + \gamma_5 \text{propensity}_i^2 + \gamma_6 C_i + \gamma_7 U_i + \gamma_{8t}' + \gamma_{9p}' + \gamma_{10}' X_i \\ &+\gamma_{11} \text{post}_t \times \text{propensity}_i + \gamma_{12} \text{Que}_p \times \text{propensity}_i + \eta_i, \end{aligned} \tag{2.10}$$

where *propensity* denotes the propensity of the mother working in the absence of the policy; to obtain it, I estimate a probit model on the pre-policy sample using pre-determined variables, and I predict the propensities for the whole sample. *post* equals 1 if the child was born in an exposed cohort (alternatively, if the family is observed in the post-policy period), and *Que* equals 1 if the household resides in Quebec. γ_1 shows the policy impact for mothers with zero predicted propensity. A significantly positive estimate of γ_2 indicates that the policy impact is increasing with the predicted propensity to work in the absence of the policy. The propensity score is predicted by pre-determined variables not altered by the policy change, as by the interaction of parents' education, a full set of province and child age fixed effects, the parents' age, household size, and a linear time trend.[28]

2.4.2 Recovering Structural Parameters

In Section 1.2.3 I showed that by assuming a multiplicative functional form on the good-specific productivity errors, the Marshallian demands can be transformed into the regression equation (2.7), where the coefficient on the schooling and net wage variables are nonlinear functions of the structural parameters ρ_j (the substitution parameter) and δ_j (the time efficiency parameter), for $j = K, H, L$ I aim to recover. In order to recover ρ_K and δ_K, the specific model I estimate by Ordinary Least Squares (OLS) and Two-Stage Least Squares (2SLS) methods is the following:

$$\log\left(\frac{\frac{\beta_K}{\theta_K}\bar{T}(1+c)}{T_{Ki}^*} - 1\right) = \lambda_0 + \lambda_1 \log S_i + \lambda_2 \log\left[(w_i - m_i)\left(1 + m_i^{\frac{1}{\rho_X - 1}}\right)\right] + \lambda_{3t}' + \lambda_{4p}' + \lambda_5' X_i + \xi_i, \tag{2.11}$$

where $\lambda_1 = \frac{\rho_K}{\rho_K - 1}\delta_K$, $\lambda_2 = -\frac{\rho_K}{\rho_K - 1}$, and the error term is $\xi_i = \frac{\rho_K}{\rho_K - 1}\log\varepsilon_{Ki}$; S is 2 if the mother obtained at least some post-secondary education and 1 otherwise, w is the hourly wage of the mother and m is the hourly daycare expenditure[29], λ_{3t} corresponds to a full set of year dummies, λ_{4p} corresponds to a full set of province dummies, and X includes the mother's and father's age, and full set of dummies

[27] The confidence intervals for the main estimates in Appendix Tables A.20-A.23 show that the standard errors are robust to the Wild-bootstrapping method of Cameron et al. (2008), accounting for small number of clusters.

[28] The estimation details in the NLSCY, the Census and the GSS datasets can be seen in Appendix Table A.24.

[29] Since the GSS does not contain information on wages or daycare expenditures, while the Census does not contain information on daycare expenditures, these variables needed to be imputed; details are available in the Appendix.

corresponding to the age and gender composition of the children. Similarly, in order to recover ρ_H and δ_H, the specific model I estimate by OLS and 2SLS methods is the following:

$$\log\left(\frac{\frac{\beta_H}{\theta_H}\bar{T}(1+c)}{T^*_{Hi}}-1\right)=\lambda_0+\lambda_1\log S_i+\lambda_2\log\left(w_i-m_i\right)+\lambda'_{3t}+\lambda'_{4p}+\lambda'_5X_i+\xi_i. \tag{2.12}$$

I do not estimate the parameters of $\frac{\beta}{\theta}$, ρ_X or $c=\frac{\tau^F w^F}{w-m}$, but assume the value of $\frac{\beta_K}{\theta_K}=\frac{0.3}{1}$, $\frac{\beta_H}{\theta_H}=\frac{0.2}{1}$, $\rho_X=-1$ and $c=0.6$, and perform sensitivity analysis.[30]

The excluded instrument is policy$_{tp}$, an interaction between the child being born in an exposed cohort—alternatively, if the family observed in the post-policy period—, and residing in Quebec. Thus, the corresponding first-stage in case of T^*_{Ki} is

$$\log\left[\left(w_i-m_i\right)\left(1+m_i^{\frac{1}{\rho_X-1}}\right)\right]=\varsigma_0+\varsigma_1\log S_i+\varsigma_2\text{policy}_{tp}+\varsigma'_{3t}+\varsigma'_{4p}+\varsigma'_5X_i+\iota_i. \tag{2.13}$$

2.4.3 The 'Human Capital Gap-Regressions'

I next examine the importance of parental investments in explaining human capital gaps in early childhood. To highlight the importance of the chosen age range and to justify the focus on parental (quality) time in this chapter, I assess the widening behavioral human capital gap by age and I will refer to them as 'gap-regressions'. Using children aged 0-11 in the NLSCY, I run the following regression:

$$\begin{aligned} y_i \quad &= \tau_0+\tau_1\text{high-educ}_i+\tau_2\text{high-educ}_i\times\text{age}(3-5)_i+\tau_3\text{high-educ}_i\times\text{age}(6-8)_i \\ &\quad +\tau_4\text{high-educ}_i\times\text{age}(9-11)_i+\tau_5 HS_i+\tau'_{6t}+\tau'_{7p}+\tau'_{8a}+\tau'_9X_i+\iota_i, \end{aligned} \tag{2.14}$$

where i indexes household, t indexes time, p indexes province, a indexes the child's age, y is the average age-standardized behavioral score (measured by hyperactivity), high-educ indicates high-educated, defined as the mother having either some college, or university education, age$(3-5)$ indicates the child being between 3 and 5 years old, and HS indicates the mother having at most a high school degree. τ_{6t} corresponds to a full set of year (or wave) dummies, τ_{7p} corresponds to a full set of province dummies, τ_{8a} corresponds to a full set of child's age dummies, and X includes the gender of the child, number of older and younger siblings (capped at 3 and 2, respectively), the size of the household, the mother's and father's age, and the father's education. Standard errors are clustered at the (province$\times$post)-level.

The coefficients of interest are $\tau_1-\tau_4$, corresponding to the behavioral/developmental human capital gap—also called the noncognitive gap—across age-categories. τ_1 shows the gap for children

[30] ρ_X is the substitution parameter between daycare time and other market goods in X_K; since empirically I find that both of them increase as daycare prices fall, I assume a negative value -1 for ρ_X. β_K and β_H are taken to be 0.3 and 0.2, respectively, and since I use the mother's time in the estimation, I consider $\theta_K=\theta_H=1$. Assuming $\tau^F=\frac{1}{4}$, 0.6 corresponds to average wages observed in the data. The sensitivity analysis with respect to alternative parameter values is in Appendix Tables A.25 and A.26.

aged 0-2, while $\tau_1 + \tau_2$ shows the gap for children aged 3-5; so, for instance, τ_2 shows whether the gap significantly widens after the first 3 years of life. I present the estimation results of (2.14), and assess how the inclusion of parental practices, maternal health, maternal employment, daycare attendance and parents' reading practices change the estimated coefficients $\hat{\tau}_1 - \hat{\tau}_4$.

The aim of this solely descriptive exercise is to see through which channel maternal education is most correlated with child development; if, for instance, the estimated coefficients shrink by including parental reading practices, that is an indication of parental reading being an important transmission mechanism between socio-economic background and child development. Recognizing these channels not only help us understand which parental investment measures are crucial in shaping the gap, but also help us to see in which measures educational differences matter the most. This approach is similar to controlling for mother's ability, family income, and family structure to see by how much the average anti-social behavior score percentile by income quartile or race is reduced, chosen by Cunha et al. (2006); or, reporting the 'conditional difference' by Baker and Milligan (2016), by controlling for some observable characteristics after reporting the means of test scores at ages 4-5 across gender.

2.4.4 Identification

The main requirement for the DiD model to provide a consistent estimator of the policy impact is that the counterfactual time trend in outcomes in Quebec is parallel to the observed time trend in the rest of Canada (i.e. the trends in the absence of the policy would have been parallel in and outside of Quebec). Figures 2.1 and 2.2 show the time trends in the mother's employment rate, working hours and regulated daycare use, respectively. The pre-trends for Quebec and the rest of Canada are either parallel or converging for labor supply, and slightly diverging for daycare use; however, once the policy is implemented the trends diverge sharply.

Besides the aforementioned graphs suggesting that the parallel trend assumption is reasonable, Baker et al. (2008) provide evidence that there were no detectable differential trends for several important demographic/control variables in Quebec, relative to the rest of Canada. Kottelenberg and Lehrer (2013) document similar trends for maternal labor supply and daycare usage in Quebec, relative to the rest of Canada, before the introduction of the policy. To address any Quebec-specific shocks coincident with the Quebec daycare policy, (i) I check and confirm that the estimates shown in Section 6 are robust to the inclusion of province-specific economic conditions, (ii) that there are no policy impacts on older children who were never exposed to the policy, and who do not have any younger siblings exposed, and (iii) there are no or substantially lower impacts on household food and domestic expenditures for households without children.[31]

When estimating the structural parameters, there is a concern in the models (2.11) and (2.12) that the variable on wage is potentially endogenous. It is plausible that innate parental ability, although orthogonal to education, is related to the wage through some general ability. The model also provides guidance on the sign of the OLS estimator's asymptotic bias of λ_2, and predicts an upward bias for child time, based on the following reasoning. First, holding wage and education constant, parents

[31] The robustness and placebo check results of (i) and (ii) are available on request.

with better parental abilities are predicted to spend less time with their children; thus, *ceteris paribus*, there is a positive correlation between the error term and the outcome variable $\log\left(\frac{\frac{\beta_K}{\theta_K}\bar{T}(1+c)}{T^*_{Ki}} - 1\right)$, in which optimal parental time T^*_{Ki} is in the denominator. Second, there is a concern of a positive correlation between the function of the net wage and parental ability in the error term.

However, this potentially endogenous variable can be instrumented with policy_{tp}, the policy variable, provided that $Cov\left(\log \varepsilon_{Ki}, \text{policy}_{tp}\right) = 0$, where policy_{tp} equals 1 if the household lives in a province with a daycare policy in effect, and 0 otherwise. The identification assumptions are that the policy change was orthogonal to parental ability, but related to net wage. Thus, the source of the identifying variation is the same in the reduced-form, as in the parameter estimation. However, while the first set of estimation results suggest the signs of the structural parameters based on the model, the second set of results reveal their magnitude, to perform policy simulations on childhood skill gaps.

2.5 Data, Measurement and Sample Selection

In this section I provide basic information on the five Canadian datasets used in the main analysis in Section 6. I then describe the measurement of the outcome variables, as well as the eligibility variable.

2.5.1 Datasets

To measure the total effect of a daycare price change on parental and child outcomes, I use Canadian data from five sources: the first seven waves of the NLSCY, the 1996/2001/2006 waves of the Canadian Census, the 1994-2006 cycles of the Canadian LFS, the cycles from 1986/1992/1996-2009 of the publicly available SHS (previously named as Survey of Family Expenditures or FAMEX), and the cycles from 1998, 2005 and 2010 of the Time Use Diary of the GSS. 2-11 and 0-4 years old children in two-parent families are selected from the NLSCY, the Census and the LFS, and two-parent households with having at least one 0-4 year old child are selected from the GSS and the SHS.

The NLSCY follows the development and well-being of Canadian children from birth to early adulthood and collects information about factors influencing a child's social, behavioral and emotional development. An initial sample of 0-11 years old children was sampled in the first wave in 1994 and followed for fourteen years until 2008; I do not use the panel structure. Starting at Cycle 2 a new cohort of 0-1 years old children was added in each cycle, following the expansion of the NLSCY emphasizing early childhood development. Data in the first cycle are representative of children between 0-11 in 1994, data in the second cycle are representative of children between 0-13 in 1996, and so on.

The LFS provides detailed labor market data on the civilian population 15 years of age and over (excluding persons living on Aboriginal reserves, full-time members of the Canadian Forces and the institutionalized population). Basic demographic information, such as gender and exact date of birth, is available for all family members in the household. The LFS uses a rotating panel sample design: selected households remain in the sample for six consecutive months.[32]

The two main objectives of the GSS are to collect data on social trends to monitor changes in

[32] More information is available at http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3701.

the living circumstances of Canadians, and to collect responses on a rotating set of particular topics.[33] One of the rotating topics is time use, measured in 1998, 2005 and 2010. The target population consists of all non-institutionalized persons 15 or older, living in the Canadian provinces. Until 1998 the target sample of respondents included around 10,000 individuals, before 1998 around 25,000.[34]

Besides household expenditures, the SHS collects information on annual income of household members from personal income tax data, demographic characteristics of the household, certain dwelling and household equipment characteristics. The SHS combines a questionnaire with recall periods based on the type of expenditure and a daily expenditure diary. The target sample is similar to the LFS'.[35]

2.5.2 The Measurement of Outcome Variables and Eligibility

The age-standardized behavioral score for ages 2-11 is the age-standardized hyperactivity score from the NLSCY, where larger values indicate worse behavioral/developmental outcomes as more severe hyperactivity. The daycare use measures—also from the NLSCY—are all binary variables and indicate whether the child is in institutional care (daycare centers, nursery school/preschool and before/after schools), daycare in own home (provided by a relative/non-relative), daycare in others' home (provided by a relative/non-relative), or in any care from the aforementioned ones. The home-environment measures from the NLSCY include age-standardized[36] hostility/consistency/positive parenting scores[37] the age-standardized depression score of the mother (a high score indicating the presence of depression symptoms), and the age-standardized family functioning score[38]. The parental labor supply measures, stemming from the NLSCY, the Census, and for robustness checks from the LFS, include an indicator variable of being employed, an indicator variable of being employed part-time, and actual hours worked. The time outcomes include work-related time[39], child time, home production time and leisure time; measured in fraction of all time (in percentages) in the GSS Time Use Diary, and in categories of hours spent per week in the Census. The household expenditures (from the SHS) include the fraction of food expenditures (all/from store/from restaurant), of expenditures on domestic help, and on the child as daycare expenditures and expenditures on toys and games, out of all expenditures (in percentage).

Eligibility is defined as follows: in the NLSCY, no 0-4-years old child is eligible in waves 1 and 2 (1994 and 1996); in wave 3 (1998) the 4-year old children were eligible, while in later waves (2000,

[33] These include commuting to work, labor market outcomes and perceptions, society and community, time use and unpaid work, and are usually repeated every 5 years.

[34] More information is available at http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=4503. Time devoted to the child includes putting children to bed, getting them ready for sleep, personal care, helping/teaching/reprimanding/reading to children, talking to them/having conversations with them, medical/emotional care. Work time includes work for pay at main job and other jobs, overtime work, looking for work, unpaid work in a family business, travel during work, waiting/delays at work, meals/snacks at work, idle time before/after work, coffee/other breaks at work, travel to/from paid work. Leisure time includes personal care, relaxation, entertainment, social/educational/civic/volunteering/communicational/sport activities, reading/watching television/listening to media.

[35] More information is available at http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3508.

[36] Note that age-standardization is with respect to the child's age.

[37] This last measure captures the frequency of positive encounters with the child, as laughing with or praising her.

[38] This scale is used to measure various aspects of family functioning, e.g., problem-solving, communications, roles, affective involvement, affective responsiveness and behaviour control by capturing how cooperatively and constructively the family solves problems, how close the relationship is between the parents, how much is alcoholism a problem in the family, etc.

[39] Work-related time includes not only hours of actual work, but commuting and other work-related time.

2002, 2004 and 2006) all children are eligible. In the LFS, no 0-4 year old child is eligible before 1997; 4-year old children are eligible in 1997 and 1998, 2-4-year old children are eligible in 1999, and all children are eligible from 2000 on. In the GSS, pre-policy year is 1998, post-policy years are 2005 and 2010. In the Census, pre-policy year is 1996, post-policy years are 2001 and 2006. In the SHS, pre-policy years are 1986-1997, post-policy years are 1998-2009.

2.6 Results

I first present the impact of decreasing daycare prices on maternal labor supply, daycare use, parental time use, and household expenditures, for all observations and also by education. Next, I show the policy impacts by the mother's predicted propensity of working in the absence of the policy. Then, I assess the widening skill gap across ages. Finally, I present the estimated structural parameters, along with counterfactuals on how the behavioral gap would evolve under targeted or universal daycare regimes, or if high-educated (low-educated) parents had no time efficiency advantage (disadvantage).

2.6.1 Reduced Form Results: Policy Impacts for All and by Education

In each table from 2.1 to 2.6, I show the basic DiD results for numerous outcomes. Table 2.1 confirms existing evidence on daycare use and maternal labor supply responses to decreasing daycare prices. Tables 2.2 to 2.6 show the empirical contributions of this chapter on the following responses: parental time use, the propensity to read to the child and household expenditures on food, domestic help, daycare and child toys. The first row of each table shows the outcome variables (in bold), *Panel A* shows the coefficients and standard errors on the policy and education coefficients from estimating the DiD models (2.8) and (2.9), where, for each outcome variable, the first column corresponds to (2.8), and the second column corresponds to (2.9). The coefficients on the interaction variables show whether there is a differential policy impact by socio-economic status, measured by the mother's education. In particular, they show whether the policy impact is significantly different for families where the mother has some post-secondary education or a university degree from families where the mother has at most a high school degree without any post-secondary studies. *Panel B* shows the weighted baseline means and the estimated policy impacts, by education, to see the size of the policy impact; for each outcome variable, the first column shows the weighted baseline mean in the estimation sample for observations where $\text{policy}_{tp} = 0$, while the second column shows the estimated policy impact. The *p-values* in *Panel C* show whether the policy significantly impacted high-educated families, in terms of the corresponding outcome variable; i.e. it corresponds to the test: $H_0 : \beta_1 + \beta_2 = 0$; $H_1 : \beta_1 + \beta_2 \neq 0$.[40]

As a reminder, the model predicts the following after a decrease in the price of daycare: first, demand for home production goods, child human capital goods and leisure goods increases. If inputs in home production are substitutable, while in child human capital production they are complementary,

[40] To see the policy impact graphically, Figures A.1, A.2 and A.3 show the estimated difference between Quebec and the Rest-of-Canada across years, with a 95% confidence band, stemming from a variant of model (2.8); in these models, instead of interacting the eligibility-by-cohort indicator variable with indicator variable for Quebec, year-indicators are interacted with the variable indicating residence in Quebec.

then, as a response, home production time decreases, while parental time increases. This time reallocation is larger for high-educated parents, if their time efficiency advantage in non-market activities is sufficiently large to outweigh their higher opportunity cost of time in non-market activities.

Daycare use

Table 2.1 shows that, consistent with previous evidence and the aim of the policy, the Quebec daycare policy significantly increased daycare utilization in regulated (institutional) daycare and in any care by 18 and 13 percentage points (200 and 29 percent of the baseline mean), respectively.[41] The policy impact on institutional daycare utilization is significantly larger for high-educated families in absolute terms, but not relative to their baseline mean (it is 247 and 190 percent for low- and high-educated families, respectively). The policy impact on daycare in any care is again significantly larger for high-educated families, but relative to their baseline mean the impact is approximately 29 percent for both groups.[42] Despite the larger change in the financial incentives for high-educated, the relative impact on any daycare use is the same for low- and high-educated groups; hence differential intent-to-treat results can be attributed to differential impact, as opposed to differential uptake of the policy.

Maternal Labor Supply

Table 2.1 (and the Appendix Tables A.4-A.5) show the policy impact on parents' labor supply along the extensive margin using data from the NLSCY (and, for robustness check from the Census and LFS); a daycare price decrease significantly increased the mothers' immediate propensity of being employed, by 3.4 to 7.6 percentage points (by 6-13 percents relative to the baseline mean).[43] These effects are driven by high-educated mothers both in absolute and relative terms; low-educated mothers are more likely to work by 1-4 percentage points (significant depending on sample size).[44] The impact on the father's labor supply is considerably lower (0.6-1.3 percentage points, on a base of 88 percent).

Table 2.1 (and the Appendix Tables A.4 and A.6) show that a daycare price decrease significantly increased mothers' working hours, by 1.2 hours on a base of 20 hours[45]; the increase is driven by high-

[41] Appendix Table A.1 shows that the Quebec daycare policy significantly decreased daycare usage in own home and in other's home by 3 and 2.4 percentage points (13 and 22 percent of the baseline mean).

Using waves 1-2 and 4-5 of the NLSCY, Baker et al. (2008) find that, after the implementation of the policy, the proportion of 0-4 year olds in any daycare rose by 14 percentage points in Quebec relative to the rest-of-Canada, that is very close to my point estimate of 0.131, using waves 1-7. Using waves 1-2 and 5-7, Kottelenberg and Lehrer (2013) find a point estimate of 0.196; this considerably larger point can be understood in light of Figure 2.2, as regulated (institutional) daycare use—that the policy targeted—rose the most between waves 3 and 5, and then 6 and 7.

[42] On the intensive margin, Appendix Table A.2 shows that the daycare price decrease induced on average a significant 5.9 hours (a 49 percent) increase in time spent in daycare, while the fraction of children being in daycare for at least 20 hours per week increased by 16 percentage points (by 59 percent). Similarly to the extensive margin, the policy impact on daycare hours is larger for high-educated families in absolute terms, but is not substantially different relative to the baseline mean (44 and 50 percent increase in hours for low-and high-educated families). Appendix Table A.3 also reveals that the bigger part (approximately 75 percent) of the increase in (any) daycare use comes from mothers working simultaneously, while primarily low-educated mothers increased their daycare use without working at the same time (by 4.9 percentage points on a base of 4.4 percent). Similarly, the bigger part (approximately 75 percent as well) of the decrease in no-daycare use comes from decreasing non-working.

[43] Regarding mothers' labor supply on the extensive margin, Baker et al. (2008), Kottelenberg and Lehrer (2013) and Lefebvre and Merrigan (2008) find a point estimate of 0.077, 0.110 and 0.073, respectively.

[44] Using the SLID for years 1993-2002, Lefebvre and Merrigan (2008) only find significantly positive impacts for mothers with more education than a high school diploma, with a point estimate of 0.065.

[45] Respondents who do not work any hours are included, as well. According to Appendix Table A.4, that is based on Census data, fathers' hours supplied decreased slightly, only for partners of low-educated mothers. However, Appendix Table A.6, that is based on LFS data, shows no detectable impact on fathers' labor supply on the intensive margin.

educated mothers. By taking the lower bound of the estimated extensive margin impact (0.034) and using the fact that, on average, 75% of the working women work full-time (Table A.5), this increase on the intensive margin comes entirely from changes on the extensive margin.

Parental Time Use

This section shows the most important empirical contribution of this chapter. I show that eligible parents increased total time spent with their child, at the expense of their home production and—for high-educated parents—leisure time. I show complementary parental quality time outcomes to the ones used by Kottelenberg and Lehrer (2016); I confirm that eligibility for universal daycare program in Quebec decreased parents' propensity to read their child daily, but I show that it increased reading time at the lower end of the reading distribution, by decreasing the propensity to never read to child, and increasing the propensity to read once or several times per week.

Table 2.2 reveals that parents decreased their propensity of never reading to their children by almost 3 percentage points, corresponding to a large 33 percent decrease. Parents also increased their propensity of reading once a week to their children by 2 percentage points (50 percent, relative to to the baseline), while increased reading 2-3 times a week by 3.6 percentage points (18 percent), with the former response coming entirely from high-educated families. Reading daily to the children decreased by 2.6 percentage points (3.9 percent), with no detectable differences across education groups.

Table 2.3 shows the policy impact on the parents' child and home production time allocation, measured in the Census. In response to decreasing daycare prices mothers increased their average time spent with their children by 0.87 hours (on a base of 44.8 hours); the response is significantly larger for high-educated mothers. The same pattern holds for fathers; the corresponding impact is 0.72 hours (on a base of 24 hours), which doubles for partners of high-educated mothers.[46] At the same time, mothers decreased their home production time by 2 hours per week (on a base of 30 hours), while fathers decreased their home production time 1 hour per week (on a base of 14 hours).[47] Note that the increase in child time is almost 33% larger than the difference between the average child time spent by high- and low-educated families.

Table 2.4 presents the policy impact on time use for mothers, estimated from the Time Use Diary[48]: the policy induced mothers to increase time spent with their children by 1.5 percentage points (significant at 1%, at a base of 13 percent) and decrease home production and leisure time by

[46] Appendix Table A.7, where the outcome is dichotomous, reveals that mothers were significantly less likely to spend at most 15 hours, but more likely to spend at least 30 hours with their child, by approximately 2 percentage points. Fathers show a 2.7 percentage point decrease in spending at most 15 hours per week with their child, and a 1.1 (1.9) percentage points increase in spending 15-30 (31-60) hours per week with the child (see Appendix Table A.9).

[47] Appendix Table A.8 shows that mothers increased spending at most 15 hours per week on home production by 5.3 percentage points, and decreased spending 15-30 hours (31-60 hours) on home production by 0.9 (2.8) percentage points, respectively. The decrease in spending 15-30 hours on home production comes entirely from high-educated mothers, otherwise the differences across education are not statistically significant. Fathers' home production time response can be seen in Appendix Table A.10.

[48] There is a decrease in work time for low-educated mothers (significant at 10% confidence level), while the increase in work time is not statistically different from zero, presumably due to small sample size. While this might seem to contradict the results on hours in Appendix Table 2.1, a reminder that in the GSS only, work time also includes commuting, and anything work-related; if only hours are considered in the GSS, the point estimate stays negative and cannot be statistically distinguished from zero. The results on the extensive margin (not shown for the GSS) are consistent across the datasets. These results are available on request.

0.65 and 0.9 percentage points, respectively.[49] These responses come from high-educated mothers: low-educated mothers significantly decreased work-time in favor of their leisure time, while high-educated mothers significantly decreased home production time and leisure time (by 1 and 2.2 percentage points) in favor of their child time (corresponding to a 2.28 percentage point or a 17 percent increase).[50]

Household Expenditures

The policy impact on the fraction of particular household expenditures (measured in percentage of all expenditures) can be seen in Tables 2.5 and 2.6. For each outcome variable, the first column corresponds to (2.8) including all years of 1986, 1992 and 1996-2009, the second column corresponds to (2.8) including only years when education was measured (years 1986, 1992, 1996, 2007-2009), and the third column corresponds to (2.9) in the same smaller sample. As before, the coefficients of the interaction variables show whether there is a differential policy impact by the mother's education. *Panel B* shows the weighted baseline means in the estimation sample and the estimated policy impacts, by education. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.[51]

Table 2.5 shows that, as a response to decreasing daycare prices, two-parent families increased all food expenditures significantly by 0.8-0.9 percentage points (by 7 percent, relative to the baseline mean); expenditures on food from stores increased by 0.55 percentage points (5 percent), while eating-out expenditures increased by 0.34 percentage points (14 percent). The increase is significantly larger (in absolute and in relative terms) for low-educated families. Table 2.6 reveals that for high-educated families—more likely to use regulated, more expensive daycare before—daycare expenditures decreased (by 28 percent), while for low-educated families—more likely to use informal, even free daycare arrangements before— they increased (by 26 percent). At the same time, the increase in expenditures on child games and toys (representing child market goods), and domestic help (representing home production goods) is not significantly different across education groups.

These results confirm the model's first prediction: as a response to decreasing daycare price, demand for home production goods (as eating out and domestic help) and child goods (as child games and toys) increase. These, together with results on increasing child time and decreasing home production time suggest that parents substitute their time away from activities where time is substitutable with market goods to activities where it is complementary to market goods. Higher-educated parents' larger child time (home production time) increase (decrease) suggests that their time efficiency advantage in child human capital and home production dominates their higher opportunity cost of time in non-market activities.

[49] Although the latter impacts are imprecisely estimated, primarily due to the small sample size. The child time increase comes primarily from increasing time devoted to playing with the children, putting them to bed, and helping/teaching/reprimanding; for higher-educated mothers there is an increase in travel-time to the daycare facility, while for lower-educated mothers there is an increase in reading to the children. The decrease in leisure time comes primarily from decreasing personal care, relaxation, sleep, civic and voluntary activities and sports; these results are not shown for the sake of brevity, but are available upon request.

[50] Appendix Table A.11 shows that although that fathers in high-educated families increased their child time, this happened at the expense of home production time and work time.

[51] In Appendix Figure A.3 the estimated difference between Quebec and Rest-of-Canada can be seen for households with and without children aged 0-4, for general household expenditures (food and domestic help) as a falsification test.

2.6.2 Reduced Form Results: Policy Impacts by the Propensity of Mother Working in the Absence of the Policy

The results so far suggest that on average, as a response to decreasing daycare prices, mothers simultaneously increased their labor supply, work time and child time at the expense of their home production and leisure time. However, the question remains whether this time trade-off has actually been made also by mothers who would not be working in the absence of the policy, but increase their labor supply due to the policy change. Mothers who would work full-time presumably do not adjust their labor supply, only make use of the positive income effect for market goods, affecting their intra-household time allocation. Then, since high-educated mothers are more likely to work and use public daycare in the absence of the policy, their observed larger child time increase might be solely due to composition. To investigate this, I now pursue an analysis in which I compare mothers who would be more likely to work and use daycare in the absence of the policy, with mothers who would be less likely to do so. The data structure does not allow me to observe mothers' working history, therefore I use a propensity score-based approach. First, using a probit model, for each household I combine a set of predetermined variables—such as education, age, household structure, number of children and their age—into a single index, that is the estimated propensity of the mother working in the absence of the policy. Then, I estimate policy impacts by this propensity, using model (2.10); by controlling linearly for the predetermined variables in X that were used for the score estimation, the variation left comes from higher-order interactions and the non-linearity of the probit model. Finally, I graph the policy impacts, where the horizontal axis shows the predicted propensities and the vertical axis shows the estimated policy impact; low propensities correspond to mothers unlikely to work in the absence of the policy, for whom both income and substitution effects are present.

Figures 2.3 and 2.4 graph the estimated policy impacts by predicted propensities for high- and low-educated families, respectively, with a 95% confidence band, based on estimates from the linear model (2.10)[52]. The first two panels of Figure 2.3 (lines in blue) show that for high-educated mothers the policy impact on working now is significantly decreasing with the propensity of working in the absence of the policy, accompanied by the same pattern for using any daycare. Comparing a high-educated mother with a low predicted propensity, say 0.5, with a very similar high-educated mother with high predicted propensity, say 0.85, the policy impacts are 9.15 *versus* 1.35 percentage points on working, and 14.5 *versus* 9.66 percentage points on (any) daycare use. The corresponding impacts on working time are 1.49 *versus* 0.15 hours per week.[53]

The previous results suggest that high-educated mothers unlikely to work in the absence of the policy are drawn into the labor market. Strikingly, the increase in the propensity to read to the child

[52] The underlying coefficients can be seen in Appendix Tables A.14-A.18. Appendix Figures A.4 and A.5 show it for the quadratic model. The distribution of the predicted propensity in the NLSCY, Census and GSS datasets, by education, can be seen in Figure A.6, where the 5 ticks in each graph correspond to the 10th, 25th, 50th, 75th and 90th percentile of the predicted propensity distribution, respectively.

[53] The underlying coefficients are in Appendix Tables A.14 (mother working and daycare use) and A.13 (hours). Using data from the Census and the LFS, respectively, Appendix Tables A.12 and A.13 substantiate the previous result from the NLSCY that for high-educated mothers the policy impact on working now is significantly decreasing with the propensity of working in the absence of the policy (see the linear model).

is larger for these mothers.[54] Comparing a high-educated mother with a propensity of 0.5, with the 'average' mother (with a predicted propensity of 0.72), the former is by 1.2 percentage points more, while the latter is by 1.2 percentage points less likely to read several times per week to the child. The positive policy impact on child time is decreasing (line in red), while the negative policy impacts on home production and leisure time are increasing with the estimated propensity (lines in black, although for home production time the estimates are very imprecise).[55] These results suggest that among high-educated mothers, the largest intra-household time reallocation was made by the ones drawn back or drawn into the labor market by less expensive daycare.

Regarding parenting and mental health investments, panels in the third row of Figure 2.3 reveal that the non-likely worker high-educated mothers' parenting behavior (measured by positive parenting and negative hostility) and mental health (measured by negative depression score) deteriorated to a significantly smaller extent after the policy was implemented (lines in green); for mothers with lowest propensities these outcomes even improved.[56] For instance, the policy impacts on negative hostility score for a high-educated mother with propensities of 0.4, 0.6 and 0.8 are 0.047, -0.07 and -0.19 standard deviations, respectively; the same impact on positive parenting are -0.013, -0.115 and -0.22 standard deviations. High-educated mothers drawn into the labor market even with a predicted propensity of 0.55 actually show better mental health, with a 2.7 percent of a standard deviation decrease in depression score; the corresponding policy impacts on negative depression score for a high-educated mother with a predicted propensity of 0.4, 0.6 and 0.8 are 13.5, -1 and 15.4 percent, respectively.

The previous results suggest that high-educated mothers drawn into the labor market are not only the ones who invest more in reading, but relatively more into parenting and—if one's depression score reflects one's mental health investments,—also into mental health. Consistent with these larger investments, panels in the fourth row of Figure 2.3 show that the behavioral outcomes of these high-educated mothers' children deteriorated to a significantly smaller extent after the daycare price fall, and in some cases even improved (brown lines).[57] For instance, the policy impacts on the child's negative hyperactivity score for a high-educated mother with a predicted propensity of 0.4, 0.6 and 0.8 are 0.043, -0.058 and -0.16 standard deviations, respectively; the same impacts on (negative) anxiety score are 0.031, -0.08 and 0.19 standard deviations, and on (negative) aggression score they are 0.044, -0.042 and 0.13, respectively.

It is unclear whether the same pattern—less deteriorating parental and child outcomes for more working mothers—holds for low-educated mothers, according to Figure 2.4. First, the policy impact on working now and working hours is significantly increasing with the propensity of working in the absence of the policy for them.[58] Second, both home production and leisure time is estimated to increase with predicted propensity for them (although the former is imprecisely estimated).[59] Third, only their propensity to read often (in the quadratic model) increases with propensity, while parenting

[54]The underlying coefficients can be seen in Appendix Table A.15.
[55]The underlying coefficients can be seen in Appendix Table A.16.
[56]The underlying coefficients can be seen in Appendix Table A.17.
[57]The underlying coefficients can be seen in Appendix Table A.18.
[58]The underlying coefficients are in Appendix Tables A.14 (mother working) and A.13 (mother working and hours).
[59]The underlying coefficients are in Appendix Table A.16.

and mental health outcomes do not.[60] Finally, the interaction point estimates for negative hyperactivity and negative aggression scores are positive, but imprecisely estimated, and for negative anxiety score are even negative.[61]

2.6.3 Behavioral Skill Gap Regressions

I now turn to the parental investments explaining the behavioral skill gap in early childhood. The purpose of the following descriptive 'skill gap-regressions' is to see (1) at which ages the gap significantly widens, and (2) which parental investment measures explain the gap.

The first column (the 'base' specification) in Table 2.9 shows the estimation result of model (2.14). There is a significant gap between high- and low-educated parents' children already in their first three years of life (in the 'critical period'): 6 percent of a standard deviation of the average age-standardized hyperactivity score in favor of high-status children. It significantly widens to 12.6 percent once the children reached 5 years of age, and the further changes in the gap cannot be statistically differentiated from zero; this means that the hypotheses $\alpha_2 = \alpha_3$ and $\alpha_2 = \alpha_4$ can not be not rejected.

To see through which channel parental education and child development is most related, additional columns show the estimation result for the same model by adding the corresponding variable of parenting practices, maternal health, maternal work, daycare use or parental reading. The initial gap at is essentially unaffected with the inclusion of maternal work or daycare use, suggesting that these are not important channels through which high-educated parents' children develop better than low-educated parents' children do. However, comparing two similar children receiving the same positive parenting practices at home, the disadvantaged child has worse behavioral outcomes on average by 4.8 percent of a standard deviation; thus, 15% of the baseline gap in the critical period is operating through positive parenting practices, that are not impacting the gap at later ages. The early childhood gap reduces to 3.6-3.9 percent and becomes significant only at 10 percent, once controlling for maternal mental health or family functioning, accounting for $\frac{1}{3}$ of the variation in the hyperactivity score. The early childhood gap reduces to 4.3 once controlling for parental reading practices, accounting for $\frac{1}{4}$ of the variation. Inclusion of any of the parenting, family functioning, maternal health, daycare use or labor supply measures do not alter by how much the childhood gap widens at later ages.

The previous results justify the chosen age-range and the focus on parental (quality) time in this chapter. In addition, they indicate that maternal health, parenting practices and reading to the child are the most important transmission mechanisms through which parental education and child development are related. Also, they suggest that high-educated mothers drawn (back) to the labor market invest more in parenting measures that matters most for the shape of the gap. This compensatory parenting behavior of high-educated parents–compensating more work time with more positive parenting, more bedtime reading and better mental health–is one of the mechanisms behind the widening gap.

[60] The underlying coefficients are in Appendix Tables A.15 (child time) and A.17 (parenting and mental health).

[61] The underlying coefficients are in Appendix Tables A.18. Appendix Table A.19 reveals that the increase in household expenditures is not solely, and not even primarily, due to income effects: families where the mother is drawn into the labor market are significantly more likely to increase food and daycare expenditures, than do families where the mother is already working (the coefficients' signs for food expenditures are consistent, but the estimates are noisy).

2.6.4 Structural Parameters and Policy Simulation

The reduced-form results so far are consistent with a scenario where time and market goods are complementary in child human capital production ($\rho_K < 0$), and substitutable in home production ($1 > \rho_H > 0$); they are also consistent with a scenario where high-educated parents' time is more efficient in child human capital and home production ($\delta_K > 0, \delta_H > 0$). Now I turn to the estimation of these structural parameters to see their magnitude and use them to perform counterfactual analysis.

Tables 2.7 and 2.8 show the estimation result of the models (2.11) and (2.12) for child time and home production time, respectively, using the GSS and Census datasets. The first-stage F-statistics are 46 and 71 in the case of child time, and 3 and 23 in the case of home production time for the GSS and Census datasets, respectively. Although time use is measured differently in the two datasets, the point estimates are relatively similar for child time and imply a substitution parameter for child human capital production of approximately -0.4 in the Census. This estimate suggests a modest complementarity between time and market goods in child human capital production. For home production time the point estimates are almost identical: $\hat{\rho}_H = 0.67$ in the Census, implying strong substitutability between time and market goods in home production.[62]

The estimated time efficiency parameters ($\hat{\delta}_K > 0$ and $\hat{\delta}_H > 0$) imply that higher-educated mothers' time is more productive in both child human capital and home production. Thus, a one unit time investment made by a higher-educated mother increases her child's human capital level more than a one unit time investment made by a lower-educated mother. This by itself would imply a behavioral human capital gap in the cross-section; the further difference in the levels of child human capital stems from different choices between higher- and lower-educated parents. The reduced-form estimates show that both types of parents increase daycare and parental time inputs after the daycare price falls, and if child human capital has only these two as inputs, the level of child human capital is predicted to increase for both higher- and lower-educated parents' children. Even though higher-educated parents are found to increase their inputs even more, the skill gap between 0-4 aged children might shrink or expand, depending on the concavity of the human capital production function, and the degree of time efficiency advantage of higher-educated parents.

To see the life-cycle implications of these estimated parameters, Figure 2.5 shows the gap in child human capital before and after the subsidized daycare policy change.[63] I assume the policy to take place at the age of 2.5, which age corresponds to the largest widening in the behavioral skill gap seen in Table 2.5. The first panel shows how the gap widens in the absence of the policy. For ages 0-2, the level of child human capital (K) for higher- and lower-educated parents' children in the control group are calculated with the observed average daycare hours (X_K) and hours spent with the child

[62] As a reminder, I do not estimate the parameters of $\frac{\beta}{\theta}$, ρ_X or $c = \frac{\tau^F w^F}{w-m}$, but assume the value of $\frac{0.3}{1}$, -1 and 0.6, when estimating the models (2.11) and (2.12) by 2SLS. Appendix Tables A.25 and A.26 show the robustness check for different values of ρ_X and β. The implied substitution parameter ρ_K decreases with β_K and ρ_X, but it is always negative, and is at most -0.23. Similarly, that the implied time efficiency parameter δ_K decreases with β_K and ρ_X, but it is always positive, and at least 0.18. The implied substitution parameter ρ_H increases with β_H up to $\beta_H = 0.25$ and then stays around 0.66. The implied time efficiency parameter δ_H decreases with β_H up to $\beta_H = 0.25$ and then stays around 0.32.

The alternative additive specification for the good-specific error, mentioned in footnote 14, yields an estimate of around -2 for ρ_K and 25 for δ_K, and 0.98 for ρ_H and 13 for δ_H (slightly depending on starting values in the optimization).

[63] The summary table of the policy simulation results can be found in Appendix Table A.29.

(T_K) as inputs observed in the NLSCY and Census datasets, in the CES production function (2.2), using the estimated parameters $\hat{\rho}_K = -0.4$ and $\hat{\delta}_K = 0.36$.[64] The prior level of K for ages 0-2 is 3.11 for lower- (L), and 4.23 for higher-educated (H) parents' children ($K^L_{012} = 3.11$, $K^H_{012} = 4.23$), for whom no daycare policy is in effect. Thus, the prior gap for ages 0-2 (gap_{012}) is 1.12. The second line shows how the gap would have evolved in the absence of the policy change, using the observed average daycare hours and hours spent with the child as inputs for ages 3-4 in the control group with no policy in effect. The no-policy level of K ($\tilde{K}$) for ages 3-4 is 3.8 for lower-, and 5.01 for higher-educated parents' children ($\tilde{K}^L_{34} = 3.8$, $\tilde{K}^H_{34} = 5.01$); thus the no-policy gap for ages 3-4 ($\tilde{gap}_{34}$) is 1.21. Using only daycare and parental time as inputs and abstracting from self-productivity in child skill formation, the gap from ages 0-2 to 3-4 widens by 8 percent in the absence of the policy.

At age 2.5 the subsidized daycare policy takes place: for ages 3-4, the level of K has been recalculated with the observed change in daycare hours ($\triangle X_K$) and the observed change in hours spent with the child ($\triangle T_K$) as inputs, just for ages 3-4 in the CES production function (2.2), keeping $\hat{\rho}_K = -0.4$ and $\hat{\delta}_K = 0.36$.[65] As shown in the second panel, the actual level of K for ages 3-4 is 4.33 for lower-, and 5.88 for higher-educated parents' children ($K^L_{34} = 4.33$, $K^H_{34} = 5.88$), resulting in the actual skill gap of 1.55 (gap_{34}). After the policy change, from ages 0-2 to 3-4, the level of K actually increases by 39 percent for both higher- and lower-educated parents' children; the gap from ages 0-2 to 3-4 widens also by 39 percent as a result of the policy.[66] Compared to the no-policy level of child human capital $\tilde{K}^H_{34}$, $\tilde{K}^L_{34}$ for ages 3-4, the actual level of K is by 17 and 14 percent higher for higher- and lower-educated parents' children as a result of the policy; the actual skill gap of 1.55 is by 29 percent larger than the no-policy gap of 1.21.[67] If the policy had not been universal but targeted only at low-educated families, 56 percent of the actual skill gap could be eliminated.[68]

To quantify the importance of the time efficiency advantage of higher-educated parents, the third panel shows the counterfactual level of K^H_{34} if higher-educated parents had no time efficiency advantage—i.e. if $\delta = 0$ or equivalently in this model, $S = 1$—, that is only 5.73 ($\check{K}^{H1}_{34} = 5.32$). As a reminder, the difference between the actual and no-policy level of K^H_{34} is 0.87; this suggests that approximately 36 percent of this difference stems from higher-educated parents' increase in inputs, while 64 percent is due to their time efficiency advantage.[69] If higher-educated parents had no time efficiency advantage, the actual skill gap would shrink by 36 percent, from the level of 1.55 to 0.99.[70] Finally, the fourth panel suggests that the difference between the actual and no-policy level of K^L_{34} for lower-educated parents' children would be by 80 percent higher if their parents had no time efficiency disadvantage, in the counterfactual case of having more schooling with $S = 2$.[71] If lower-educated

[64]The underlying conditional means for daycare hours and parental time, for ages 0-2 and 3-4, for children in the control group with no policy in effect, can be seen in Appendix Table A.27.

[65]The underlying estimates can be seen in Appendix Table A.28.

[66]This is by comparing K^e_{012} with $K^e_{34.}$ for $e = L, H$, and gap_{012} with gap_{34}.

[67]This is by comparing $\tilde{K}^e_{34.}$ with $K^e_{34.}$ for $e = L, H$, and $\tilde{gap}_{34}$ with gap_{34}.

[68]This is because removing the policy would decrease the skill level of high-educated parents' children, by moving them from the solid black line (from K^H_{34}=5.88) to the dashed black line (to $\tilde{K}^H_{34}$=5.01).

[69]The difference between the actual $K^H_{34} = 5.88$ and no-policy $\tilde{K}^H_{34} = 5.01$ is 0.87; the difference between the counterfactual $\check{K}^{H1}_{34} = 5.32$ and no-policy $\tilde{K}^H_{34} = 5.01$ is 0.31. Thus, $\frac{0.31}{0.87} \sim 36\%$ of the (actual-no policy) difference is stemming from increased inputs and the rest is due to time efficiency advantage.

[70]This is by $1 - \frac{1.55-0.99}{1.55} \sim 36\%$.

[71]The difference between the actual $K^L_{34} = 4.33$ and no-policy $\tilde{K}^L_{34} = 3.80$ is 0.53; the difference between the counter-

parents had no time efficiency disadvantage, the actual skill gap would shrink by 27 percent.[72]

2.7 Discussion of Alternative Models

As shown in Section 2, time efficiency differences in non-market activities across education groups can explain both parents' cross-sectional choices, and their responses to a daycare price change. As shown in Section 6, the reduced-form predictions of the model with time efficiency differences across education groups are supported by empirical evidence (I further refer to this as 'baseline model'). The question arises, however, what alternative models would arrive at the same predictions on parents' cross-sectional choices; especially, on higher-educated parents spending more time with their children. Therefore, instead of differences between education groups in time-efficiency, I explore alternative sources of heterogeneity. In particular, I allow for (1) differences in preferences for child human capital, and (2) differences in the substitution parameter between time and market good in child human capital production, both across education groups. In addition, I show that (3) a variant of the original model including a large daycare quality decrease concurrent with the daycare price decrease gives predictions inconsistent with the data. These alternative models can be rejected based on *responses* to a price change; they thus help rule out competing explanations to the cross-sectional observation on *choices*.[73]

Regarding (1), suppose that instead of the time efficiency γ, heterogeneity is introduced into the utility parameter β_K, that is assumed to depend positively on schooling S and a preference error/shock $\varepsilon_{K\beta}$: $\beta_K = S_i^{\delta_{K\beta}} \varepsilon_{K\beta i}$. This model allows for the possibility that higher-educated parents value child human capital more. To reject this model, I make use of the functional form assumption of Cobb-Douglas preferences, in the following way. First, since the base child time response depends positively on β_K and negatively on $w - m$, this (1) and the baseline model give the same prediction for how the response of T_K to decreasing m depends on schooling. However, their prediction differs regarding how the response of $\log T_K$ depends on schooling: since $\frac{\partial \log T_{Ki}}{\partial(-m)}$ is independent of β_{Ki}, it depends on schooling only through the wage, unambiguously negatively. This prediction can be tested by running a DiD model on $\log T_{Ki}$, time fixed effects and province fixed effects. The main insight here is that if the policy had no impact on either schooling or parental preferences either in Quebec or in the rest of Canada, then the preference channel can be differenced out, and the coefficient on the policy variable will give exactly the DiD estimate for $\log T_K$. However, this estimate is constant–and does not decrease by schooling–, thus alternative model (1) is not supported by the data.

At this point I would like to re-iterate that the rejection of (1) is based on the functional form assumption of Cobb-Douglas utility function (2.1), and thus the log-separability of the utility parameter β_j in the Marshallian demands (2.5 and 2.6). Note that the Marshallian demands in this problem are almost identical to the the standard Marshallian demands in case of Cobb-Douglas utility, except that in this problem there is no direct price of the goods child human capital K, home production good H and leisure good L, but an equilibrium price index is formed from the underlying input prices ($w - m$

factual $\check{K}_{34}^{L2} = 4.75$ and no-policy $\tilde{K}_{34}^{L} = 3.80$ is 0.95. Thus, the (actual-no policy) difference would be by $\frac{0.95}{0.53} - 1 \sim 80\%$ larger if low-educated parents had no time efficiency disadvantage.

[72] This is by $1 - \frac{1.55-1.13}{1.55} \sim 27\%$.

[73] The detailed explanations and derivations can be found in Appendix 4.

on the time input T_K and 1 one the time market good X_K). Other than this difference, the demands are linear in the utility parameter β_j and in income, and the demand for the j-th good only depends on the input prices for the j-th good. If, for instance, instead of a Cobb-Douglas utility function, a Stone-Geary utility function was used, the Marshallian demands (i) would not be log-additive in the utility parameter anymore, but additive in the "committed" amount on each good, and (ii) would depend on all prices. Thus, they would considerably lose from their tractability and simplicity, and the alternative model (1) would no longer be rejectable based on the aforementioned argumentation. Consequently, my results may depend on the chosen functional form of Cobb-Douglas utility.

Suppose that heterogeneity by education is introduced into the substitution parameter ρ, and ρ_j-s are assumed to depend on years of schooling S (negatively) and a good-specific error $\varepsilon_{K\rho}$: $\rho_K = \Upsilon_\rho(S, \varepsilon_{K\rho})$. This model allows for the possibility that the time of higher-educated parents is less substitutable with daycare time, or, the substitution of time and market goods in producing child human capital is perceived to be lower for higher-educated parents—for instance, if they think that market-purchased daycare is a poor substitute for their own quality time in child human capital production. As I show in Appendix A.4, both channels predict that higher-educated parents increase child goods, and also daycare to a smaller extent, contradicting the estimates presented in Section 2.6.

Regarding (3), suppose that the CES function of child market good X_K takes the following form: $X_K = \left[B_K^{\rho_X} + (QD)^{\rho_X}\right]^{\frac{1}{\rho_X}}$, where ρ_X is the substitution parameter between daycare time D and other child goods B, while Q denotes daycare quality. Denote the fraction of other child goods α, then, in optimum, the optimal ratio of B relative to D depends on the relative prices and daycare quality, according to: $\frac{1-\alpha}{\alpha} = m^{\frac{1}{\rho_X - 1}} Q^{\frac{\rho_X}{1-\rho_X}}$. As I show in Appendix A.4, if daycare quality dramatically decreases parallel to a daycare price decrease and if higher-educated parents care about daycare quality more, then they are predicted to increase daycare use considerably less than low-educated parents increase daycare use. Again, this is not consistent with estimates presented in Section 2.6.

2.8 Income and Substitution Effects: The German Experiment

Using the policy change in Quebec, the total effect of a daycare price change on parents' time and resource allocation can be measured. To disentangle income and substitution effects, I make use of the daycare policy change in 2006 in Thuringia; this latter policy environment enables me to estimate the impact of a compensated daycare price change and a pure income shock on parents' time allocation.

Between 1st July 2006 and 2010, the government of Thuringia had a daycare policy in effect, corresponding to a compensated daycare price increase. Under this system parents of 2-year-old children received a subsidy of EUR150 if their child does not attend a publicly subsidized daycare. If the eligible child attended full-time public daycare, the full amount went the daycare provider; if part-time, the parents and the provider shared the amount proportionally.

The policy impacted families differently depending on whether they would use public daycare in the absence of the policy. Those who would, the policy change represents a fully compensated daycare price increase: public daycare becomes more expensive relative to other daycare modes, but it is fully compensated by the government. Parents with an eligible 2-year-old child in full-time daycare would pay an additional EUR150 per month to the facility, relative to the the baseline of approximately

EUR80. Parents who would stop to use public daycare could use the windfall income to pay for other private childcare arrangements or staying at home. Parents who would not use public daycare in the absence of the policy are expected to respond differently, depending on whether they would not want to or would not get a spot. If they would not want to, they are predicted to continue not to use public daycare, due to a negative substitution effect and a pure income shock. If they would not get a spot in the absence of the policy, they might use the opportunity of new spots becoming available.

To shed light on income and substitution effects, I measure the policy impact of compensated price increase by the predicted propensity of using public daycare in the absence of the policy. I complement the analysis of Gathmann and Sass (2012), who measure the average policy impact with a DiD strategy using data on East-German bundeslands between 2000 and 2009. They find that parents decreased public daycare use and informal daycare use, and mothers decreased their labor supply in response to the policy change, but have not analyzed parental time use in depth.

The preceding policy was means-tested, conditional on income and independent of daycare choices: parents received a monthly subsidy of EUR300 if at least one parent worked less than 30 hours per week, and the monthly household income was below a threshold of EUR1,375 for two-parent families and EUR1,125 for single parents. Gathmann and Sass (2012) provide evidence that the new policy had little effect on the supply side, in terms of daycare places supplied, opening hours, attendance rates of non-eligible one- and three-year old children, quality or daycare fees; and there were no further changes in the legislation or regulation of publicly subsidized daycare facilities in Thuringia. However, a major parental leave reform took place in 2007, and although parental leave legislation is federal in Germany, four bundeslands (Thuringia, Saxony, Baden-Wurttemberg, Bavaria) have complementary policies.[74] Thus, to mitigate concerns about contemporaneous bundesland-specific changes in daycare/parental leave legislation, I exclude Lower-Saxony, Bavaria and Baden-Wurttemberg.

For this analysis, the waves between 2000 and 2010 in the German Socio-Economic Panel (GSOEP) are used.[75] The GSOEP is the German household panel study similar to the Panel Study of Income Dynamics in the United States and the British Household Panel Study; it is a representative longitudinal study of private households in Germany since 1984, initially containing only West-German households and expanding to include the states of the former German Democratic Republic (GDR) in June 1990. In this dataset all members of the first-wave survey households and all their offsprings are followed, and all household members 17 years and older are interviewed.

Using data on families with at least one 2-year-old child between 2000 and 2010, and excluding data on Lower-Saxony, Bavaria and Baden-Wurttemberg, I estimate the following model:

$$\begin{aligned} Y_i \quad &= \delta_0 + \delta_1 \text{policy}_{tp} + \delta_2 \text{policy}_{tp} \times \text{propensity-daycare}_i + \delta_3 C_i + \delta_4 U_i \\ &\quad + \delta'_{5t} + \delta'_{6p} + \delta'_7 X_i + \delta_8 \text{post}_t \times \text{propensity-daycare}_i \\ &\quad + \delta_9 \text{Thur}_p \times \text{propensity-daycare}_i + \delta_{10} \text{propensity-daycare}_i + \varsigma_i, \end{aligned} \tag{2.15}$$

where i indexes household, t indexes time (year), p indexes bundesland (German state), *propensity-daycare* denotes the propensity score of using public daycare in the absence of the policy, *post* equals

[74] For instance, they paid in addition a means tested parental benefit extended to the 3rd year of parental leave, or in Thuringia before 2006, childcare allowance extended a federal policy for all parents with children under the age of two.

[75] More information on the GSOEP can be found in Wagner et al. (2007).

1 if the family is observed after 2006 July, *Thur* indicates the family residing in Thuringia, and X includes the child's gender, parents' age, household structure and household size. A significantly negative estimate of δ_2 indicates that policy impact is larger for families less likely to use public daycare in the absence of the policy. The propensity score is predicted by the interaction of parental education of the mother and the father, a full set of province and child age fixed effects, a linear time trend, mother's age, father's age and household size. Similarly to the propensity of the mother working in the absence of the policy used in section 2.6.2, *propensity-daycare* is obtained by estimating a probit model on the pre-policy sample and predicting propensities for the whole sample.

According to Table 2.10, the impacts of the Thuringian Daycare Policy differ substantially whether the family would be more or less likely to use public daycare in the absence of the policy. For instance, a very unlikely daycare-user family (with a predicted propensity of 0.043 that corresponds to the 10th percentile in the distribution of the predicted propensity) is 22 percentage points more likely to use public daycare after the policy is implemented. On the other hand, a very likely daycare user family (with a predicted propensity of 0.82, corresponding to the 95th percentile), for whom only the substitution effect is present, is 4 percentage points less likely to use public daycare after the compensated price increase. The same impacts for the mother's labor supply on the extensive margin are 46 and -1.5 percentage points, and on hours worked per week are 12 and -2.8. These results suggest that families respond to compensated daycare price increases as predicted by the theory. In addition, they suggest that families less likely to use public daycare in the absence of the policy are now drawn into it, with the mother more likely to work, and supplying more market hours—presumably these families take advantage of the outflow of children from public daycare in a rationed daycare market.[76] At the same time, these mothers increase their child time at the expense of their home production and leisure time—the same substitution pattern seen in the Canadian experiment. The increase in child time is 1 hour on a usual weekday (imprecisely estimated), and the significant decreases in time spent on housework, errands and hobbies are -1.4, -0.29 and -0.57 hours on a usual weekday, respectively.

2.9 Conclusion

This chapter contributes to the theoretical literature on the "parental time-education gradient" puzzle by extending the classic Beckerian framework with time efficiency differences across education groups in child skill formation. It presents a new set of results on parental time and household expenditure allocation using exogenous daycare price variation. It also presents estimates of the model's structural parameters on substitutability between time market goods for child human capital and home production, and on the between-education-group heterogeneity in non-market time efficiency.

This chapter provides the following explanation to the puzzle documented by Guryan et al. (2008): keeping efficiency of time investments into children's human capital constant, higher-wage parents are more able to afford child market goods (such as daycare); if there is complementarity between time and market goods in child human capital production, they will also spend more time with their child, despite their time efficiency advantage. This framework also predicts that higher-educated parents increase their child time more after daycare prices fall, if their time efficiency advantage in non-market

[76]Evidence on whether there have been daycare supply constraints and/or rationing in Thuringia is mixed; according to Wrohlich (2006) there have been, contradicting evidence presented by Gathmann and Sass (2012).

activities is large enough to outweigh their larger opportunity cost of time in non-market activities.

Using the Quebec daycare policy change as a negative exogenous shock to the price of daycare, I show substantial labor supply responses, driven primarily by higher-educated mothers not working in the absence of the policy. When higher- and lower-educated mothers increase their work-time, they tend to increase their time spent with their children (including reading more to the children), at the expense of their home production and leisure time. Consistent with the model's predictions, families increase household expenditures that proxy for home production goods. My findings underline the pivotal role of substitutability between time and market goods, implying complementarity in child human capital production and substitutability in home production. Higher-educated parents' larger time reallocation suggests their time efficiency advantage in child human capital and home production.

I document and formally test the widening noncognitive gap between high-status and low-status children, between ages 0 and 11. I find the largest widening after the first three years of life, often labeled as the 'critical period'. I also find that parental reading to the child, together with maternal health, family functioning and positive parenting, are more important transmission mechanisms between family background and behavioral scores in early childhood, than are daycare attendance or maternal work. I argue that a compensatory parenting behavior of high-status parents—compensating more work time with more parental time and possibly other parental inputs—is one of the mechanisms behind the widening gap. As a suggestive evidence of this, I show that high-educated Canadian mothers drawn into the labor market invest more in reading, parenting and mental health; consistently with these larger investments, the behavioral outcomes (anxiety, hyperactivity, aggression) of these mothers' children deteriorated to a significantly smaller extent, and in some cases even improved, after the introduction of the Quebec Daycare Policy.

2.10 Tables

General Notes for Tables 2.1-2.6: in each table, *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9), on the Intent-to-Treat effect of subsidized daycare provided by the Quebec "5\$/day" policy (1997). The policy variable is an interaction between residing in Quebec and being eligible to the policy by cohort; the high-educated dummy refers to the mother having at least some post-secondary studies; province and year fixed effects and standard set of X-s are controlled for (as the mother's education, gender and age of the child, age structure of the parents, number and age structure of siblings, number of other household members). In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample in the control group (policy=0), while the second column shows the estimated policy impact, for all and by mother's education. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$. Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.1: Effect of a Daycare Price Decrease on Daycare Use (Extensive Margin) and Maternal Labor Supply (Both Margins); Policy Impact for All and by Education

	1: in institutional care		**1: in any care**		**1: mother working**		**mother hours**	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	0.184***	0.130***	0.131***	0.088***	0.076***	0.042**	1.081***	0.414*
	(0.007)	(0.015)	(0.009)	(0.019)	(0.008)	(0.016)	(0.172)	(0.221)
β_2 : policy		0.070***		0.061***		0.048**		0.756**
×high-educ.		(0.014)		(0.020)		(0.018)		(0.361)
β_3 : college	0.048***	0.035***	0.153***	0.166***	0.174***	0.176***	5.160***	4.857***
	(0.014)	(0.012)	(0.011)	(0.017)	(0.011)	(0.015)	(0.396)	(0.357)
β_4 : university	0.076***	0.062***	0.195***	0.209***	0.191***	0.194***	6.506***	6.208***
	(0.011)	(0.010)	(0.008)	(0.013)	(0.017)	(0.017)	(0.606)	(0.471)
R^2	0.123	0.124	0.117	0.117	0.094	0.095	0.074	0.074
N	61,962	61,962	61,962	61,962	61,496	61,496	186,941	186,941
data	NLSCY	NLSCY	NLSCY	NLSCY	NLSCY	NLSCY	LFS	LFS
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	0.092	0.184	0.437	0.131	0.614	0.076	19.296	1.081
low-educ.	0.053	0.131	0.312	0.088	0.481	0.042	15.097	0.414
high-educ.	0.106	0.201	0.484	0.141	0.663	0.09	21.376	1.170
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.000		0.000		0.000		0.000

Source of data: NLSCY waves 1-7, LFS 1994-2006, 0-4 years old children in two-parent families, both parents at most 50 years old.

Table 2.2: Effect of a Daycare Price Decrease on Reading to the Child; Policy Impact for All and by Education

	1: never reading		**1: reading weekly**		**1: reading often**		**1: reading daily**	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	-0.032***	-0.029**	0.0194***	0.003	0.036***	0.042***	-0.026***	-0.022
	(0.0037)	(0.0117)	(0.0017)	(0.0063)	(0.0032)	(0.0124)	(0.0035)	(0.0261)
β_2 : policy		-0.003		0.022***		-0.011		-0.004
×high-educ.		(0.0142)		(0.0076)		(0.0200)		(0.0366)
β_3 : college	-0.047***	-0.042***	-0.019***	-0.019***	-0.032**	-0.046**	0.099***	0.108***
	(0.005)	(0.010)	(0.004)	(0.005)	(0.012)	(0.018)	(0.015)	(0.030)
β_4 : university	-0.070***	-0.066***	-0.031***	-0.032***	-0.072***	-0.086***	0.174***	0.184***
	(0.013)	(0.014)	(0.003)	(0.006)	(0.016)	(0.022)	(0.021)	(0.037)
R^2	0.118	0.119	0.015	0.015	0.035	0.035	0.139	0.139
N	60,858	60,858	60,858	60,858	60,858	60,858	60,858	60,858
dataset	NLSCY	NLSCY	NLSCY	NLSCY	NLSCY	NLSCY	NLSCY	NLSCY
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	0.088	-0.032	0.037	0.019	0.202	0.036	0.672	-0.026
low-educ.	0.129	-0.029	0.055	0.003	0.248	0.042	0.567	-0.022
high-educ.	0.073	-0.032	0.030	0.025	0.184	0.031	0.712	-0.026
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.000		0.000		0.002		0.0344

Source of data: NLSCY waves 1-7, 0-4 years old children in two-parent families, both parents at most 50 years old.

Table 2.3: Effect of a Daycare Price Decrease on Mother's and Father's Child Time and Home Production Time Use; Policy Impact for All and by Education

	mother child time		**mother home production time**		**father child time**		**father home production time**	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	0.870***	0.591**	-2.046***	-2.233***	0.718***	0.442*	-1.084***	-1.215***
	(0.1052)	(0.2385)	(0.130)	(0.271)	(0.135)	(0.220)	(0.059)	(0.188)
β_2 : policy		0.677**		0.528		0.484*		0.278
×high-educ.		(0.2766)		(0.334)		(0.241)		(0.260)
β_3 : college	1.175***	1.809***	-1.232***	-0.793***	1.505***	1.714***	0.097	0.106
	(0.374)	(0.143)	(0.345)	(0.262)	(0.223)	(0.136)	(0.201)	(0.115)
β_4 : university	-0.313	0.315*	-4.272***	-3.848***	1.741***	1.947***	-0.451**	-0.456***
	(0.293)	(0.164)	(0.575)	(0.347)	(0.108)	(0.248)	(0.199)	(0.119)
R^2	0.067	0.067	0.023	0.023	0.028	0.028	0.016	0.016
N	698,490	698,490	698,490	698,490	698,490	698,490	698,490	698,490
data	Census	Census	Census	Census	Census	Census	Census	Census
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	44.824	0.870	30.523	-2.046	24.104	0.718	14.115	-1.084
low-educ.	44.409	0.591	32.403	-2.233	23.203	0.442	14.169	-1.215
high-educ.	45.020	1.268	29.637	-1.805	24.530	0.927	14.089	-0.935
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.000		0.000		0.000		0.000

Source of data: Census (1996,2001,2006), two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

Table 2.4: Effect of a Daycare Price Decrease on Mother's Time Use; Policy Impact for All and by Education

	work time		**child time**		**home production time**		**leisure time**	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	0.059	-2.480*	1.493***	-1.008	-0.649	0.339	-0.903	3.150**
	(0.8497)	(1.2718)	(0.4010)	(1.1648)	(0.4227)	(0.9489)	(0.6593)	(1.1468)
β_2 : policy		3.361*		3.289*		-1.292		-5.358***
×high-educ.		(1.8350)		(1.6850)		(1.1086)		(1.6155)
β_3 : college	2.966***	2.711*	-0.707	-0.318	-1.872***	-2.204**	-0.387	-0.189
	(0.527)	(1.461)	(0.707)	(1.625)	(0.485)	(0.856)	(0.830)	(0.801)
β_4 : university	4.751***	4.523***	-0.505	-0.035	-3.407***	-3.787***	-0.839	-0.701
	(0.743)	(1.502)	(1.126)	(1.658)	(0.427)	(0.957)	(1.493)	(1.008)
R^2	0.083	0.082	0.163	0.164	0.033	0.032	0.024	0.025
N	2,001	2,001	2,001	2,001	2,001	2,001	2,001	2,001
dataset	GSS	GSS	GSS	GSS	GSS	GSS	GSS	GSS
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	10.993	0.059	13.122	1.493	16.792	-0.649	59.093	-0.903
low-educ.	8.485	-2.480	13.121	-1.008	18.180	0.339	60.215	3.150
high-educ.	12.050	0.881	13.123	2.281	16.207	-0.953	58.620	-2.208
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.465		0.007		0.056		0.021

Source of data: GSS (1998,2005,2010), two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

Table 2.5: Effect of a Daycare Price Decrease on Food Expenditures (%); Policy Impact for All and by Education

	food(%)			**food - store(%)**			**food - restaurant(%)**		
Panel A: Regression Results									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β_1 : policy	0.817***	0.939***	2.108***	0.436***	0.558***	1.403***	0.322***	0.343***	0.623***
	(0.114)	(0.144)	(0.234)	(0.117)	(0.127)	(0.182)	(0.057)	(0.098)	(0.148)
β_2 : policy			-1.097***			-0.637**			-0.407***
×high-educ.			(0.282)			(0.256)			(0.097)
β_3 : college	-0.922**	-1.001***	-0.804***	-1.020***	-1.098***	-0.950***	0.103	0.100	0.145*
	(0.319)	(0.301)	(0.207)	(0.279)	(0.260)	(0.171)	(0.069)	(0.066)	(0.074)
β_4 : university	-1.562***	-1.689***	-1.508***	-1.666***	-1.782***	-1.658***	0.101*	0.088	0.142*
	(0.291)	(0.262)	(0.198)	(0.323)	(0.297)	(0.148)	(0.050)	(0.054)	(0.069)
R^2	0.13	0.17	0.17	0.13	0.17	0.18	0.04	0.04	0.04
N	22,725	7,228	7,228	22,725	7,228	7,228	22,725	7,228	7,228
Panel B: Means and Policy Impacts									
		mean	*impact*		*mean*	*impact*		*mean*	*impact*
all		13.095	0.939		10.603	0.558		2.457	0.343
low-educ.		14.845	2.108		12.269	1.403		2.539	0.623
high-educ.		12.047	1.011		9.608	0.766		2.406	0.216
Panel C: P-values of Testing Coefficients									
$\beta_1 + \beta_2 = 0$			0.000			0.001			0.021

Source of data: SHS (1986,1992,1996-2009), two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

Table 2.6: Effect of a Daycare Price Decrease on Child Good and Home Production Good Expenditures (%); Policy Impact for All and by Education

	daycare(%)			**games-toys(%)**			**domestic help(%)**		
Panel A: Regression Results									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
β_1 : policy	-0.125**	-0.374***	0.428***	0.089***	0.155***	0.086	0.023**	0.094***	0.049**
	(0.047)	(0.073)	(0.136)	(0.018)	(0.017)	(0.068)	(0.008)	(0.006)	(0.018)
β_2 : policy			-1.093***			0.134			0.032
×high-educ.			(0.179)			(0.090)			(0.026)
β_3 : college	0.116	0.163*	0.109	0.028	0.015	0.143*	0.002	0.006	-0.002
	(0.088)	(0.087)	(0.134)	(0.048)	(0.049)	(0.068)	(0.019)	(0.018)	(0.024)
β_4 : university	0.512***	0.603***	0.542**	0.038	0.017	0.159**	0.161***	0.172***	0.165***
	(0.121)	(0.127)	(0.181)	(0.057)	(0.054)	(0.062)	(0.029)	(0.031)	(0.019)
R^2	0.04	0.05	0.05	0.05	0.04	0.05	0.03	0.06	0.06
N	22,725	7,228	7,228	22,725	7,228	7,228	20,578	5,081	5,081
Panel B: Means and Policy Impacts									
		mean	*impact*		*mean*	*impact*		*mean*	*impact*
all		2.088	-0,374		0.851	0.155		0.119	0.094
low-educ.		1.627	0.428		0.866	0.086		0.0402	0.049
high-educ.		2.365	-0.665		0.842	0.220		0.158	0.081
Panel C: P-values of Testing Coefficients									
$\beta_1 + \beta_2 = 0$			0.000			0.001			0.000

Source of data: SHS (1986,1992,1996-2009), two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

Table 2.7: Estimation of Structural Parameters - Child Human Capital Production

	GSS Time Use Diary		Census	
	OLS	**2SLS**	**OLS**	**2SLS**
$\log(w_i - m_i)(1 + m_i^{\frac{1}{\rho_X - 1}})$	-0.1027	-0.3598	0.0652***	-0.2826***
	(0.0786)	(0.2464)	(0.0113)	(0.0464)
$\log S_i$	0.1330	0.2343	-0.1728***	0.1045***
	(0.1353)	(0.1678)	(0.0495)	(0.0108)
R^2	0.1819	0.1855	0.0225	0.0502
N	1,809	1,809	680,420	696,290
first-stage F-statistics		45.77		71.18
implied ρ		-0.5621		-0.3939***
		(0.601)		(0.0902)
implied δ		0.6510*		0.3697***
		(0.3486)		(0.0395)

Note: this table shows the estimation result of (2.11), with the corresponding first-stage equation of (2.13), to recover the structural parameters ρ_K (the substitution parameter between time T and market good X in child human capital production) and δ_K (that shows how time efficiency in child human capital production γ_K depends on schooling). The excluded instrument is the policy variable (an interaction between residing in Quebec and being eligible to the policy by cohort). The standard errors on the implied parameters ρ and δ are calculated by the delta method. Standard errors are in parentheses and are clustered at the (province×post)-level. The coefficient on $\log S_i$ is $\frac{\rho_K}{\rho_K - 1}\delta_K$, while the coefficient on the function of $w_i - m_i$ is $-\frac{\rho_K}{\rho_K - 1}$. *** p<0.01, ** p<0.05, * p<0.1. Source of data: GSS (1998,2005,2010), and Census (1996,2001,2006), mothers in two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

Table 2.8: Estimation of Structural Parameters - Home Production

	GSS Time Use Diary		Census	
	OLS	**2SLS**	**OLS**	**2SLS**
$\log(w_i - m_i)$	0.4240***	2.2938	0.2271***	2.0580***
	(0.1198)	(2.0508)	(0.0274)	(0.4532)
$\log S_i$	0.0239	-0.8242	-0.1679***	-0.7184**
	(0.1005)	(0.9234)	(0.0518)	(0.2027)
R^2	0.0473	0.0501	0.0866	0.608
N	1,866	1,866	668,070	668,070
first-stage F-statistics		3.1258		23.31
implied ρ		0.6964***		0.6730***
		(0.1890)		(0.0485)
implied δ		0.3593***		0.3491***
		(0.0889)		(0.0310)

Note: this table shows the estimation result of (2.12), to recover the structural parameters ρ_H (the substitution parameter between time T and market good X in home production) and δ_H (that shows how time efficiency in home production γ_H depends on schooling). The excluded instrument is the policy variable (an interaction between residing in Quebec and being eligible to the policy by cohort). The coefficient on $\log S_i$ is $\frac{\rho_H}{\rho_H - 1}\delta_H$, while the coefficient on the function of $w_i - m_i$ is $-\frac{\rho_H}{\rho_H - 1}$. Standard errors are in parentheses and are clustered at the (province×post)-level. The standard errors on the implied parameters ρ and δ are calculated by the delta method. *** p<0.01, ** p<0.05, * p<0.1. Source of data: GSS (1998,2005,2010), and Census (1996,2001,2006), mothers in two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

Table 2.9: Documenting the Behavioral Skill Gap between High-status and Low-status Children over Ages 0-11

controlling for:		*parenting practices*		*maternal health*		*maternal work and care*		
	base	**positive**	**family func.**	**depression**	**1:excellent**	**1:work**	**1:in care**	**1:reading**
τ_1 : high-educ.	-0.0559***	-0.0477**	-0.0390*	-0.0364*	-0.0395**	-0.0553**	-0.0558***	-0.0429***
	(0.0186)	(0.0185)	(0.0189)	(0.0185)	(0.0182)	(0.0203)	(0.0160)	(0.0148)
τ_2 : high-educ.	-0.1261***	-0.1290***	-0.1302***	-0.1274***	-0.1306***	-0.1297***	-0.1257***	-0.1297***
×age(3-5)	(0.0225)	(0.0234)	(0.0223)	(0.0233)	(0.0207)	(0.0202)	(0.0229)	(0.0238)
τ_3 : high-educ.	-0.1049**	-0.1145***	-0.1202***	-0.1232***	-0.1030**	-0.1059**		
×age(6-8)	(0.0399)	(0.0392)	(0.0390)	(0.0380)	(0.0394)	(0.0399)		
τ_4 : high-educ.	-0.1023**	-0.1119**	-0.1176**	-0.1204**	-0.1045**	-0.1027**		
×age(9-11)	(0.0479)	(0.0467)	(0.0472)	(0.0462)	(0.0481)	(0.0486)		
R^2	0.035	0.039	0.038	0.043	0.041	0.035	0.022	0.024
N	104,370	104,370	104,370	104,370	103,232	103,449	71,749	71,749

Note: this table shows the result of estimating the base model (2.14) on the average age-standardized behavioral score; controlling for a full set of year (or wave), province, child age dummies, and standard set of X-s, such as the gender of the child, number of older and younger siblings (capped at 3 and 2, respectively), the size of the household, the mother's and father's age, and the father's education. high-educ. indicates high-educated, defined as the mother having either some college, or university education, age$(3-5)$ indicates the child being between 3 and 5 years old. The coefficients of interest are $\tau_1 - \tau_4$, corresponding to the behavioral/developmental human capital gap—also called the noncognitive skill gap—across age-categories. τ_1 shows the gap for children between 0 and 2 years old, while $\tau_1 + \tau_2$ shows the gap for children between 3 and 5 years old; so, for instance, τ_2 shows whether the gap significantly widens after the first 3 years of life. In the first column I present the base estimation results of (2.14), and in further columns I assess how the inclusion of parental practices, maternal health, maternal employment, daycare attendance and parents' reading practices (highlighted in bold in the headlines of each column) change the estimated coefficients $\hat{\tau}_1 - \hat{\tau}_4$. Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: NLSCY waves 1-7 (1994-2006), 0-11 years old children in two-parent families, both parents at most 50 years old.

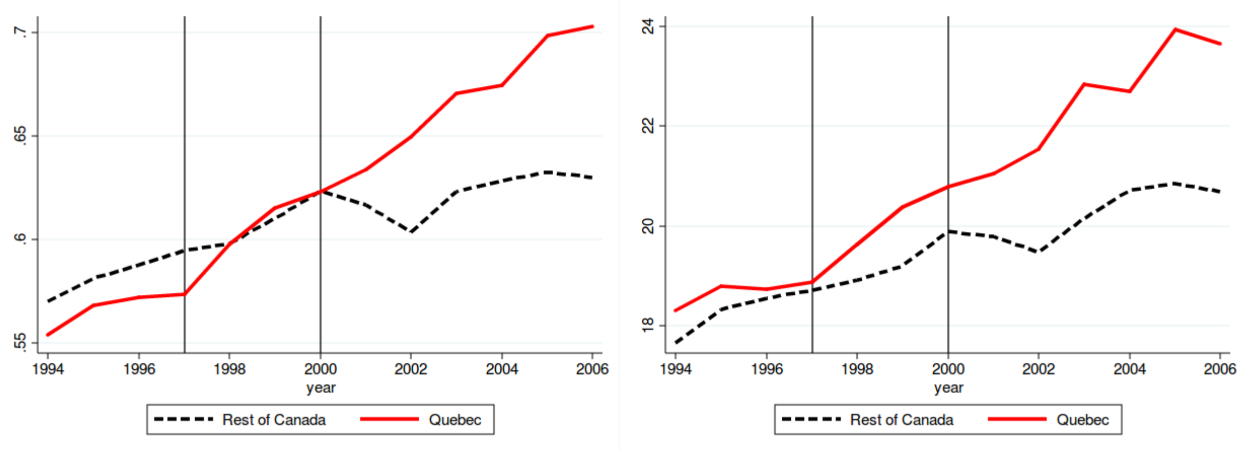
Table 2.10: Effect of a Daycare Price Increase on Public Daycare Use, Mother's Labor Supply, and Mother's Time Allocation on Work, Children, Home Production and Hobbies; Policy Impact by the Propensity of Using Public Daycare in the Absence of the Policy

	public daycare	**mother working**	**work**	**child**	**housework**	**errand**	**hobby**
Panel A: Regression Results							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
δ_1 : policy	0.238***	0.485***	12.650***	1.073	-1.505***	-0.276**	-0.598***
	(0.055)	(0.023)	(0.939)	(0.841)	(0.140)	(0.102)	(0.175)
δ_2 : policy	-0.340***	-0.610***	-18.885***	-1.714	2.087***	-0.259	0.720**
×propensity-daycare	(0.100)	(0.053)	(2.908)	(1.752)	(0.240)	(0.257)	(0.309)
δ_3 : college	0.028	0.025	2.545***	0.873*	-0.127	0.001	-0.345*
	(0.026)	(0.019)	(0.908)	(0.446)	(0.170)	(0.080)	(0.196)
δ_4 : university	0.068	0.012	2.609**	-0.363	-0.611***	-0.184**	-0.136
	(0.042)	(0.026)	(1.155)	(0.448)	(0.182)	(0.089)	(0.168)
R^2	0.31	0.09	0.15	0.08	0.17	0.06	0.07
N	2,915	2,915	2,871	2,908	2,904	2,896	2,859
Panel B: Means							
	0.318	0.1086	10.005	10.018	3.157	1.469	1.381

Note: *Panel A* shows the result of estimating the Difference-in-Differences model (2.15), on the Intent-to-Treat effect of a compensated daycare price increase, provided by the Thuringian daycare policy (2006). Standard errors are in parentheses and are clustered at the (bundesland×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: GSOEP (2000-2010), families with at least one 2 years old child, both parents at most 50 years old, excluding bundeslands Lower-Saxony, Bavaria and Baden-Wurttemberg. The outcomes are in bold in the headline of each column. The first two outcome variables are binary, work hours are measured per week, while the remaining activities (child, home production and leisure activities) are measured in hours per usual weekday. *Panel B* shows the mean for each outcome in the estimation sample. propensity-daycare denotes the predicted propensity of using public daycare in the absence of the policy.

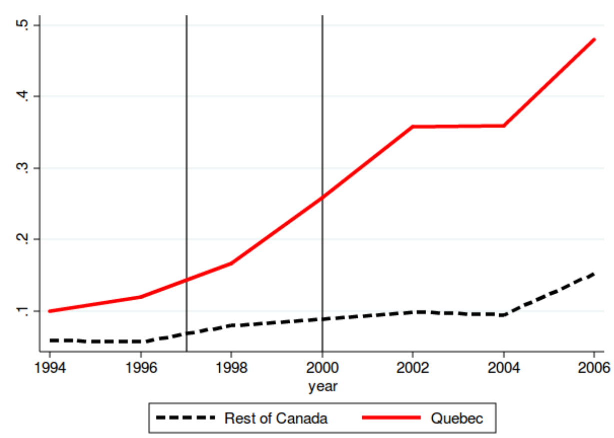
2.11 Figures

Figure 2.1: Trends in Quebec and the Rest-of-Canada in Maternal Employment (Both Margins)



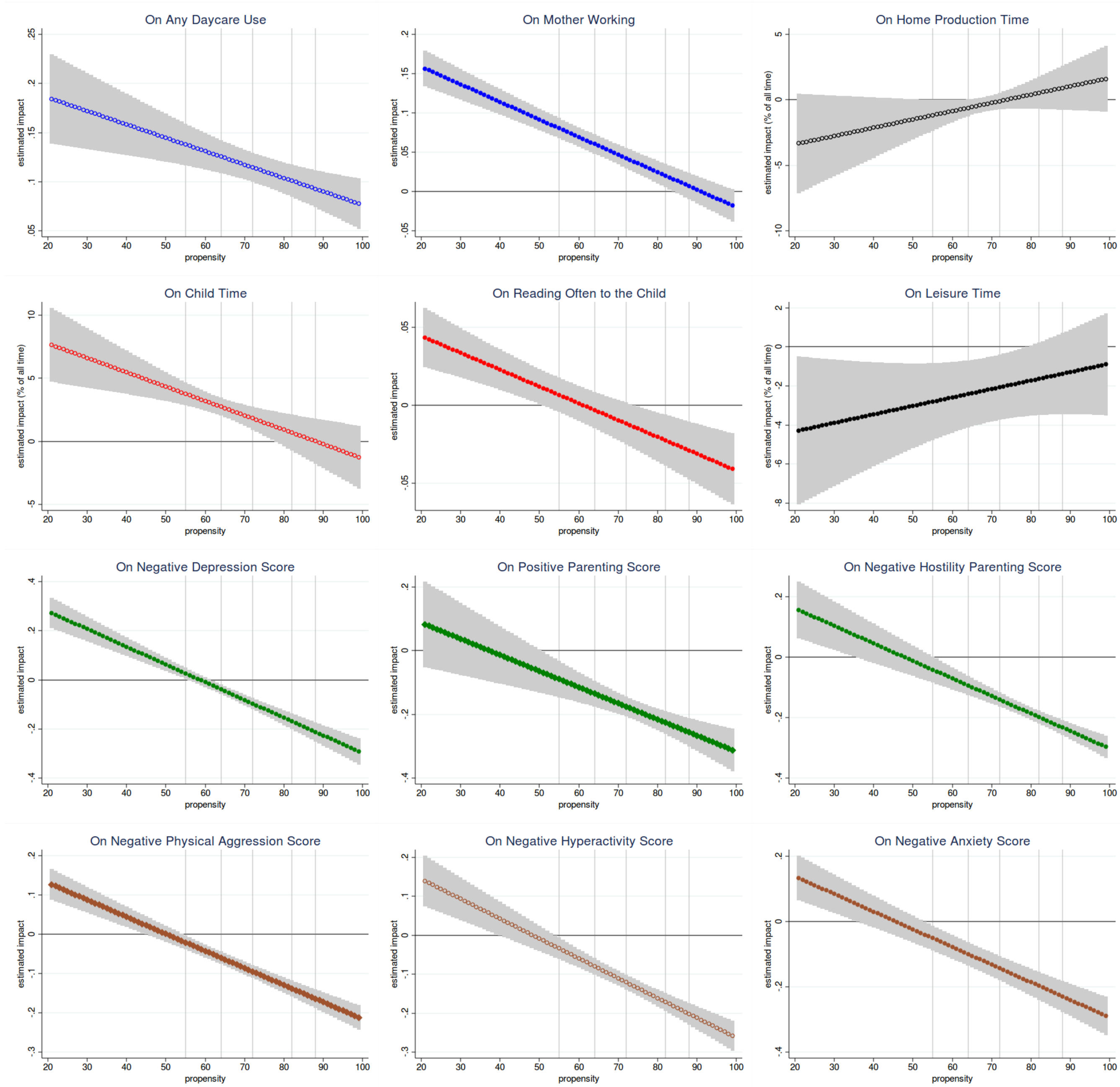
Note: these graphs show the weighted mean of women's employment rate and hours worked (below the age of 50, living in two-parent households), for Quebec and the rest of Canada between 1994 and 2006. Source of data: LFS (1994-2006), mothers with 0-4 years old children in two-parent families, both parents at most 50 years old. In 1997 the Quebec "5$/day" subsidized universal daycare policy was phased-in gradually for children aged 0-4 until 2000.

Figure 2.2: Trends in Quebec and the Rest-of-Canada in Regulated (Institutional Daycare Use)



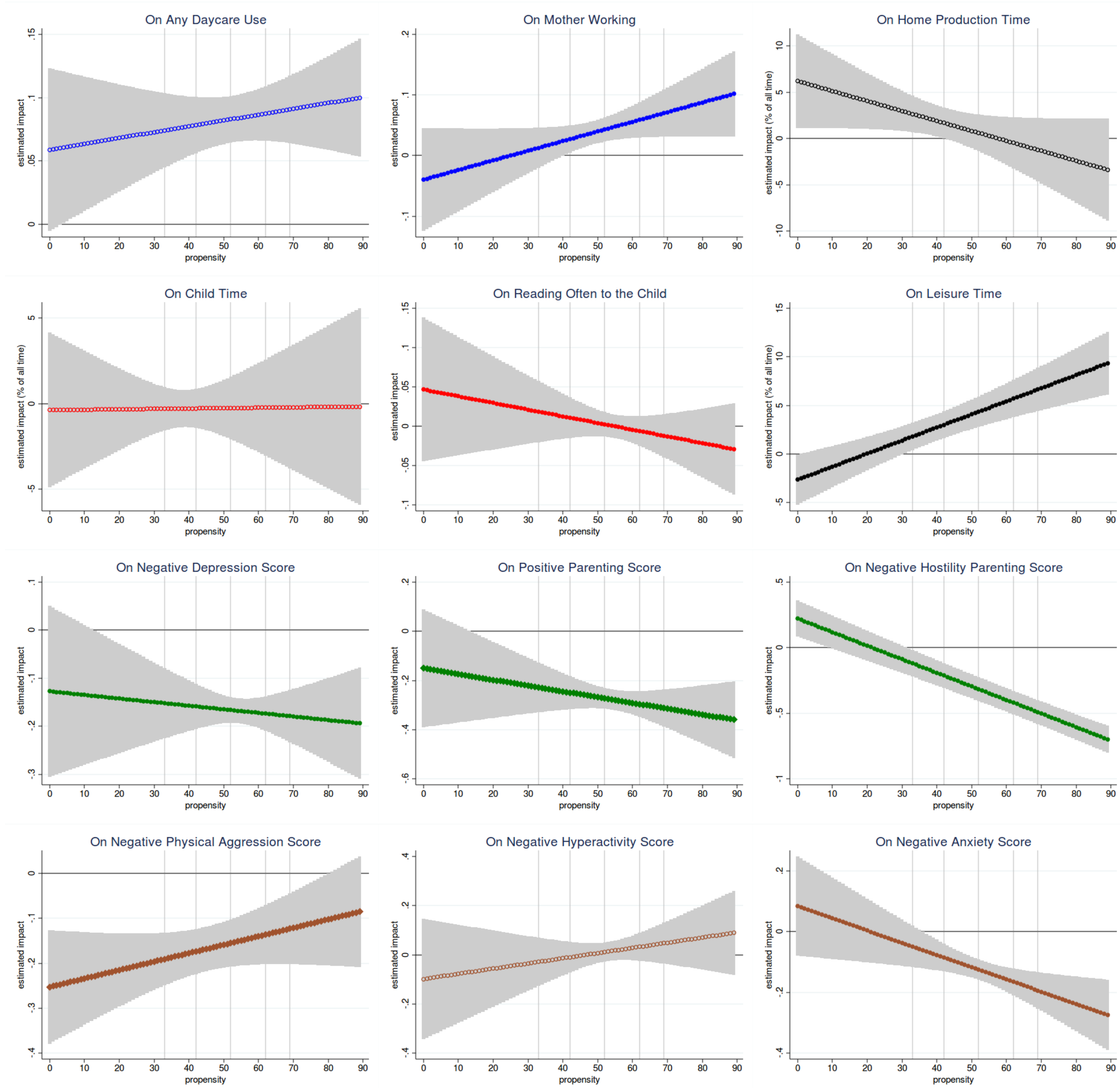
Note: this graph shows the weighted mean of regulated (institutional) daycare use, for Quebec and the rest of Canada between 1994 and 2006. Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old. In 1997 the Quebec "5$/day" subsidized universal daycare policy was phased-in gradually for children aged 0-4 until 2000.

Figure 2.3: Estimated Policy Impacts by Propensity of the Mother Working in the Absence of the Policy, with a 95% Confidence Band, for High-Educated Mothers (Linear Model)



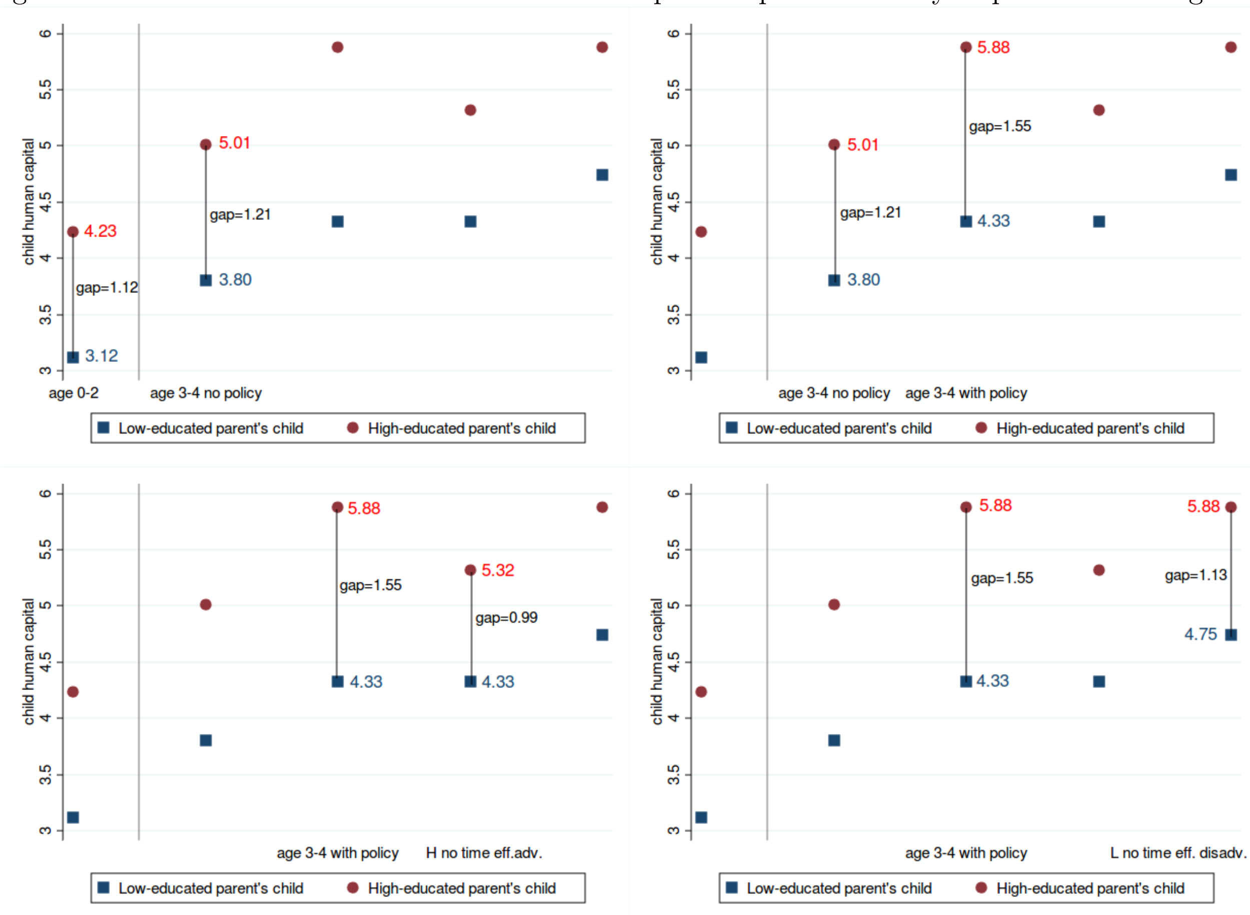
Note: these figures show the estimated policy impacts for high-educated families by the predicted propensities of the mother working in the absence of the policy, with a 95% confidence band; based on estimates from the linear model (2.10), where the underlying coefficients can be seen in Appendix Tables A.14-A.18. Source of data: NLSCY waves 1-7 (1994-2006) and GSS (1998,2005,2010), 0-4 years old children in two-parent families, both parents at most 50 years old. Propensities 55, 64, 72, 82 and 88 correspond to the 10th, 25th, 50th, 75th, 90th percentile of the predicted propensity distribution for high-educated mothers in the NLSCY.

Figure 2.4: Estimated Policy Impacts by Propensity of the Mother Working in the Absence of the Policy, with a 95% Confidence Band, for Low-Educated Mothers (Linear Model)



Note: these figures show the estimated policy impacts for low-educated families by the predicted propensities of the mother working in the absence of the policy, with a 95% confidence band; based on estimates from the linear model (2.10), where the underlying coefficients can be seen in Appendix Tables A.14-A.18. Source of data: NLSCY waves 1-7 (1994-2006) and GSS (1998,2005,2010), 0-4 years old children in two-parent families, both parents at most 50 years old. Propensities 33, 42, 52, 62 and 69 correspond to the 10th, 25th, 50th, 75th, 90th percentile of the predicted propensity distribution for low-educated mothers in the NLSCY.

Figure 2.5: Counterfactuals on the Child Human Capital Gap When Policy Implemented at Age 2.5



Note: these figures show the gap in child human capital before and after the policy change.

Panel1 (up-left): in the first panel for ages 0-2, the level of child human capital (K) for higher- and lower-educated parents' children in the control group are calculated with the observed average daycare hours (X_K) and hours spent with the child (T_K) as inputs in the CES production function (2.2) from Table A.27, using the estimated parameters $\hat{\rho}_K = -0.4$ and $\hat{\delta}_K = 0.36$.

For instance, K^L_{012}=3.1153=$\left[\left(1^{0.36} \times 40.515\right)^{(-0.4)} + 9.448^{(-0.4)}\right]^{\left(\frac{1}{-0.4}\right)}$.

The second vertical line shows how the gap would have evolved in the absence of the policy change, using the observed average daycare hours and hours spent with the child for ages 3-4 for the control group as inputs from Table A.27, using the estimated parameters $\hat{\rho}_K = -0.4$ and $\hat{\delta}_K = 0.36$.

Panel2 (up-right): the second panel shows how the gap evolves under the policy change, using the observed average daycare hours and hours spent with the child for ages 3-4 for the treated group as inputs from Table A.28, using the estimated parameters $\hat{\rho}_K = -0.4$ and $\hat{\delta}_K = 0.36$.

For instance, K^{34}_L=4.3264=$\left[\left(1^{0.36} \times (47.322 + 0.632)\right)^{(-0.4)} + (11.818 + 2.595)^{(-0.4)}\right]^{\left(\frac{1}{-0.4}\right)}$;

or K^H_{34}=5.8811=$\left[\left(1^{0.36} \times (47.322 + 0.632 + 0.228)\right)^{(-0.4)} + (11.818 + 2.595 + 5.840)^{(-0.4)}\right]^{\left(\frac{1}{-0.4}\right)}$.

Panel3 (down-left): the second vertical line in the third panel shows the counterfactual level of K if high-educated parents had no time efficiency advantage; i.e. if $\delta = 0$ or equivalently in this model, $S = 1$.

Panel4 (down-right): the second vertical line in the fourth panel shows the counterfactual level of K if low-educated parents had no time efficiency disadvantage, in the counterfactual case of $S = 2$.

Chapter 3

Returns to Starting School Later: Academic Redshirting vs. Lucky Date of Birth

3.1 Introduction

"When the Harvard sociologist Hilary Levey Friedman was expecting her first child, one thing worried her: her due date, January 3rd. It was uncomfortably close to January 1st, an often-used age cutoff for enrollment in academics and sports. "I was determined to keep him in until after January 1st," she said. And if the baby came early? "I actively thought about redshirting," she said."[77]

There is a non-trivial and potentially increasing fraction of parents postponing the entrance of their age-eligible child into kindergarten, a practice known as 'academic redshirting'.[78] For instance, while in 1968 96 percent of 6 years old children were enrolled in kindergarten in the United States, in 2005 only 84 percent.[79] The exact fraction of redshirted children in the United States is debated, but is in a range of 5 to 10 percent of kindergartners (Bassok and Reardon (2013), Stipek (2002), Aud et al. (2013)). Parents might have several reasons for justifying redshirting: the child might not be school-ready, or even if school-ready, they excessively value pre-school services. Alternatively, they may not want their child to start primary school as one of the youngest children, but hope that an extra year would allow her/him to excel relative to younger peers in the classroom.

Academic redshirting is essentially a voluntarily delayed school entry[80], differing from the involuntarily delayed school entry prescribed for children born after the school enrollment cutoff date. Children born after and complying with the school enrollment date enter kindergarten a year older, usually at the age of 6. Redshirted children enter kindergarten a year older than prescribed, usually at the age of 6 instead of 5. If not school-ready, an extra year can help the child to make up for any learning or developmental/behavioral deficits and enter school not only a year older, but also with boosted human capital. However, redshirting can be a beneficial compensating alternative even

[77]The New Yorker, September 19, 2013: http://www.newyorker.com/online/blogs/elements/2013/09/youngest-kid-smartest-kid.html.

[78]Redshirting is "*named for the red jersey worn in intra-team scrimmages by college athletes kept out of competition for a year*." The New Yorker, September 19, 2013.

[79]According to Deming and Dynarski (2008), the school attendance rate of six-year-olds has not decreased, but they are more likely to be enrolled in kindergarten rather than first grade. In addition, they show that only 25 percent of the aforementioned change can be explained by changes in school entry laws that increased kindergarten entrance ages. Since 1965, there is a trend of moving school enrollment cutoff dates from September-February to June-September in the US states; see Figure on page 2 in Cannon and Lipscomb (2008). Note that several states do not have a uniform cutoff date for all school districts (but delegate it to the discretion of Local Education Agencies) and some states do not have a legislation regarding kindergarten entry age.

[80]In what follows, 'being redshirted', 'withheld from primary school' or 'voluntarily delayed primary school entry' for an additional year are synonyms.

for a school-ready child: s/he could experience the gains of entering school a year older—gains an involuntarily delayed child has just from being luckily born after the school enrollment cutoff date.

This chapter is, to the best of my knowledge, the first attempt to measure the causal impact of academic redshirting on child outcomes, using a natural experimental design. I exploit an administrative barrier to academic redshirting in the Hungarian educational system: a school-readiness evaluation, compulsory for potentially redshirted children born before January 1st. By the research design, I measure the Local Average Treatment Effect (LATE) for children who would have not been redshirted if born before January 1st, but are redshirted if born after; the compliers are children who are or are not school-ready, and for whom the administrative barrier has a deterring impact. This LATE is then compared to the effect of involuntary delay of primary school entry by making use of the school enrollment cutoff date of June 1st; here the compliers are school-ready children assigned to enter primary school a year older. I utilize Hungarian administrative testscore, secondary school track choice and mental health survey data for years 2008-2014.

The rest of the chapter is organized as follows. Section 3.2 presents the details of the relevant institutional features of the Hungarian educational system that is necessary to understand the identification strategy in section 3.3. Section 3.4 discussed data and measurement issues, section 3.5 presents the results, while section 3.6 concludes.

3.2 Institutional Background

In this part of the chapter I present the details of the relevant institutional features of the Hungarian educational system, with a focus on transition from child care into primary school. I discuss the flexible regulation of both child care and primary school entry, and the possibility of academic redshirting. Most importantly, I describe the three regimes of primary school start based on month of birth that justify the importance of the two cutoff dates in the identification strategy, January 1st and June 1st.

Compared to North American educational systems, Hungary has a universal child care system[81] more integrated into the public education system, with 90 percent of the children spending at least 3 years in child care before entering primary school. The last child care year is the school-preparation year, approximately equivalent to the North-American kindergarten year, nevertheless still part of the child care system.[82] Redshirting a child means making her repeat this school-preparation year in kindergarten and postponing primary school entry.

Table 3.1 shows the most important institutional features of school entry in the Hungarian educational system, by month of birth.[83] *Regime1* pertains to children born in the months September

[81]Hungary has a universal child care system, with a minimal (less than 4 percent) fraction of private institutions. Public child care institutions in general provide an 8-hour long educational service per day free of charge, except for meal fees. According to the Act No. LXXIX of 1993 on Public Education (Public Education Act), 24(1), child care institutions educate children from the age of 3 until the start of compulsory school attendance. According to the enactment of the Ministry of Education, 137/1996. (VIII. 28.), the child care institutions' board of teachers is required to satisfy the physical and mental needs of child care-relevant aged children: to assure a healthy lifestyle, to provide emotional security, socialization, and to help mental skill enhancement.

[82]The minimum compulsory child care attendance time amount is 4 hours per day starting from 1st September for all children turning 5 in that particular calendar year (Public Education Act, 24(3)).

[83]Besides the date of birth affecting primary school entry age, it also influences, although to a lesser extent, child care

to December: these children are prescribed to enter primary school at the age of 6, and for them the administrative barrier to academic redshirting is present. *Regime2* pertains to children born in the months January to May: these children are prescribed to enter primary school at the age of 6, and for them there is no administrative barrier to academic redshirting. *Regime3* pertains to children born in the summer: these children delayed involuntarily, are prescribed to enter primary school at the age of 7 and for them academic redshirting is not allowed.

The end of child care attendance is bound together with the start of compulsory school attendance. All children turning 6 before June 1st in any calendar year are required to start primary school on 1st September of that particular calendar year (Public Education Act 6(1)). Thus, the school enrollment cutoff date is June 1st in Hungary. As a consequence, if complying with the prescribed primary school starting age, children born in the months January to May start primary school at the average age of 6 years 7.5 months to 6 years 3.5 months, children born in the months June to August start primary school at the average age of 7 years 2.5 months to 7 years 0.5 months and children born in the months September to December start primary school at the average age of 6 years 8.5 months months to 6 years 11.5 months. The regulation results in children born in May generally being the youngest and children born in June generally being the oldest in the classroom.

Nevertheless, with some restrictions and depending on month of birth, there is the possibility of academic redshirting, i.e. repeating the school-preparation year. As opposed to the United States, where redshirting is decided by the parents, together with significant input from child care providers, in Hungary redshirting is decided primarily by the parents, together with the evaluation of the child care institution's board of teachers and, depending on month of birth, possibly the local *Developmental Advisory Board* (DAB). While both the child care institution and the local DAB are part of the public education system, the latter is an independent pedagogic institution maintained by the local governments. According to the Public Education Act, 35(4), it is obliged to evaluate children's learning, social/integrational and behavioral skills until reaching adulthood, to propose and conduct skill enhancing activities involving the parents and child care teachers and, most importantly for the purpose of this chapter, evaluate school-readiness if asked by the child care institution and/or the parents.

In *Regime1*, children born in the months September to December are allowed to start their last

entry age. September-to-December-born children are more likely to enter child care at the age of 2, since these children are preferred by the child care institutions if there is a queue for entering child care at the age of 2. In particular, the child care institution can admit 2 years old children, provided that all the 3 years old and older applicants living in the particular municipality or capital district are admitted (Public Education Act 24(1)). However, children who are turning 3 within 6 months of acceptance are preferred. Acceptance usually happens at the beginning of June and start of child care happens only at the beginning of September (with very few exceptions). January-to-May-born children are most likely to enter at the age of 3, while summer-born children are most likely to enter at the age of 4. Table B.2 shows the distribution of children by month of birth and age at child care entry, together with the fraction of children entering school at the age of 7 and average years spent in child care in each cell. The first observation is that keeping month of birth fixed, the fraction of delayed children is monotonically increasing by child care entry age, and correspondingly, the average years spent in child care is monotonically decreasing by child care entry age. The second observation is that child care entry age is closely related to quarter of birth. September-to-December-born children are more likely to enter child care at the age of 2 and and less likely to enter at ages 3,4,5 than what their overall fraction would indicate. The reason behind is that these children are turning 3 within 6 months of acceptance, thus, they are preferred to enter at the age of 2 if local supply constraints permit it. January-to-May born children are more likely to enter child care at the age of 3, while summer-born children are more likely to enter at the age of 4.

year in child care on 1st September in the calendar year when they are turning 7, provided that the parents ask for the child's additional year in child care and both the child care institution's board of teachers and the local DAB supports this decision (Public Education Act, 24(5)). As a consequence, these children start primary school at an age between 7 years 8.5 months and almost 8 years, if redshirted.

In *Regime2*, children born in the months January to May are allowed to start their last year in child care on the 1st September of the calendar year when they are turning 6, provided that the parents ask for this and the child care institution's board of teachers supports this decision[84]; i.e. there is no need for the local DAB's evaluation or support in this case. As a consequence, these children start primary school at an age between 7 years 3.5 months and 7 years 7.5 months, if redshirted.

In *Regime3*, children born in the summer months are assigned to start primary school at the default age of 7 and are not allowed to be redshirted; they are the involuntarily delayed children.

The school-readiness examination conducted by the local DABs is free of charge and standardized. First, the developmental experts examine how the child behaves in social situations with other children. Usually during drawing exercises they investigate how the child is able to work while being distracted, how much she is influenced, able to adapt and handle new situations. Then, in a one-on-one situation the developmental experts investigate how much the child knows about her environment, how developed her thinking and remembering abilities are, and how easily she becomes tired. Finally, the results are shared and discussed with the parents and under most circumstances, with the child care teachers as well, in order to achieve a consensus that best meets the child's interests.

According to the interviews I conducted with the experts of 4 local DABs in Budapest[85], Hungary in June 2013, the major costs of a school-readiness examination are the following. First, in most cases the accompanying parent needs to miss work for the time of the examination. Additionally, there are traveling costs for parents living in municipalities where there is no DAB located (on average there are 4-5 DABs per county and Hungary has 20 counties). Second, parents might be afraid of the result, and of their child's potentially negative result being listed later, that might have—as they presume—harmful consequences on her educational career. Third, parents might be reluctant to go in front of an unknown formal committee; especially if the default option is to discuss their child's development and school entry with the child care teachers, with whom the parents are already familiar. Higher-educated parents have presumably lower costs of taking this examination, as for them traveling costs might be a lower share of the household budget, they might work more flexible hours, and might be less influenced by their child's potentially negative result being listed later, or going in front of an unknown formal committee. This impacts that the results as far as which parents are systematically more likely to show up for examination, and who are the compliers, for whom costs might have a deterring impact.

[84] According to the enactment of the Ministry of Education, 11/1994. (VI. 8.), 15(5), part "b", the child care institution's board of teachers is obliged to suggest an additional year in child care for every child who is assigned to start primary school on 1st September in that particular calendar year, but who does not meet the requirements of school-readiness according to the opinion of the board of child care teachers.

[85] The developmental experts emphasized during the interviews that their evaluation is suggestive, rather than decisive. The locally available child care services (number of child care places or teachers) are generally not taken into account by the local DABs; only developmental aspects.

3.3 Identification

In this section I discuss identification issues; first, identification of individual returns to starting school a year older, second, on how to disentangle absolute age effects from within-class relative age effects.

3.3.1 Identification of Individual Returns

Primary school entry age is not only determined by an exogenous rule based on date of birth, but depends additionally on the child's abilities at her entry-relevant age or the parents' preferences; both unobserved. As a consequence, Ordinary Least Squares (OLS) estimates of the effect of school entry age on school achievement and other related outcomes are biased.

Unobserved abilities bias the estimates downward. On one hand, unobserved child abilities around the age of 6 are negatively correlated with school entry delay. On the other hand, keeping everything else fixed, a child with worse abilities is expected to have worse student outcomes. Irrespective of the ultimate sign of the bias, it is expected to be smaller around the cutoff date of June 1st. The reason is that summer-born children are not allowed to be redshirted (at most "brought forward") and for them entering primary school at the age of 7 is prescribed by the regulation. At the same time, for non-summer born children, redshirting is a possible practice.

Unobserved parental valuation of child care leads to a bias with ambiguous sign. Parents with higher valuation of child care services are more likely to delay their child's primary school entry. These parents may be those who try to compensate for insufficient skill-enhancing resources at home, and whose children presumably score lower on achievement tests, *ceteris paribus.* In this case unobserved parental valuation of child care leads to a downward bias. Alternatively, these parents may be those who, besides other early childhood education institutions and resources, view child care as equally important and would want to provide it in the largest possible amount. The children of these parents presumably score higher on achievement tests, *ceteris paribus.*

From the previous description of the institutional background it follows that there are two main discountinuity points in the months of birth that greatly influence children's educational paths and opportunities. First, a child born on June 1st *vs.* 31st May is prescribed to enter primary school at the age of 7 *vs.* 6. Second, there is no need for the local DAB's evaluation or support for being redshirted for a child born on January 1st; an administrative barrier that is present for a child born on 31st December. To further support the importance of these discontinuity points, child care entry age is increasing when moving from children born in September to children born in August, and higher child care entry age is also associated with higher propensity of being delayed.

This chapter uses the Wald estimand for fuzzy Regression Discontinuity design (RDD) that, under conditions discussed below, captures the causal effect on compliers. Compliers are defined as children whose school starting age (∼treatment status) changes as moving from the left of the particular discontinuity point to the right. Since, as discussed soon, there is no information available about the exact birth date in the data set, each bin contains children born in the same month. A 3-months window, leading to 3-3 bins, is used on the two sides of the discontinuity points. Month of birth is controlled for linearly and the linear trend in month of birth is allowed to have different slopes on the two sides of the discontinuity points.[86]

The first-stage relationship is

$$D_i = \beta_0 + \beta_1 1\{X_i \geq x_d\} + \beta_2 X_i + \beta_3 1\{X_i \geq x_d\} \times X_i + \beta_4 C_i + \sum_{t=1}^{T-1} \mu_{1t} F_t + \eta_i, \tag{3.1}$$

and the second-stage relationship is

$$Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 X_i + \gamma_3 1\{X_i \geq x_d\} \times X_i + \gamma_4 C_i + \sum_{t=1}^{T-1} \mu_{1t} F_t + \upsilon_i, \tag{3.2}$$

where Y_i is the outcome variable from the set of {testscore, grade repetition, secondary school track choice, mental health}, D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6), X_i is month of birth (=linear trend, recentered at x_d), the discontinuity point x_d is either January 1st or June 1st for voluntary and involuntary delay, respectively, $1\{X \geq x_d\}$ is the discontinuity dummy, C_i denotes the vector of control variables and F_t variables denote year dummies corresponding to the year when child i was born (year is defined to start in September and end in August). Only $1\{X_i \geq x_d\}$ is used as an instrument and is excluded from the second-stage. The coefficients of interests are β_1 and γ_1.

It is implausible that the potential outcomes and thus the effect of school starting age is homogeneous in the population. Therefore, following Angrist and Pischke (2008, pp.111), the requirements for the aforementioned instruments to be valid in a heterogeneous framework are that (i) (*relevance or the existence of the first-stage relationship*) the instrument is related to the endogenous variable; (ii) (*independence*) the instrument is as good as randomly assigned: it is independent of the vector of potential outcomes and potential treatment assignments; (iii) (*excludability*) the instrument has no direct relationship to student achievement, only an indirect relationship through the decision made about primary school entry age. In the following I discuss each of these requirements in detail.

The existence of the first-stage relationship requires that month of birth should be systematically related to school starting age. More specifically, children who were born on/after January 1st and June 1st should have significantly higher probability of starting primary school at the age of 7 than children born before it, *ceteris paribus*.[87]

An assumption, needed for the independence requirement to hold, is that month of birth around the two cutoff points is effectively random. This requirement could fail in the following two cases. First, if children born on one side of the discontinuity points would have better-than-average innate abilities than children born on the other side. Or, to phrase it differently, if innate ability would not be uniformly distributed across month of birth near the cutoff points. Second, it fails if parents with particular preferences for early childhood investments plan conception systematically differently on the two sides of the discontinuity points. For instance, if they delay the birth of the child from

[86]The results continue to hold if using a 2-months window or a quadratic trend; these robustness checks are not included in this chapter for the sake of brevity.

[87]Table B.2 suggests the strength of the first-stage relationship in advance: there is a 17 and a 28 percentage points jump in the fraction of delayed children between children born in December and January and May and June, respectively.

May to June so that the child likely would be among the older students in the classroom, and at the same time they can provide the essential early childhood resources in their first years of lives. Table 3.2, discussed later in detail, provides evidence that neither the fraction of boys, nor the fraction of children with different parental education is changing discontinuously at the cutoff dates.

Bound and Jaeger (2001, pp.96-99) review the potential factors associated with month of birth, as health or personality traits. For instance schizophrenia, mental retardation, autism, multiple sclerosis or manic depression has been found to vary across quarter of birth, while an association has been found between shyness and month of birth. The evidence on the association between IQ and month of birth is inconclusive. Buckles and Hungerman (2013) show large differences in maternal characteristics for births throughout the year. They document that women giving birth in the winter months are younger, less educated, and less likely to be married than other women. They present evidence in favor of the relationship between month of birth and later outcomes being driven by differences in fertility patterns across socioeconomic groups, rather than by natural phenomena or schooling laws. I compare children born at most in two consecutive quarters, with one of the discontinuity points being in the middle of the winter, the other being at the beginning of the summer season, therefore differences between summer and winter born children documented in the literature are less relevant.

Dickert-Conlin and Elder (2010) test whether parents systematically plan childbirth to capture the option value of sending their child to school at a younger age and thus avoiding an additional year of child care costs. They find no evidence of the option value influencing time of birth. Using exact birth dates, McEwan and Shapiro (2008) and, although in a somewhat different context, McCrary and Royer (2011), do not find that either birth frequencies or students' observed socio-economic characteristics would vary sharply around enrollment cutoff dates, suggesting that unobserved characteristics would also vary smoothly around these dates.

The excludability requirement fails if children born after the appropriate cutoff points are affected in some way other than through an increased likelihood of delaying school entry. Clear potential threats are absolute or relative age effects: comparing two, otherwise identical children with the same school starting age, older students score higher.[88] As a consequence, month of birth after the appropriate cutoff date is correlated with testscore for at least two reasons: an increased likelihood of spending an extra in child care and a disadvantage stemming from being younger at the time test is taken. Even if month of birth were randomly assigned among children indicating that parents do not plan conception systematically differently, and even if innate abilities were evenly distributed across months, the channel of the effects of the instruments would still not be unique. This concern can be eliminated by comparing two children born just before or just after the cutoff date by narrowing down the window around the discontinuity point as much as possible. This approach is possible when exact birth date is

[88] For instance, consider first the cutoff date of January 1st and consider two otherwise identical children: the first born in October, the second born in February in the next year. Suppose that both children are redshirted. Even though the second child had a higher likelihood of being redshirted, the first one is still 4.5 months older than the second at the time the test is taken. Next, consider the June 1st cutoff and consider otherwise two identical children, the first born in March, the second born in July in the same year. Suppose that the first child is redshirted and the second complied to the prescribed school starting age of 7. Even though the second child had a higher likelihood of starting school at the age of 7, spending an additional, 4th year in child care, the first one is still 4.5 months older than the second at the time the test is taken.

known[89], unfortunately not available in the data set used in this chapter. However, a trend in month of birth can be controlled for and thus the age-at-the-test concern can be lessened. More importantly, month of birth needs to be controlled for to account for the steadily increasing fraction of delayed children born in September to August, seen in Figure 3.1[90], and in order to identify off exclusively of the jump in the fraction of redshirted children at the cutoff dates.

Figure 3.1 shows the identification strategy in a nutshell. The first-stage relationship between month of birth and the fraction of delayed children is marked with the dotted line. The reduced form relationship between month of birth and student achievement is marked with the dashed line. Finally, the discontinuity cutoffs of January 1st and June 1st are marked with solid reference lines.

The fraction of delayed children born in the months September to August is steadily increasing. However, there are clear jumps at the discontinuity points and exactly these jumps at January 1st and June 1st are used as instruments for voluntary and involuntary delay, respectively. The reduced-form relationship reveals that the average testscore by month of birth is steadily decreasing from June to December, then, the slope of the testscore sharply changes. Specifically, the average testscore is decreasing for children born between December and September, then jumps by 0.023 standard deviation between children born in December and January. Note that without the possibility of redshirting and with 1st of June as the school enrollment cutoff date, the average testscore by month of birth would presumably be steadily decreasing from June to May, from the oldest to youngest.

From the subgroup analysis by child care entry age it can be seen that the jump at January 1st is less present for children who started child care at the age of 2 (approx. 17 percentage of all children, 28 percentage of December-born children and 9.5 percentage of January-born children). From the subgroup analysis it can also be seen that the initial decrease in testscores for children born between September and December is driven by children who started child care at the age of 2; for children who started at the age of 3, this fraction is steadily increasing for children born between September and August. Despite these striking differences across children with different child care entry age, I do not control for child care entry age as that would induce endogenous selection based on parents' choice.

This chapter's identification strategy is most similar to the strategy used by Puhani and Weber (2007) and Kollo and Hamori (2011). To overcome the endogeneity of primary school starting age, they apply an Instrumental Variable (IV) identification strategy using month of birth as the instrument. Their instrument, theoretical or predicted school entry age has the following form:

$$I(b_i, s_i) = \left\{ \begin{array}{ll} \frac{(72+s_i)-b_i}{12} & \text{if } 1 \leq b_i \leq c \\ \frac{(84+s_i)-b_i}{12} & \text{if } c < b_i \leq 12 \end{array} \right\},$$

where b_i is child i's month of birth, s_i is the month of primary school start and c is the enrollment cutoff month. Note that Puhani and Weber (2007)'s instrument is exactly of the same form as this chapter's instrument around the cutoff date of June 1st. However, despite the fact that their instrument is a function of month of birth, they do not control for a linear trend in month of birth, which might

[89] Although Puhani and Weber (2007, pp.371) use this approach even in the absence of information about exact birth date. They use students born in the 2-months adjacent to the respective school entry cut-off point in their IV regressions to eliminate any direct seasonal effects and any bias from parental timing of birth.

[90] See also Table B.2.

question the validity of $I(b_i, s_i)$. On page 367 they mention that if b_i and s_i are exogenous, the theoretical school entry age is exogenous and can be used as an instrument for the actual age of school entry. They discuss factors that would invalidate their instrument, such as parental planning of conception, or that month of birth "*might exert some direct effect on physical and psychological health*" (Puhani and Weber, 2007, pp.371). However, as discussed previously, exogeneity is not enough for identification: even if month of birth was randomly assigned among children indicating that parents do not plan conception systematically differently, and even if innate abilities were evenly distributed across months, the channel of the causal effects of the instruments would not be unique if absolute or relative age effects are present. Also, trends around the cutoff date might be present.

In order for the IV estimator to provide the Local Average Treatment Effect (LATE), the *monotonicity* requirement has to be satisfied. Monotonicity requires that while the instrument may not influence school starting age for some children, all of those who are influenced are done in the same direction. Around the discontinuity point of January 1st monotonicity requires for all children to be true that if she is redshirted when being born before January 1st (in the presence of the administrative barrier to redshirting), then she would be definitely redshirted when being born on/after January 1st (in the absence of this barrier). Around the discontinuity point of June 1st monotonicity requires for all children to be true that if she starts school at the age of 7 when being born before June 1st (that is equivalent to being redshirted), then she would be definitely starting school at the age of 7 when being born on/after June 1st (as prescribed).[91] I believe the assumption for both cutoffs is an innocuous one, and is very likely to hold.

Given relevance, independence, excludability and monotonicity, the Wald estimand for fuzzy RDD has the interpretation of the local average effect of starting school at the age of 7, relative to 6, on complier children, whose treatment status is influenced by the change in the instrument around the discontinuities. Perhaps my main insight is that the impact of redshirting incorporates not only the age-impact (being older at the time of the test), but potentially also the impact of boosted human capital in that extra year before school for non-school-ready children. By comparing children born around January 1st, I measure the combined impact of age and boosted human capital due to redshirting, for complier children who might or might not have struggled with school-readiness problems. By comparing children born around the June 1st school enrollment cutoff date, I measure the sole age impact of starting school a year older for the complier children who, by definition, did not struggle with any school-readiness problems—they start late if born after, but would have started early if born before the enrollment cutoff date. My results suggest this distinction is important, despite that the school starting age literature concentrated almost exclusively on the sole age impact so far.

Around the discontinuity point of June 1st the compliers are school-ready children who enter primary school as prescribed by the regulation: 3 years if born before, but 4 years if born on/after June 1st. Perhaps it is easier to capture who are the non-compliers in this case. Either the particularly weak or immature students who are redshirted if born before June 1st, or the particularly strong students who are brought forward if born on/after June 1st. Alternatively, non-complier parents have a very clear opinion when their child should enter primary school and whose behavior is not influenced by

[91] Or, equivalently, it cannot happen that a child's school entry is delayed from age 6 to 7 when being born before June 1st, but is brought forward from 7 to 6 when being born on/after June 1st.

any enrollment cutoff date.

Around January 1st, the first group of complier children are those who are not school-ready, and would be redshirted if the parent only would have to discuss that issue with the child care teachers, but are not redshirted if the local DAB's approval is needed in addition. For these parents it is too costly to look for, visit, and let the child take the school-readiness examination; either because they work, or traveling costs are too high, or they are less informed, and/or distrustful with the system.

The second group of complier children are school-ready and whose parents could achieve their redshirting in the child care institution if only the child care teachers' approval is needed. However, these parents could be convinced by the developmental experts in the local DAB to let their school-ready child into primary school on time. These parents are either the ones who possess weaker lobbying power, or the ones who are more sophisticated/less concerned, so that they can be convinced.

Although it is impossible to identify the compliers in the data set, it is possible to assess their average characteristics *ex post.* Following Almond and Doyle (2011, pp.12-13), first let us define two, independent binary variables, the instrument Z_j and the endogenous variable D^j as follows:

$$Z_j = \left\{ \begin{array}{ll} 1 & \text{if born on/after the cutoff day} \\ 0 & \text{if born before the cutoff day} \end{array} \right\}; \; D^j = \left\{ \begin{array}{ll} 1 & \text{if starting school at the age of 7} \\ 0 & \text{if starting school at the age of 6} \end{array} \right\};$$

with j=1,2 corresponding to the January 1st and June 1st cutoffs, respectively. In addition, let us denote $D^j_{Z_j}$ as the value D^j would take if Z_j were either 1 or 0. Then compliers are those with $D^j_1 = 1$ and $D^j_0 = 0$. Let us denote their fraction π_{C_j}. Then, the observable characteristics of the compliers can be written as

$$E\left(X|D^j_1 = 1, D^j_0 = 0\right) = \frac{\pi_{C_j} + \pi_{A_j}}{\pi_{C_j}} \left[E\left(X|D^j = 1, Z_j = 1\right) - \frac{\pi_{A_j}}{\pi_{C_j} + \pi_{A_j}} E\left(D^j = 1, Z_j = 0\right)\right], \quad (3.3)$$

where π_{A_j} is the fraction of always-takers (for whom $D^j_1 = 1$ and $D^j_0 = 1$) and $\pi_{C_j} = 1 - \pi_{A_j} - \pi_{N_j}$, where π_{N_j} is the fraction of never-takers (for whom $D^j_1 = 0$ and $D^j_0 = 0$). Defiers (for whom $D^j_1 = 0$ and $D^j_0 = 1$) are assumed away by the monotonicity assumption. Using the independence between Z_j and D^j, $\pi_{A_j} = Prob(D^j_0 = 1)$ and $\pi_{N_j} = Prob(D^j_1 = 0)$.

Note that this chapter does not attempt to disentangle the composite effect of being one year older in class (and, at the time of the test) from having spent one additional year in pre-school for the following reasons. First, disentangling the composite effect in the Hungarian child care system would require fixing either the endogenous child care entry margin or the endogenous child care exit margin, leading to non-random sorting on the two sides of the cutoffs in birth date.[92] Second, there are institutional setups in other countries that provide the suitable natural experiment to compare children in the same grade having the same age-at-the-test, but having spent different amount of time in pre-school; see, for instance, Dustmann et al. (2013a) for evidence in the United Kingdom.[93] The

[92] For instance, by fixing the time spent in child care at 3 years the effect of starting school at the age of 7 could be estimated, if time spent in child care would be exogenous.

[93] Dustmann et al. (2013a) exploit variation in age-at-school-entry and effective length of first year in school. The effect of longer exposure to kindergarten, without controlling for age-at-the-test, is measured by DeCicca and Smith (2013) who exploit the introduction of the "dual entry" scheme into kindergarten mandated in 1990/1991 in British Columbia,

setup in Carlsson et al. (2015) is perhaps the closest to the ideal natural experiment to disentangle age-at-the-test impacts from impacts of time spent in school: they exploit conditionally random variation in the assigned test date for intelligence tests for 18 year-old Swedish males in preparation for military service. In their setup, both age at test date and number of days spent in school vary randomly after holding date of birth, parish, and expected graduation date fixed.

Also, this chapter can not disentangle the effect of school starting age from the effect of age-at-the-test, since all children are tested at the same date. Using Norwegian data, Black et al. (2011) are able to disentangle the aforementioned effects by using scores from IQ tests taken outside of school and exploiting the variation in the mapping between birth date and the year the test is taken.

Finally, it is not possible to compare a delayed child's student achievement with a non-delayed child of the same age. First, among non-grade-repeaters, there is no variation in school starting age by age-at-the-test, so additional information about students in a grade lower (5th/7th/9th) or a grade higher (7th/9th/11th) would be needed to do this exercise. Second, although there is some variation in school entry starting age by age-at-the-test, this variation comes solely from grade-repeaters.

3.3.2 Disentangling Absolute and Relative Age Effects

In contrast to the reviewed literature's data, the Hungarian administrative data used in this chapter has information about all students' month of birth and school starting age in a given class[94], thus the child's relative age rank in class is known. To separate the effect of being one year older from the relative age rank in class, I augment (3.2) and include the child's relative age rank in her class:

$$Y_i = \pi_0 + \pi_1 D_i + \pi_2 RR_i + \pi_3 X_i + \pi_4 1\{X_i \geq x_d\} \times X_i + \pi_5 C_i + \sum_{t=1}^{T-1} \tau_{1t} F_t + \varsigma_i, \tag{3.4}$$

where D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6), RR_i is the child's relative age rank in her class (in percentage terms), X_i is month of birth (=linear trend, recentered at x_d), the discontinuity point x_d is June 1st, $1\{X \geq x_d\}$ denotes the discontinuity dummy, C_i denotes the vector of control variables, F_t denotes year dummies corresponding to the year when child i was born (year is defined to start in September and end in August) and Y_i denotes the outcome variable (testscore). As an alternative to RR_i, I use a binary variable Q_i that equals 1 if the child is in the highest quartile in the age distribution in her class and 0 otherwise.

The coefficients of interests are π_1 and π_2. π_1 corresponds to the absolute age effect and shows the impact of entering school a year older, by comparing two very similar children who also have the same position in the age distribution in their class. π_2 corresponds to the relative age effect and shows

Canada. This provisional policy experiment unintentionally decreased time in kindergarten to 6 months for children born between November and December, 1985, while increased time in kindergarten to 16 months for those born between January and April, 1986.

[94] The class identifiers do not have any meaning across years, therefore they cannot be used for estimating class fixed effects.

the impact of moving up in the relative age rank by 1 percentage point, by comparing two very similar children from the same cohort who started school at the same age (both 6 or both 7).

Both D_i and RR_i/Q_i are endogenous. Besides $1\{X_i \geq x_d\}$, the two additional instrumental variables are the fraction of summer-born children in the class and its interaction with $1\{X_i \geq x_d\}$. The intuition for these instruments is the following: summer-born children enter primary school at the age of 7 by law, thus, in the presence of not many redshirted peers, they are the oldest in the class. A larger fraction of summer-born children decreases the relative rank of any individual summer-born child in the class, *ceteris paribus*. Note that this decrease is smaller for June-born children than for August-born ones. The impact of this fraction for children born before the school enrollment cutoff date depends on whether they are redshirted or not: for redshirted children the impact is positive, for on-time children the impact is negative. Since nearly 60 percent of March-to-May-born children are redshirted, the net impact is expected to be positive.

The fraction of summer-born children is a valid instrument if it is related to testscore only through the endogenous variables, D_i and RR_i/Q_i. As opposed to the child's relative age rank, the propensity of starting school at the age of 7 is less likely to be related to the fraction of summer-born children. However, this fraction violates the excludability assumption if children non-randomly sort into classrooms within a school, based on student achievement. Since summer-born children generally have higher testscores, classes concentrating high-achievement students would very likely have summer-born children overrepresented in them. It is essential to see whether results on separating absolute and relative age effects are robust to restricting the sample to children who study in schools that do not sort children systematically into classes within the school.

To separate non-sorting schools from sorting schools, I follow a slightly modified method of Horvath (2015): for each school I test whether classroom assignment is significantly related to (i) student achievement (mathematics testscore) or (ii) being born in the summer months. Technically, I run an F-test on the joint significance of classroom effects within school on outcomes (i) and (ii) (with year fixed effects). The nullhypothesis corresponds to no sorting, while the alternative hypothesis corresponds to sorting. These tests relates within-class sum-of-squares of testsores/fractions to between-class sum-of-squares and a large enough within-variation is indicative of heterogeneous classes within, but roughly similar between, therefore sorting can be not be accepted. As a last step I identify "sorting" schools as those with p-values of the F-test less than 0.05. As a consequence I end up with a sub-sample of schools that do not sort children into classes based on prior student achievement (called non-sorting schools "A") and a sub-sample of schools that do not sort children into classes based on being born in the summer (called non-sorting schools "B").[95]

[95] To illustrate sorting further, for each school I decompose the Total Sum-of-Squares of mathematics testscore, $\sum_{i=1}^{n_s}\left(Y_i - \bar{Y}_s\right)^2$ into Between Sum-of-Squares, $\sum_{i=1}^{n_s}\left(\bar{Y}_{cs} - \bar{Y}_s\right)^2$ and Within Sum-of-Squares $\sum_{i=1}^{n_s}\left(Y_i - \bar{Y}_{cs}\right)^2$, where $\bar{Y}_s$ is the average mathematics testscore in school s, n_s is the number of students in school s and $\bar{Y}_{cs}$ is the average mathematics testscore in class c in school s. Finally, I present the average fraction of Within Sum-of-Squares for all schools and non-sorting schools.

3.4 Data and Measurement

This chapter uses primarily data from two sources, one administrative and one survey dataset. First, data for 2008-2014 for 6th-, 8th- and 10th-graders on testscore in mathematics and reading, grade repetition and secondary school track choice, birth date and child care attendance is used, along with background characteristics from the Hungarian National Assessment of Basic Competences (HNABC). The time period 2008-2014 refers to the (uniform) date at which the test was taken in May. Second, data on mental health outcomes is used from the Hungarian Life Course Survey (HLCS).

The HNABC does not measure the students' knowledge regarding the compulsory curriculum; rather it measures how students are able to apply acquired skills in realistic situations, and to what extent they possess the necessary competences for further development. The tests are low-stake, the completion of the mathematics and reading test sheets was mandatory for all children, and the result of almost all students was centrally processed, counting towards the school's average achievement.[96]

Data on background variables, age-at-the-test, (imputed) school starting age and child care attendance is obtained from the student background survey of HNABC. The data set from this survey contains detailed information about the students' demographical and family background, but completion of the survey was non-compulsory. Unfortunately, mainly due to voluntary completion of the background survey, information cannot be used about all students at grades 6, 8 and 10 in years 2008-2014. The analysis is restricted to children for whom information about their time spent in child care, gender, year and month of birth is available, who have valid testscore either in reading or in mathematics and whose parents' highest educational attainment can be observed. The final sample contains 76.5-81 percent of the original sample.[97]

The excluded observations are very likely to be non-random. The missing response analysis of the HNABC 2006 and 2007 made at the Hungarian Academy of Sciences, Institute of Economics, Economics of Education Research Unit[98] reveals that among students who did not have valid testscores, those with lower parental education are overrepresented. Additionally, students with valid testscores obtained systematically higher grade from mathematics the year before. There was a significant negative relationship between testscores and non-response behavior in the background survey. As a consequence, excluded students with non-privileged family background and worse student achievement are presumed to be overrepresented. Since the returns of later school entry is expected to be larger for the excluded students, the estimates shown in this chapter are presumably a lower bound.

The HLCS is an individual panel survey conducted annually. The original sample (10,022 respondents, every 10th child from the population) was chosen in 2006 from the population of 8th-graders with valid testscores from the HNABC. Students were followed throughout their school career and the questionnaire contained detailed questions on ethnicity, schools, family background including poverty

[96]Exceptions were made for students with special educational needs, e.g. corporeally/sensually/mentally disabled or autistic students. Students who suffered from physic developmental deficits (e.g. suffered from behavioural problems or dyslexia/dysgraphia/dyscalculia) were required to complete the test sheet, but their results were not taken into account in the calculation of the school's achievement. Children who suffered from a temporary injury that made them physically unable to do this were not required to complete the test sheet, neither those who missed the class on the testing day.

[97]Table B.1 shows the details of sample selection.

[98]The missing response analyses are available at http://www.econ.core.hu/kutatas/edu/produktumok/kostb.html.

and home environment, etc. Each wave had a special block as early childhood environment, secondary school application, ethnicity, alcohol and drug usage, social network, and prejudices.

The outcomes are measured according to the following. Student achievement is measured by standardized testscore, where standardization was made for the Final Sample by all interactions of grade (6/8/10), year (2008-2014) and subject of the test (mathematics and reading). Grade repetition is measured by a binary variable indicating whether the child repeated at least one grade by grade 6, 8 and 10. Secondary school track choice is measured by a binary variable indicating whether the child attends the j-th track at grade 10. Track #1 is the lowest secondary school track, granting a vocational, but no high school degree upon completion. Track #2 is the middle secondary school track, granting both a vocational and a high school degree upon completion. Track #3 is the highest (academic) secondary school track, granting a high school degree upon completion. Track #1 does not allow continuation on tertiary level. Mental health outcomes are captured by an indicator variable denoting that child i has the j-th mental health condition, where $j = 1$ refers to having headache/stomachache/backpain every day, $j = 2$ refers to being afraid or being anxious or having sleeping problems every day or several days per week, and $j = 3$ refers to feeling dizzy or exhausted or nausea every day or several days per week.

The background survey of the HNABC does not contain information on whether the child was evaluated by the school-readiness committee, or whether there were concerns about school-readiness around the age of 6. The exact day of birth is unfortunately not available either. Although the background survey of the HNABC asks directly primary school starting age, the information of school starting age together with grade repetition show data inconsistencies in several cases.[99] Therefore, school starting age from birth date information (year of birth and month of birth) and grade repetition are imputed as follows. First, in the absence of exact birth date, assuming that all children were born on 15th of the appropriate month and that the achievement test was taken on 15th of May in all years 2008-2014, the age-at-the-test measured in months is computed according to the following formula:

$$\text{months-at-the-test} = (\text{year-of-the-test} - \text{year-of-birth} - 1) \times 12 + (12.5 - \text{month-of-birth}) + 4.5.$$

Second, assuming that school started on 1st of September in all years between 2008 and 2014, all children who have not repeated any grades lived 68.5 months since their primary school entry. Similarly, all children who repeated 1, 2 and 3 grades lived 80.5, 92.5 and 104.5 months since their primary school entry, respectively. Subtracting the appropriate months since school entry from age-at-the-test leads to the imputed primary school starting age.

Information about years spent in child care results in the imputed child care starting age. In cases where imputed child care starting age is smaller than 2 or larger than 5, it has been replaced by 2 and 5, respectively, and the number of years spent in child care has been modified accordingly.

[99] In these cases, if one believes the provided school starting age and grade repetition, it is impossible for the child to be in the 6th grade in the particular year, when testscore information is available for her. Moreover, families presumably answered these type of age-related questions differently based on whether the child already turned to the age of 6 when starting school, or she was in her 7th year of age. These differences can be especially striking between families where the child was born in September (thus almost e.g. 7, but technically 6 when starting school) or in May (thus unambiguously 6 when starting school).

The child background control variables are the followings: gender, highest parental educational attainment (at most primary education, vocational education, secondary education and tertiary education), indicators about the family's wealth situation (whether the child is considered to be disadvantaged or excessively disadvantaged[100], whether the child considers her family poor, whether the child is eligible for free meals and/or free books at school, whether the child has an own desk at home), the composition of the household (the number of the members of the household, whether the child lives together with biological father and mother vs. stepfather and stepmother), the nature and quality of interactions in the family (whether the parents regularly help in the homework, whether the child regularly discusses with the parents what happened at school or what she is currently reading), the amount and quality of cultural goods at home (access to internet, number of books at home, number of books the child has on its own), indicators about parental interest (how often the parents visit the school in order to discuss the child's development and other school issues with the teachers) and the municipality characteristics the family lives at (region and type of settlement).

Additionally, I control for the local child care situation[101] at the relevant age of school entry (at the age of 6 or 7) and at the age of 2 for the following reasons. First, if there is a scarcity of locally available child care services, then it may be the case that a child prescribed to enter primary school is not allowed to be delayed by the director of child care institutions and/or the child care teachers. Second, local child care circumstances likely influence child care entry age. Child care control variables stem from 3 data sources: KIR-Stat[102], the municipality-level demographical data set by cohorts of the Hungarian Statistical Authority (received from Gabor Kertesi) and the Local Government Treasury Dataset. If applicable, I aggregated institution-level data to municipality-level and then I matched to each observation the variables in the municipality she lived at her age of predicted primary school entry; i.e. at her age of 6 or 7, depending on month of birth. The following child care variables are used as controls: average child care class size, number of child care teachers and child care places relative to child care-relevant aged (3-6 years old) population, the ratio of 3, 4, 5, 6 years old children attending child care, a binary variable indicating if there exists no child care institution in the municipality the child lived at her age of predicted primary school entry, number of child care expenditures per capita, child care expenditures per available child care places and child care non-salary and salary expenditures relative to all local government expenditures.

[100] According to Act No. LXXIX of 1993 on Public Education (121(1)) a child is disadvantaged if her family is eligible for child protection support based on her social circumstances; a child is excessively disadvantaged if, in addition, one of her parents' highest educational attainment is primary school.

[101] Note that the decision of delaying childcare entry could be influenced by the primary school quality the parents intend to take their child into first grade. For instance, if there is a waiting list in the local good quality school, but excess supply of spots in the local worse quality school, parents would have an increased incentive to hold their child back in child care. Therefore, a school-specific primary school quality measure is estimated from an individual-level regression where standardized testscore is regressed on year fixed effects, gender, highest parental educational attainment, family background variables, childcare control variables and missing response characteristics controls; the school residuals are estimated out for each child. All the results shown in this chapter are robust whether the aforementioned measure, or its average version (within municipality), or school fixed effects or school-year fixed effects are included as controls.

[102] KIR-Stat provides the most comprehensive data about the Hungarian educational system. Every educational institution every year is required to fill out a data form according to the enactment of the Ministry of Education 229/2006. (XI. 20.).

3.5 Results

In this section I present the main results of the chapter. First, I show the results on individual returns to starting school a year older on several student and mental health outcomes. Second, I present the results on whether absolute age effects or within-class relative age effects dominate in the impact of starting school older due to the school enrollment cutoff date.

3.5.1 Results on Individual Returns

After showing some descriptive statistics of delayed children, in this part I present the fuzzy RDD/IV results for all (based on (3.2)) and by gender and parental education (based on (3.5)) for standardized testscores at grades 6, 8 and 10; grade repetition by grade 6 and 10; secondary school track choice after grade 8 and mental stability measures, as anxiety and exhaustion, at grade 8.[103] The second-stage relationship, where the treatment dummy D_i is interacted by gender and parental education is

$$\begin{gathered} Y_i = \gamma_0 + \gamma_1 D_i + \gamma_2 boy_i \times D_i + \gamma_3 lowp_i \times D_i + \gamma_4 boy_i \times lowp_i \times D_i \\ +\gamma_5 boy_i + \gamma_6 lowp_i + \gamma_7 boy_i \times lowp_i + \gamma_8 X_i + \gamma_9 1\left\{X_i \geq x_d\right\} \times X_i \\ +\gamma_{10} C_i + \textstyle\sum_{t=1}^{T-1} \xi_{1t} F_t + \vartheta_i, \end{gathered} \tag{3.5}$$

where D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6), X_i is month of birth (=linear trend, recentered at x_d), the discontinuity point x_d is either January 1st or June 1st, $1\left\{X \geq x_d\right\}$ denotes the discontinuity dummy, boy_i denotes the child's gender, $lowp_i$ denotes that the child has low-educated parents (the child's parents have at most vocational, but no high school degree), C_i denotes the vector of control variables, F_t denotes year dummies corresponding to the year when the child was born (year is defined to start in September and end in August), and Y_i denotes the outcome variable. In what follows, child i is considered and called low-status or disadvantaged if $lowp_i = 1$.

The first-stage relationship for all groups is in (3.1), where only $1\left\{X_i \geq x_d\right\}$ is used as an instrument and is excluded from the second-stage relationship. When assessing the first-stage relationship by gender and parental education, $1\left\{X_i \geq x_d\right\}$ and its interactions with gender and parental education are used as an instrument and are excluded. In the latter case there are four first-stage equations, corresponding to the endogenous variables D_i, $boy_i \times D_i$, $lowp_i \times D_i$, and $boy_i \times lowp_i \times D_i$. The first-stage results for D_i will be presented separately:

$$\begin{gathered} D_i = \beta_0 + \beta_1 1\left\{X_i \geq x_d\right\} + \beta_2 boy_i \times 1\left\{X_i \geq x_d\right\} + \beta_3 lowp_i \times 1\left\{X_i \geq x_d\right\} \\ +\beta_4 boy_i \times lowp_i \times 1\left\{X_i \geq x_d\right\} + \beta_5 boy_i + \beta_6 lowp_i + \beta_7 boy_i \times lowp_i \\ +\beta_8 X_i + \beta_9 1\left\{X_i \geq x_d\right\} \times X_i + \beta_{10} C_i + \textstyle\sum_{t=1}^{T-1} \mu_{1t} F_t + \eta_i. \end{gathered} \tag{3.6}$$

[103] The other outcomes that I considered but found no effect is high school completion, the first labor market outcomes and prosocial measures (alcohol, drug usage, smoking, teenage pregnancy and committing various crime types).

Descriptive Statistics of Delayed Children

Table 3.2 shows the fraction of boys and children with different parental education by quarter of birth and school starting age around the January 1st and June 1st cutoff dates.[104] Most importantly, there is no indication of any discontinuity of the fraction of boys or children with various family backgrounds around the cutoff dates. On both sides the fraction of boys is 49 percent, while the fraction of children with parents at most 8 years of schooling, vocational training, high school degree or tertiary degree is 11, 29, 31, 29 percent, respectively. The fraction of boys is approximately 14 percentage points higher among delayed children, while children with low-educated (high-educated) parents are overrepresented (underrepresented) among them.

The fraction of children with different developmental obstacles among redshirted and non-redshirted children can be seen in Table 3.3. Children are significantly more likely to be redshirted if they experienced a family shock (e.g. they were separated from their mothers/fathers before the age of 5 or the family had a negative income shock before then), a shock at birth (e.g. low birth weight, too early or birth with complications), chronic illnesses (ear or nerve system disease at ages 0-5) or developmental problems (attention disorder, cognition disorder, hyperactivity at ages 5-6).

First-stage Results and Complier Analysis

Table 3.4 shows the first-stage results on voluntary and involuntary school entry delay, using the administrative data.[105] In Table 3.4, *PanelA* shows the estimation result - coefficients, standard errors - for all groups (based on (3.1)), while PanelB1 shows the estimation result, based on (3.6). *PanelB2* shows the estimated effects of $Z_i = 1\{X_i \geq x_d\}$ by subgroups, and their significance level. *PanelB3* shows whether the effects of Z_i significantly differs across various subgroups. The values of the joint *F-statistics,* each of them above 300, suggest a very strong first-stage and no concerns passing tests of weak instruments (Stock et al. (2002) and, in the presence of clustering, Olea and Pflueger (2013)).

Children born after the January 1st and June 1st cutoff dates are by 11-13 and 18-25 percentage points (by 50 and 25 percent) more likely to be delayed than children born before. Boys are significantly more likely to be redshirted than girls, if born on or after January 1st than if born just before; this holds irrespective of their grade, family background and, in the case of high-status boys, their higher baseline propensity of being redshirted. For instance, a 6th-grader boy with high-educated parents has approximately 19 percentage points higher likelihood of being redshirted if born on or after January 1st than if born just before, while a very similar girl has only approximately 9 percentage points higher likelihood of being redshirted. The same effects for low-status boys and girls are 14.5 and 9 percentage points, respectively. Thus, being born on or after January 1st induces high-status boys (girls) being redshirted to a significantly larger (smaller) extent than very similar low-status boys (girls). The effect of quarter of birth is largest for high-educated parents' sons also when comparing it to their baseline fraction of redshirted students: in 6th grade, for instance, their estimated coefficient is 2/3 of their baseline fraction. The same fraction for high-educated and low-educated parents' daughters is 1/2,

[104] The table contains data from the administrative data, grade 6; the fractions are very similar in grade 8 and 10, therefore I omit those grades for the sake of brevity.

[105] Table B.3 shows the first-stage coefficients on all the endogenous variables.

while for low-educated parents' sons it is around 40 percent.

Girls are significantly more likely to be (involuntarily) delayed than boys, if born on or after June 1st than if born just before; a possible reason behind this is the girls' lower propensity of being redshirted if born in March to May. The effect of quarter of birth is highest for low-educated parents' daughters: they are approx. 35 percentage points more likely to start school at the age of 7 if born in the summer months, than if born before June 1st (the estimated effect is around 48 percent of their baseline propensity). The same effect for high-status girls, high-status boys and low-status boys is 28, 16 and 16 percentage points, respectively (40, 20 and 24 percent of their baseline propensity). For both genders, high-educated families are significantly more responsive to quarter of birth, although for boys the difference is significant only in grade 10.

Table 3.5 shows the first-stage results on voluntary and involuntary school entry delay, using the survey data.[106] Although, mainly due to the smaller sample size, the value of the joint *F-statistics* decreased, each of them is still above 20. Children born after the January 1st and June 1st cutoff dates are by 15 and 30 percentage points more likely to be delayed than children born before.

Similarly to the results using the administrative data, boys are significantly more likely to be redshirted than girls, if born on or after the corresponding cutoff date of January 1st; irrespective of their family background. The estimated coefficient for high-educated parents' sons is 4/5 of their baseline fraction of redshirted children, while for low-educated parents' sons it is 2/3. The same fraction for girls is around 55 percent. There are no significant differences by parental background for redshirting, presumably partly due to small sample size.

As in the case of the administrative data, girls and high-educated parents' children are significantly more likely to be induced to enter school at the age of 7 than boys and low-educated parents' children, if born in the summer than if born before. The effect of quarter of birth, in terms of their baseline fraction, is 68 and 57 percent for high-status and low-status girls, while it is around 33 percent for high-status and low-status boys.

The average characteristics of the compliers can be seen in Table 3.6, and can not be distinguished from the average characteristics of all the students. This result is reassuring in the sense that the results presumably do not correspond to a very specific sub-population.

Results on Testscore

Table 3.7 reveals that redshirted children achieved significantly, on average 0.24 of a standard deviation, higher score in mathematics on the 6th-grade and 8th-grade testings than on-time children, and the effects, though drop to 0.16, persist in secondary school. The effect of involuntary delay is very similar in grade 6, but drop more, to 0.1 standard deviation in grades 8 and 10. This suggests that boosting human capital of non-school-ready children have a longer-lasting impact on mathematics score, than being solely a year older at the time of the test. Table 3.8 shows that this is not the case for reading: the impact of school entry delay on 6th-, 8th-, and 10th-grade reading score is 0.23, 0.2 and 0.17 standard deviations, irrespective of whether it is voluntary or involuntary.

Redshirted disadvantaged boys score significantly higher on the 6th-grade mathematics test, than

[106]Table B.4 shows the first-stage coefficients on all the endogenous variables.

high-status boys, *ceteris paribus*; otherwise no significant differences in the impact of redshirting across gender or parental education can be detected. In the case of voluntary delay, disadvantaged boys benefit significantly more from being one year older at the time of the 6th-grade and 8th-grade mathematics and reading tests, than their otherwise very similar male high-status peers. Disadvantaged boys also benefit significantly more at the time of the 6th-grade and 8th-grade test, than disadvantaged girls; while high-status boys benefit more than high-status girls in grade 10.

Results on Grade Repetition

According to Table 3.9, redshirted high-status children have 3.6 percentage points smaller likelihood of grade repetition by grade 6 than on-time children, and these effects are very large compared to the baseline propensity of grade repetition of these groups. There are no detectable effects by grade 10 for them. Redshirted low-status girls have 5 percentage points smaller likelihood of repeating a grade by 6th grade than on-time low-status girls and this effect is 66 percent of their baseline propensity. Low-status boys realize a significantly larger advantage from redshirting: their point estimate is -7.5 percentage points (70 percent of their baseline propensity). Moreover, this effect persists through 10th grade. Since grade repetition is by far the biggest problem for low-status boys at grade 10, leading in several cases to dropping out from secondary school, this persistent effect is important.

The effect of starting school a year older due to the school enrollment cutoff date is more persistent for groups other than disadvantaged boys (for whom it is equally persistent). For instance, while the effects for high-status children at grade 6 are small and occasionally insignificant, high-status children who started school at the age of 7 if born in the summer (but would have started at the age of 6 if born before) have 2.5-3 percentage points smaller probability of repeating a grade at grade 10. For low-status children there are large impacts already at grade 6: their point estimate is approx. -3.5 percentage points, which is approx. 40 percentage of the baseline propensity of girls and boys, respectively. The point estimates increase further at grade 10 to -4.39 and -6.44 percentage points (32 and 37 percent of the baseline). The point estimates are significantly higher for disadvantaged boys and girls; although, their baseline propensity of grade repetition is also higher.

Despite the more persistent impacts of involuntary school entry delay to secondary school, the impacts of redshirting are larger in grades 6 and 8: redshirted children are by 4.5 percentage points less likely to repeat a grade in primary school, than on-time children, while for involuntary delay the impact is 2.5 percentage points. This suggests that boosting human capital of non-school-ready children have a larger impact on grade repetition in primary school, than solely entering school a year older.

Results on Secondary School Track Choice

The effects of voluntary and involuntary delay on secondary school track choice are in Table 3.10. Secondary school track choice is measured by a binary variable indicating whether the child attends the j-th track at grade 10. Track #1 is the lowest secondary school track, granting a vocational, but no high school degree upon completion. Track #2 is the middle track, granting both a vocational and a high school degree upon completion. Track #3 is the highest (academic) track, granting a high school degree upon completion. Track #1 does not allow continuation on tertiary level.

Redshirting significantly decreases the propensity of going to the vocational track by 9.25 percentage points (50 percent), and significantly increases the propensity of going to the academic track by 6.6 percentage points (16.5 percent), although the impact is significant only for boys and low-educated girls. For low-educated boys, the latter impact is almost half of the baseline propensity. The impact of involuntary delay is similar, although a bit smaller (larger) for the vocational (academic) track.

Redshirted high-status girls have significantly, 10 percentage points lower probability of attending the lowest track; this effect is remarkable, given the very low baseline propensity of them attending the non-academic tracks (approximately 40 percent). The impact is -8 and -7 percentage points for high-status boys and low-status girls, 70 and 25 percent of their baseline propensity, respectively. Redshirted low-status boys, overrepresented in the vocational track, have 12 percentage points (27 percent) lower likelihood of attending it, than on-time low-status boys, holding everything else constant.

High-status girls who were induced to start primary school at the age of 7 by the school enrollment cutoff date have significantly, 2.2 (6.06) percentage points lower probability of attending the vocational (middle) track at grade 10. The effect of involuntary school entry delay is -4 and -6 percentage points for high-status boys and low-status girls, 40 and 22 percent of their baseline propensity, respectively. Low-status boys have 8 percentage points lower likelihood of attending it if induced to start school at the age of 7 by the school enrollment cutoff date (20 percent of their baseline propensity).

The point estimates of the effect on lowest track attendance is significantly larger for boys, irrespective of family background; also, it is significantly larger for low-status than for high-status boys. The effect of involuntary delay on choosing the highest, academic track is between 7 and 8.4 percentage points for the various groups; in terms of their baseline propensity, the effect is the highest for low-status boys (50 percent), then for low-status girls (32 percent), high-status boys (19 percent) and high-status girls (13 percent). However, the differences in the point estimates are not significantly different from zero, either by gender or by parental background.

Results on Mental Health

The aforementioned results indicate that disadvantaged boys are the primary winners of delayed school entry, irrespective whether it was voluntary or involuntary. In addition, redshirting has a longer-lasting and larger impact than involuntary delay exactly on outcomes where disadvantaged boys enjoy the largest gains, as mathematics testscore, grade repetition in primary school and attending the vocational secondary school track. The following results suggest that mental health can be one of the mechanisms; i.e. boosted human capital in the extra year in child care leads to better mental health, associated with better student achievement.

Mental health outcomes are captured by an indicator variable denoting that child i has the j-th mental health condition, where $j = 1$ refers to having headache/stomachache/backpain every day, $j = 2$ refers to being afraid or being anxious or having sleeping problems every day or several days per week, and $j = 3$ refers to feeling dizzy or exhausted or nausea every day or several days per week. The aforementioned mental conditions likely prevent children from succeeding at school.[107]

[107] I have not found evidence so far for gender differences in the aforementioned mental conditions; Duckworth and Seligman (2008) and Duckworth et al. (2012) argue that girls outperform boys on report card results and grades due to their higher self-discipline.

According to Table 3.11, the effect of redshirting is significant only for low-educated parents' sons. For instance, a low-status boy who was redshirted in the absence of the administrative barrier but would not have been redshirted in the presence of it, is significantly less likely to suffer from everyday-pains, anxiety, sleeping problems and exhaustion than a very similar non-redshirted peer. The corresponding point estimates are -15,-29 and -21 percentage points, and are very large given the baseline propensity of low-status boys having these mental conditions. The effect of involuntary school entry delay is significant only in one case: on severe exhaustion for high-educated parents' sons.

3.5.2 Results on Absolute vs. Relative Age Effects

Figure 3.2 shows the distributions of class size, the relative age rank in class (RR, in percentiles), and the fraction of summer-born children in class, using all 6th-graders in 2008-2014 in the Hungarian administrative data, separately for all schools and for non-sorting schools.[108]

Table 3.12 contains the result of the first-stage regressions on the endogenous D_i and RR_i/Q_i variables, separately for all schools and for non-sorting schools (based on non-sorting by prior testscores (*"A"*) or month of birth (*"B"*). Being born in the summer months, as opposed to being born between March and May, significantly increases both the likelihood of starting school at the age of 7 and the child's relative age rank in the class, either measured in continuous (RR_i) or binary form (Q_i). The fraction of summer-born peers in the class is, as expected, not related to school starting age D_i. On the contrary, it is significantly related to the relative age rank in class: for summer-born children a 1 percentage point increase in the fraction of summer-born classmates leads to 0.11 percentage points decrease in their relative age rank (measured in percentiles) or a 0.1 percent decrease in the probability of ending up in the highest quartile of the age distribution in their class, *ceteris paribus.* For children born between March and May a 1 percentage point increase in the fraction of summer-born classmates leads to 0.1 percentage points increase in their relative rank (measured in percentiles) or a 0.3 percent increase in the probability of ending up in the highest quartile of the age distribution in class. These results are robust to sub-samples of only non-sorting schools.

Regarding the strength of the instruments, the *joint F-statistic* in the first-stage of the aforementioned endogenous variables is all above 100, when all schools are considered. Due to the reduction of the sample size, the aforementioned value considerably drops when only non-sorting schools are considered, nevertheless still above 40 in each case. In addition, each of the *Cragg-Donald Wald F-statistics* are above 70 and each of the *Kleibergen-Papp F-statistics* are above 40. Therefore, weak identification can be excluded. Regarding the validity of the instruments, each of the *p-values of the Hansen J-statistic* are above 0.2, therefore the nullhypothesis of valid (excludable) instruments cannot be rejected at the usual significance levels.

Table 3.13 shows the OLS and IV results for the effect of D_i and RR_i/Q_i on 6th-grade standardized mathematics testscore, separately for all schools and for non-sorting schools (based on non-sorting by prior testscores (*"A"*) or month of birth (*"B"*)). Columns (1) to (3) show the OLS results if absolute

[108]See Table B.5 for further descriptive statistics. Table B.6 shows the fraction of Within Sum-of-Squares of mathematics testscore for school-year pairs with at least 2 classes in grade 6.

and relative age effects are entered separately. It can be seen that both starting primary school at the age of 7 and the relative age rank in class are either negatively associated with student achievement, or not significantly related to it. Once absolute and relative age effects are entered together (columns (4) and (5)), the association between school starting age and testscore switches sign and becomes significantly positive. On the other hand, there is a negative association between relative age rank and student achievement, holding school starting age fixed.

Looking at the IV results, both higher school starting age and the relative age rank positively affect student achievement if entered separately. For instance, according to columns (6), entering primary school a year older leads to a 0.22-0.27 standard deviation increase in mathematics testscore in grade 6 (depending on the subset of schools considered), *ceteris paribus*. In contrary, moving up on the relative age distribution in the class by 1 percentage point leads to a 0.37-0.48 percentage points increase in the score, *ceteris paribus*. However, column (8) and (9) reveal that once starting primary school at the age of 7, the relative rank has no additional causal effect on testscores. According to column (8), a child who entered primary school at the age of 7 is predicted to have 0.2-0.35 standard deviations higher student achievement than the one who entered primary school at the age of 6, and has the same position in the relative age distribution in her class. These results indicate that absolute age effects dominate relative age effects.

3.6 Conclusion

In this chapter I measure the causal effect of academic redshirting using Hungarian administrative testscore data and survey data on mental health for years 2008-2014. The main institutional feature exploited is a school-readiness evaluation, compulsory for potentially redshirted children born before January 1st. By comparing children born around January 1st, I measure the combined impact of age and boosted human capital due to academic redshirting, for the complier children who might or might not have struggled with school-readiness problems, and for whom this administrative barrier had a deterring impact. By comparing children born around the June 1st school enrollment cutoff date, I measure the sole age impact of starting school a year older due to involuntary school entry delay, for the complier children, who by definition do not struggle with any school-readiness problems: they start late if born after school enrollment cutoff date, but would have started early if born before.

I find four striking results. First, although there are large gains for all children, disadvantaged boys benefit the most from school entry delay. Second, redshirting has a longer-lasting or larger impact than involuntary delay exactly on outcomes where disadvantaged boys enjoy the largest gains; namely grade repetition (especially in primary school), mathematics testscore and avoiding the lowest, vocational secondary school track. Third, redshirting impacts mental health, measured by anxiety and exhaustion, positively only for disadvantaged boys. Finally, exploiting natural variation in the fraction of summer-born children in class, I find the positive effects of higher school entry age to be driven by absolute, rather than within-class relative age effects.

3.7 Tables

Table 3.1: Predicted Path into Primary School in Hungary, by Month of Birth

Month of Birth	Predicted Age at School Entry	Possibility for Redshirting?	Regime
June	7 years 2.5 months	no	3
July	7 years 1.5 months	no	3
August	7 years 0.5 months	no	3
September	6 years 11.5 months	yes, with administrative barrier	1
October	6 years 10.5 months	yes, with administrative barrier	1
November	6 years 9.5 months	yes, with administrative barrier	1
December	6 years 8.5 months	yes, with administrative barrier	1
January	6 years 7.5 months	yes, no administrative barrier	2
February	6 years 6.5 months	yes, no administrative barrier	2
March	6 years 5.5 months	yes, no administrative barrier	2
April	6 years 4.5 months	yes, no administrative barrier	2
May	6 years 3.5 months	yes, no administrative barrier	2

Table 3.2: The Fraction of Boys and Children with Different Parental Education, by Month of Birth and School Starting Age, Administrative Data Grade 6

month of birth:	**October-December**		**January-March**	
x_d : *January 1st*	**non-delayed**	**delayed**	**non-delayed**	**delayed**
fraction of boys	47.474%	61.714%	44.072%	58.222%
(mean)	*49.433%*		*49.252%*	
fraction of parental education: at most 8 years in school	9.259%	24.216%	8.782%	15.271%
(mean)	*11.584%*		*11.310%*	
fraction of parental education: at most vocational training	28.652%	35.080%	28.146%	31.446%
(mean)	*29.437%*		*29.439%*	
fraction of parental education: at most high school degree	32.437%	23.467%	32.755%	28.776%
(mean)	*31.043%*		*31.196%*	
fraction of parental education: tertiary degree	29.652%	17.236%	30.317%	24.507%
(mean)	*27.936%*		*28.056%*	
month of birth:	**March-May**		**June-August**	
x_d : *June 1st*	**non-delayed**	**delayed**	**non-delayed**	**delayed**
fraction of boys	41.182%	55.743%	35.408%	49.664%
(mean)	*49.615%*		*49.471%*	
fraction of parental education: at most 8 years in school	9.083%	12.303%	9.803%	10.915%
(mean)	*11.428%*		*11.648%*	
fraction of parental education: at most vocational training	28.924%	28.888%	23.754%	29.666%
(mean)	*28.982%*		*29.627%*	
fraction of parental education: at most high school degree	32.438%	30.320%	27.720%	31.542%
(mean)	*30.937%*		*30.976%*	
fraction of parental education: tertiary degree	29.555%	28.488%	38.723%	27.877%
(mean)	*28.653%*		*27.749%*	

Note: source of data: HNABC (grade 6) 2008-2014.

Table 3.3: The Fraction of Children with Different Developmental Obstacles born between September and May by School Starting Age, Survey Data

fraction of children who:	***among redshirted***	***among on-time***	***p-value of diff.***
were separated from mother before age 5	3.68%	2.28%	0.002
were separated from father before age 5	16.56%	13.40%	0.001
were born before week 36	12.01%	6.94%	0.000
were born with complications	16.01%	13.97%	0.038
were born with low birth weight	14.35%	8.36%	0.000
were born with very low birth weight	3.39%	1.59%	0.000
said first words 1 year or older	36.98%	30.73%	0.000
had problems with speaking at age 5	25.24%	17.56%	0.000
had chronic ear disease at ages 0-3	14.21%	10.30%	0.000
had ever problems with nerve system	1.33%	0.70%	0.012
were diagnosed with dislexia/disgrafia/discalculia at ages 6-7	19.42%	12.02%	0.000
had cognition disorder at ages 5-6	3.73%	1.48%	0.017
had attention disorder at ages 5-6	5.80%	2.74%	0.000
diagnosed with hyperactivity at ages 5-6	2.84%	1.94%	0.029
had negative income shock in family before age 6	10.35%	7.95%	0.000

Note: source of data: HCLS. The last column tests whether the difference in the fraction between redshirted and non-redshirted children in significantly different from 0. *** p<0.01, ** p<0.05, * p<0.1.

Table 3.4: The Effect of Quarter of Birth - First-stage Results on School Entry Delay, with full set of interactions by Gender and Parental Education, Administrative Data, Grades 6/8/10

$D_i = 1\{\text{school entry at age 7}\}$	***Academic Redshirting***			***Involuntary Delay***		
	x_d : *January 1st*			x_d : *June 1st*		
grades	**6**	**8**	**10**	**6**	**8**	**10**
Panel A: effects for all groups						
$Z_i = 1\{X_i \geq x_d\}$	0.1312***	0.1139***	0.1149***	0.1819***	0.2174***	0.2525***
	[0.004]	[0.003]	[0.003]	[0.005]	[0.005]	[0.005]
N	223,924	206,499	219,901	227,185	219,173	214,092
R^2	0.177	0.181	0.127	0.276	0.302	0.333
mean of dependent variable:	*0.263*	*0.200*	*0.197*	*0.768*	*0.746*	*0.712*
Panel B: effects for subgroups						
Panel B1: coefficients and standard errors						
$Z_i = 1\{X_i \geq x_d\}$	0.0884***	0.0703***	0.0730***	0.2553***	0.2896***	0.3086***
	[0.005]	[0.004]	[0.004]	[0.006]	[0.006]	[0.006]
$boy_i \times Z_i$	0.0987***	0.0850***	0.0860***	-0.1326***	-0.1327***	-0.1186***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
$lowp_i \times Z_i$	0.0122**	0.0230***	0.0122**	-0.0304***	-0.0352***	-0.0088*
	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]	[0.005]
$lowp_i \times boy_i \times Z_i$	-0.0540***	-0.0335***	-0.0231***	0.0241***	0.0372***	0.0286***
	[0.007]	[0.007]	[0.007]	[0.006]	[0.006]	[0.007]
N	223,924	206,499	219,901	227,185	219,173	214,092
R^2	0.179	0.183	0.13	0.281	0.307	0.337
joint $F-statistic$	*395.49*	*356.4*	*360.93*	*689.48*	*815.57*	*1939.44*
Panel B2: estimated effects of Z_i						
girl, high parental educ.	0.088***	0.070***	0.073***	0.255***	0.290***	0.309***
mean of dependent variable:	*0.177*	*0.134*	*0.148*	*0.716*	*0.694*	*0.666*
boy, high parental educ.	0.187***	0.155***	0.159***	0.123***	0.157***	0.190***
mean of dependent variable:	*0.283*	*0.217*	*0.221*	*0.807*	*0.786*	*0.755*
girl, low parental educ.	0.134***	0.085***	0.100***	0.332***	0.375***	0.373***
mean of dependent variable:	*0.265*	*0.203*	*0.191*	*0.748*	*0.729*	*0.688*
boy, low parental educ.	0.145***	0.145***	0.148***	0.116***	0.159***	0.210***
mean of dependent variable:	*0.358*	*0.281*	*0.257*	*0.811*	*0.789*	*0.749*
Panel B3: p-values of testing differences of effects of Z_i						
girl, high-low parental educ.	0.022	0.000	0.013	0.000	0.000	0.091
boy, high-low parental educ.	0.000	0.067	0.059	0.216	0.691	0.000
high parental educ., boy-girl	0.000	0.000	0.000	0.000	0.000	0.000
low parental educ., boy-girl	0.000	0.000	0.000	0.000	0.000	0.000

Note: this table shows the result of estimating a LPM of D_i on Z_i, $boy_i \times Z_i$, $lowp_i \times Z_i$, $boy_i \times lowp_i \times Z_i$ and control variables. D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). Z_i is a binary variable denoting quarter of birth of child i (1: child was born on/after January 1st or June 1st, 0: child was born before January 1st or June 1st). Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Estimation restricted to children who were born in the three-months window around the cutoff dates of January 1st and June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HNABC (grades 6,8,10) 2008-2014, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding equation is (3.1).

Table 3.5: The Effect of Quarter of Birth - First-stage Results on School Entry Delay, with full set of interactions by Gender and Parental Education, Survey Data

$D_i = 1\{\text{school entry at age 7}\}$	***Academic Redshirting***	***Involuntary Delay***
	x_d : *January 1st*	x_d : *June 1st*
Panel A: effects for all groups		
$Z_i = 1\{X_i \geq x_d\}$	0.1533***	0.3022***
	[0.019]	[0.024]
N	6,899	6,116
R^2	0.210	0.329
mean of dependent variable:	*0.2377*	*58.92*
Panel B: effects for subgroups		
Panel B1: coefficients and standard errors		
$Z_i = 1\{X_i \geq x_d\}$	0.0990***	0.3583***
	[0.023]	[0.030]
$boy_i \times Z_i$	0.0868***	-0.0555**
	[0.023]	[0.028]
$lowp_i \times Z_i$	0.0338	-0.0647**
	[0.029]	[0.033]
$lowp_i \times boy_i \times Z_i$	-0.0171	-0.0152
	[0.040]	[0.044]
N	6,899	6,116
R^2	0.213	0.331
$joint\ F - statistic$	*20.85*	*42.85*
Panel B2: estimated effects of Z_i		
girl, high parental educ.	0.0990***	0.3583***
mean of dependent variable:	*0.176*	*0.522*
boy, high parental educ.	0.1858***	0.3028***
mean of dependent variable:	*0.238*	*0.614*
girl, low parental educ.	0.1328***	0.2936***
mean of dependent variable:	*0.245*	*0.579*
boy, low parental educ.	0.2025***	0.2229***
mean of dependent variable:	*0.318*	*0.659*
Panel B3: p-values of testing differences of effects of Z_i		
girl, high-low parental educ.	0.238	0.050
boy, high-low parental educ.	0.561	0.008
high parental educ., boy-girl	0.000	0.047
low parental educ., boy-girl	0.035	0.043

Note: this table shows the result of estimating a LPM of D_i on Z_i, $boy_i \times Z_i$, $lowp_i \times Z_i$, $boy_i \times lowp_i \times Z_i$ and control variables. D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). Z_i is a binary variable denoting quarter of birth of child i (1: child was born on/after January 1st or June 1st, 0: child was born before January 1st or June 1st). Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Estimation restricted to children who were born in the three-months window around the cutoff dates of January 1st and June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HLCS, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding equation is (3.1).

Table 3.6: Academic Redshirting and Involuntary Delay, Average Characteristics of Compliers, Administrative Data

	January 1st cutoff		*June 1st cutoff*	
characteristics	**complier**	*sample*	**complier**	*sample*
female	0.4476	0.5101	0.5105	0.5080
highest parental education: primary	0.1061	0.1107	0.1057	0.1085
highest parental education: vocational	0.2915	0.3023	0.3042	0.2984
highest parental education: secondary	0.3209	0.3196	0.3197	0.3175
highest parental education: tertiary	0.2815	0.2675	0.2703	0.2756
disadvantaged	0.2940	0.2929	0.3042	0.3013
excessively disadvantaged	0.1028	0.1026	0.1020	0.1028
discount on meal at school	0.2712	0.2774	0.2673	0.2725
free meal at school	0.2525	0.2030	0.2522	0.2380
free books at school	0.4881	0.4881	0.4947	0.4939
# of books at home : 0-50	0.1534	0.1522	0.1515	0.1523
internet at home	0.8176	0.7612	0.8084	0.7939
own books at home	0.9535	0.9541	0.9555	0.9544
own desk at home	0.9164	0.9185	0.9220	0.9207
considers own family to be poor	0.1961	0.1836	0.1884	0.1867
biological mother in household	0.9532	0.9625	0.9612	0.9610
biological father in household	0.7271	0.7610	0.7475	0.7507
stepmother in household	0.0206	0.0157	0.0156	0.0159
stepfather in household	0.1003	0.0948	0.0981	0.0968
# of individuals in household	4.4411	4.4672	4.4509	4.4682
parents, siblings help in homework	0.2985	0.2916	0.2915	0.2941
school issues discussed at home	0.7638	0.7633	0.7698	0.7682
child's current readings discussed at home	0.1795	0.1680	0.1712	0.1705
parents regularly visit school meetings	0.8442	0.8526	0.8617	0.8573
type of settlement: county centre	0.2140	0.1610	0.1574	0.1586
type of settlement: other city	0.3493	0.3526	0.3523	0.3501
type of settlement: village	0.2968	0.3739	0.3765	0.3765
region: Central Trans-Danubia	0.1109	0.1124	0.1110	0.1107
region: Western Trans-Danubia	0.0906	0.0971	0.0996	0.0987
region: Southern Trans-Danubia	0.0968	0.0961	0.0954	0.0942
region: Northern Hungary	0.1210	0.1339	0.1325	0.1330
region: Northern Great Plain	0.1810	0.1841	0.1836	0.1840
region: Southern Great Plain	0.1264	0.1366	0.1377	0.1363
child care expenditures per capita in municipality (Fts in 2005)	318.1707	314.4123	314.3697	314.9296
child care teachers per 3-6 old children in municipality	0.0830	0.0812	0.0814	0.0814
child care places per 3-6 old children in municipality	3.8154	3.2885	3.3191	3.3327
fraction of 6-year olds in child care in municipality	0.2305	0.2231	0.2267	0.2251
no child care in municipality	0.0233	0.0304	0.0302	0.0307

Note: source of data: HNABC (grade 6) 2008-2014, KIR-Stat/DEM 1998-2006, LGT 1998-2005. Relevant sample includes children who were born in the months October to March or March to August. The corresponding equation is (3.3).

Table 3.7: The Effect of Academic Redshirting and Involuntary School Entry Delay - IV Results on 6th/8th/10th-grade Mathematics testscore, with full set of interactions by Gender and Parental Education, Administrative Data, Grades 6/8/10

Y_i: mathematics testscore	***Academic Redshirting***			***Involuntary Delay***		
	x_d : *January 1st*			x_d : *June 1st*		
grades	**6**	**8**	**10**	**6**	**8**	**10**
Panel A: effects for all groups						
D_i	0.2402***	0.2493***	0.1620**	0.2146***	0.1181***	0.0998***
	[0.061]	[0.077]	[0.071]	[0.046]	[0.038]	[0.033]
N	223,851	206,415	219,782	227,127	219,081	213,983
R^2	0.195	0.193	0.223	0.209	0.216	0.236
Panel B: effects for subgroups						
Panel B1: coefficients and standard errors						
D_i	0.2181***	0.2593***	0.1529*	0.1955***	0.0930***	0.0755**
	[0.075]	[0.097]	[0.090]	[0.041]	[0.035]	[0.031]
$boy_i \times D_i$	0.005	0.0044	0.0221	0.001	-0.0006	0.0407*
	[0.041]	[0.051]	[0.049]	[0.029]	[0.027]	[0.022]
$lowp_i \times D_i$	0.0333	-0.0289	0.0206	0.0227	0.0486**	0.0289
	[0.044]	[0.054]	[0.054]	[0.023]	[0.023]	[0.020]
$lowp_i \times boy_i \times D_i$	0.0269	-0.0054	-0.0543	0.0477	0.0204	-0.0232
	[0.058]	[0.069]	[0.067]	[0.040]	[0.036]	[0.032]
N	223,851	206,415	219,782	227,127	219,081	213,983
R^2	0.195	0.193	0.223	0.21	0.217	0.236
Panel B2: estimated effects of D_i						
girl, high parental educ.	0.2181***	0.2593***	0.1529*	0.1955***	0.093***	0.0755**
boy, high parental educ.	0.2231***	0.2637***	0.175***	0.1965***	0.0924*	0.1162***
girl, low parental educ.	0.2514***	0.2304**	0.1735*	0.2182***	0.1416***	0.1044***
boy, low parental educ.	0.2833***	0.2294***	0.1413*	0.2669***	0.1614***	0.1219***
Panel B3: p-values of testing differences of effects of D_i						
girl, high-low parental educ.	0.446	0.595	0.701	0.327	0.031	0.149
boy, high-low parental educ.	0.095	0.431	0.454	0.029	0.018	0.843
high parental educ., boy-girl	0.902	0.931	0.653	0.972	0.981	0.068
low parental educ., boy-girl	0.477	0.985	0.582	0.161	0.513	0.492

Note: this table shows the second-stage result of Y_i on D_i, $boy_i \times D_i$, $lowp_i \times D_i$, $boy_i \times lowp_i \times D_i$ and control variables. D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). Z_i is a binary variable denoting quarter of birth of child i (1: child was born on/after January 1st or June 1st, 0: child was born before January 1st or June 1st). Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Testscores have been standardized to have mean 0 and standard deviation of 1 in each year and grade. Estimation restricted to children who were born in the three-months window around the cutoff dates of January 1st and June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HNABC (grades 6,8,10) 2008-2014, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding equation is (3.2).

Table 3.8: The Effect of Academic Redshirting and Involuntary School Entry Delay - IV Results on 6th/8th/10th-grade Reading testscore, with full set of interactions by Gender and Parental Education, Administrative Data, Grades 6/8/10

Y_i: reading testscore	***Academic Redshirting***			***Involuntary Delay***		
	x_d : *January 1st*			x_d : *June 1st*		
grades	**6**	**8**	**10**	**6**	**8**	**10**
Panel A: effects for all groups						
D_i	0.2307***	0.1975***	0.1736**	0.2456***	0.2119***	0.1649***
	[0.060]	[0.073]	[0.069]	[0.042]	[0.037]	[0.032]
N	223,879	206,433	219,802	227,153	219,125	214,018
R^2	0.261	0.268	0.257	0.272	0.274	0.264
Panel B: effects for subgroups						
Panel B1: coefficients and standard errors						
D_i	0.1873**	0.1742*	0.1757**	0.2218***	0.1816***	0.1334***
	[0.073]	[0.092]	[0.088]	[0.037]	[0.033]	[0.030]
$boy_i \times D_i$	0.0402	0.0704	-0.0016	0.0013	0.0201	0.0524**
	[0.038]	[0.048]	[0.049]	[0.028]	[0.026]	[0.021]
$lowp_i \times D_i$	0.0641	-0.0262	0.0084	0.0196	0.0261	0.0262
	[0.040]	[0.050]	[0.054]	[0.022]	[0.021]	[0.019]
$lowp_i \times boy_i \times D_i$	-0.0254	-0.0263	-0.0217	0.0835**	0.0639*	0.0003
	[0.054]	[0.065]	[0.068]	[0.040]	[0.036]	[0.032]
N	223,879	206,433	219,802	227,153	219,125	214,018
R^2	0.261	0.269	0.257	0.272	0.274	0.264
Panel B2: estimated effects of D_i						
girl, high parental educ.	0.1873**	0.1742*	0.1757**	0.2218***	0.1816***	0.1334***
boy, high parental educ.	0.2275***	0.2446***	0.1741***	0.2231***	0.2017***	0.1858***
girl, low parental educ.	0.2514***	0.148*	0.1841**	0.2414***	0.2077***	0.1596***
boy, low parental educ.	0.2662***	0.1921***	0.1608**	0.3262***	0.2917***	0.2123***
Panel B3: p-values of testing differences of effects of D_i						
girl, high-low parental educ.	0.107	0.600	0.877	0.383	0.221	0.164
boy, high-low parental educ.	0.268	0.218	0.769	0.001	0.002	0.337
high parental educ., boy-girl	0.293	0.143	0.973	0.962	0.438	0.013
low parental educ., boy-girl	0.727	0.397	0.695	0.011	0.004	0.038

Note: this table shows the second-stage result of Y_i on D_i, $boy_i \times D_i$, $lowp_i \times D_i$, $boy_i \times lowp_i \times D_i$ and control variables. D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). Z_i is a binary variable denoting quarter of birth of child i (1: child was born on/after January 1st or June 1st, 0: child was born before January 1st or June 1st). Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Testscores have been standardized to have mean 0 and standard deviation of 1 in each year and grade. Estimation restricted to children who were born in the three-months window around the cutoff dates of January 1st and June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HNABC (grades 6,8,10) 2008-2014, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding equation is (3.2).

Table 3.9: The Effect of Academic Redshirting and Involuntary School Entry Delay - IV Results on the Probability of Repeating a Grade by 6th/10th-grade, with full set of interactions by Gender and Parental Education, Administrative Data, Grades 6/8/10

$Y_i = 1\{\text{having repeated a grade}\}$	***Academic Redshirting***			***Involuntary Delay***		
	x_d : *January 1st*			x_d : *June 1st*		
grades	**6**	**8**	**10**	**6**	**8**	**10**
Panel A: effects for all groups						
D_i	-0.0472***	-0.0407**	-0.0256	-0.0253**	-0.0239***	-0.0375***
	[0.014]	[0.017]	[0.026]	[0.010]	[0.009]	[0.011]
N	223,924	206,499	219,901	227,185	219,173	214,092
R^2	0.089	0.084	0.074	0.099	0.088	0.077
mean of dependent variable:	*.0478*	*.0463*	*.1088*	*.0451*	*.0469*	*.0954*
Panel B: effects for subgroups						
Panel B1: coefficients and standard errors						
D_i	-0.0361**	-0.0355*	-0.02	-0.0172**	-0.0136*	-0.0257***
	[0.016]	[0.020]	[0.031]	[0.008]	[0.007]	[0.010]
$boy_i \times D_i$	-0.0005	-0.0089	-0.0103	0.0002	0.0002	-0.0033
	[0.007]	[0.009]	[0.016]	[0.005]	[0.005]	[0.006]
$lowp_i \times D_i$	-0.0131	0.0087	0.0275	-0.0153***	-0.0207***	-0.0182***
	[0.010]	[0.012]	[0.019]	[0.006]	[0.006]	[0.007]
$lowp_i \times boy_i \times D_i$	-0.0259*	-0.0175	-0.0514**	-0.0059	-0.0064	-0.0172
	[0.015]	[0.018]	[0.026]	[0.011]	[0.010]	[0.012]
N	223,924	206,499	219,901	227,185	219,173	214,092
R^2	0.087	0.084	0.074	0.099	0.088	0.077
Panel B2: estimated effects of D_i						
girl, high parental educ.	-0.0361**	-0.0355*	-0.0200	-0.0172**	-0.0136*	-0.0257***
mean of dependent variable:	*.0135*	*.0143*	*.0585*	*.0123*	*.0137*	*.0499*
boy, high parental educ.	-0.0366***	-0.0444***	-0.0303	-0.0170	-0.0134	-0.0290**
mean of dependent variable:	*.0217*	*.0248*	*.0893*	*.0191*	*.0246*	*.0771*
girl, low parental educ.	-0.0492***	-0.0268	0.0075	-0.0325***	-0.0343***	-0.0439***
mean of dependent variable:	*.0758*	*.0704*	*.1537*	*.0744*	*.0735*	*.139*
boy, low parental educ.	-0.0756***	-0.0532***	-0.0542***	-0.0382**	-0.0405***	-0.0644***
mean of dependent variable:	*.1103*	*.1086*	*.1910*	*.1058*	*.1090*	*.1719*
Panel B3: p-values of testing differences of effects of D_i						
girl, high-low parental educ.	0.207	0.468	0.152	0.006	0.000	0.009
boy, high-low parental educ.	0.000	0.476	0.151	0.020	0.001	0.000
high parental educ., boy-girl	0.935	0.309	0.518	0.967	0.963	0.604
low parental educ., boy-girl	0.059	0.109	0.007	0.600	0.529	0.054

Note: this table shows the second-stage result of Y_i on D_i, $boy_i \times D_i$, $lowp_i \times D_i$, $boy_i \times lowp_i \times D_i$ and control variables. D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). Z_i is a binary variable denoting quarter of birth of child i (1: child was born on/after January 1st or June 1st, 0: child was born before January 1st or June 1st). Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Estimation restricted to children who were born in the three-months window around the cutoff dates of January 1st and June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HNABC (grades 6,8,10) 2008-2014, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding equation is (3.2).

Table 3.10: The Effect of Academic Redshirting and Involuntary School Entry Delay - IV Results on 10th-grade Secondary School Track Choice, with full set of interactions by Gender and Parental Education, Administrative Data, Grade 10

$Y_i = 1\{\text{being in track } j\}$	***Academic Redshirting***			***Involuntary Delay***		
$(j = low, middle, high)$	*January 1st cutoff*			*June 1st cutoff*		
tracks	**low**	**middle**	**high**	**low**	**middle**	**high**
Panel A: effects for all groups						
D_i	-0.0925***	0.0263	0.0662*	-0.0456***	-0.0357*	0.0812***
	[0.029]	[0.040]	[0.036]	[0.012]	[0.019]	[0.017]
N	219,901	219,901	219,901	214,092	214,092	214,092
R^2	0.223	0.039	0.199	0.235	0.044	0.203
mean of dependent variable:	*.184*	*0.416*	*0.40*	*.171*	*.4096*	*.4193*
Panel B: effects for subgroups						
Panel B1: coefficients and standard errors						
D_i	-0.1001***	0.0394	0.0607	-0.0218**	-0.0606***	0.0824***
	[0.035]	[0.052]	[0.049]	[0.011]	[0.018]	[0.016]
$boy_i \times D_i$	0.0173	-0.0226	0.0052	-0.0206***	0.0185	0.0021
	[0.016]	[0.030]	[0.029]	[0.007]	[0.013]	[0.012]
$lowp_i \times D_i$	0.0279	-0.0508	0.0229	-0.0370***	0.0358***	0.0011
	[0.022]	[0.032]	[0.030]	[0.008]	[0.012]	[0.011]
$lowp_i \times boy_i \times D_i$	-0.0627**	0.0930**	-0.0302	-0.0007	0.0159	-0.0152
	[0.030]	[0.039]	[0.035]	[0.014]	[0.019]	[0.016]
N	219,901	219,901	219,901	214,092	214,092	214,092
R^2	0.223	0.039	0.199	0.234	0.043	0.203
Panel B2: estimated effects of D_i						
girl, high parental educ.	-0.1001***	0.0394	0.0607	-0.0218**	-0.0606***	0.0824*
mean of dependent variable:	*.052*	*.346*	*.606*	*.047*	*.332*	*.621*
boy, high parental educ.	-0.0828***	0.0168	0.0659**	-0.0424***	-0.0421*	0.0845*
mean of dependent variable:	*.117*	*.454*	*.429*	*.106*	*.446*	*.448*
girl, low parental educ.	-0.0722*	-0.0114	0.0836*	-0.0588***	-0.0248	0.0835*
mean of dependent variable:	*.295*	*.460*	*.245*	*.283*	*.459*	*.258*
boy, low parental educ.	-0.1176***	0.059	0.0586*	-0.0801***	0.0096	0.0704*
mean of dependent variable:	*.444*	*.428*	*.128*	*.426*	*.433*	*.141*
Panel B3: p-values of testing differences of effects of D_i						
girl, high-low parental educ.	0.207	0.112	0.447	0.000	0.002	0.917
boy, high-low parental educ.	0.098	0.077	0.700	0.002	0.001	0.247
high parental educ., boy-girl	0.291	0.450	0.856	0.003	0.140	0.865
low parental educ., boy-girl	0.123	0.030	0.351	0.116	0.031	0.283

Note: this table shows the second-stage result of Y_i on D_i, $boy_i \times D_i$, $lowp_i \times D_i$, $boy_i \times lowp_i \times D_i$ and control variables. D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). Z_i is a binary variable denoting quarter of birth of child i (1: child was born on/after January 1st or June 1st, 0: child was born before January 1st or June 1st). Secondary school track choice is measured by a binary variable indicating whether the child attends the j-th track at grade 10. Track #1 is the lowest secondary school track, granting a vocational, but no high school degree upon completion. Track #2 is the middle secondary school track, granting both a vocational and a high school degree upon completion. Track #3 is the highest secondary school track, granting a high school degree upon completion. Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Estimation restricted to children who were born in the three-months window around the cutoff dates of January 1st and June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HNABC (grade 10) 2008-2014, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding equation is (3.2).

Table 3.11: The Effect of Academic Redshirting and Involuntary School Entry Delay - IV Results on Mental Health Outcomes at grade 8, with full set of interactions by Gender and Parental Education, Survey Data

$Y_i = 1\{\text{having mental condition } j\}$	***Academic Redshirting***			***Involuntary Delay***		
$(j = pain, anxiety, exhaustion)$	*January 1st cutoff*			*June 1st cutoff*		
	pain	**anxiety**	**exhaustion**	**pain**	**anxiety**	**exhaustion**
Panel A: effects for all groups						
D_i	-0.1443*	-0.2716*	-0.1904	-0.0342	0.0305	-0.0986
	[0.082]	[0.155]	[0.149]	[0.041]	[0.070]	[0.078]
N	6,899	6,899	6,899	6,116	6,116	6,116
R^2				0.016	0.030	0.015
mean of dependent variable:	*0.0524*	*0.1827*	*0.1941*	*0.0519*	*0.1808*	*0.1932*
Panel B: effects for subgroups						
Panel B1: coefficients and standard errors						
D_i	-0.1469	-0.3295	-0.1779	-0.0342	0.0979	-0.0883
	[0.113]	[0.205]	[0.196]	[0.043]	[0.074]	[0.080]
$boy_i \times D_i$	0.0192	0.1055	0.0432	0.0072	-0.1216**	-0.0485
	[0.057]	[0.104]	[0.102]	[0.030]	[0.051]	[0.054]
$lowp_i \times D_i$	-0.0049	0.0775	-0.0483	-0.0033	-0.0685	0.0041
	[0.076]	[0.118]	[0.114]	[0.045]	[0.065]	[0.066]
$lowp_i \times boy_i \times D_i$	-0.0167	-0.1476	-0.0264	-0.0109	0.1004	0.0595
	[0.083]	[0.141]	[0.137]	[0.056]	[0.088]	[0.094]
N	6,899	6,899	6,899	6,116	6,116	6,116
R^2				0.015	0.029	0.014
Panel B2: estimated effects of D_i						
girl, high parental educ.	-0.1469	-0.3295	-0.1779	-0.0342	0.0979	-0.0883
mean of dependent variable:	*0.0642*	*0.1832*	*0.2059*	*0.0675*	*0.1980*	*0.2160*
boy, high parental educ.	-0.1277	-0.224	-0.1347	-0.027	-0.0237	-0.1368
mean of dependent variable:	*0.0213*	*0.1456*	*0.1910*	*0.0848*	*0.2378*	*0.2125*
girl, low parental educ.	-0.1518	-0.252	-0.2262	-0.0375	0.0294	-0.0842
mean of dependent variable:	*0.0855*	*0.2328*	*0.2108*	*0.02501*	*0.1383*	*0.1915*
boy, low parental educ.	-0.1493*	-0.2941**	-0.2094**	-0.0412	0.0082	-0.0732
mean of dependent variable:	*0.0394*	*0.1660*	*0.1671*	*0.0345*	*0.1501*	*0.1658*
Panel B3: p-values of testing differences of effects of D_i						
girl, high-low parental educ.	0.948	0.513	0.672	0.941	0.293	0.951
boy, high-low parental educ.	0.592	0.382	0.358	0.672	0.591	0.353
high parental educ., boy-girl	0.736	0.310	0.672	0.811	0.018	0.373
low parental educ., boy-girl	0.970	0.707	0.874	0.938	0.767	0.887

Note: this table shows the second-stage result of Y_i on D_i, $boy_i \times D_i$, $lowp_i \times D_i$, $boy_i \times lowp_i \times D_i$ and control variables. D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). Z_i is a binary variable denoting quarter of birth of child i (1: child was born on/after January 1st or June 1st, 0: child was born before January 1st or June 1st). Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Estimation restricted to children who were born in the three-months window around the cutoff dates of January 1st and June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HLCS, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding equation is (3.2).

Table 3.12: First-stage Results on School Entry Delay (Starting School at 7) and the Change in Relative Rank in Class, Administrative Data, Grades 6

	with sorting schools		
	D_i	$RR_i(\%)$	Q_i
$1\{X_i \geq x_d\}$	0.185***	16.22***	0.0681***
	(0.00711)	(0.536)	(0.0103)
$peers\_summer_i$	0.000258	0.107***	0.00302***
	(0.000173)	(0.0113)	(0.000181)
$1\{X_i \geq x_d\} \times peers\_summer_i$	-0.000143	-0.211***	-0.00199***
	(0.000181)	(0.0146)	(0.000281)
N	223,422	223,422	223,422
R^2	0.276	0.080	0.078
joint F-statistic	*476.73*	*356.83*	*111.03*
Hansen J-statistic p-value		*0.2327*	*0.3677*
Cragg-Donald Wald F-statistic:		*409.19*	*194.49*
Kleibergen-Paap rk Wald F-statistic:		*242.93*	*110.61*

	without sorting schools A			*without sorting schools B*		
	D_i	$RR_i(\%)$	Q_i	D_i	$RR_i(\%)$	Q_i
$1\{X_i \geq x_d\}$	0.175***	15.97***	0.0603***	0.182***	15.93***	0.0582***
	(0.0122)	(0.923)	(0.0185)	(0.00792)	(0.599)	(0.0115)
$peers\_summer_i$	0.000277	0.108***	0.00320***	0.000298	0.111***	0.00313***
	(0.000291)	(0.0192)	(0.000303)	(0.000195)	(0.0128)	(0.000204)
$1\{X_i \geq x_d\} \times peers\_summer_i$	-0.000122	-0.219***	-0.00208***	-0.000196	-0.211***	-0.00188***
	(0.000303)	(0.0245)	(0.000508)	(0.000203)	(0.0164)	(0.000315)
N	78,810	78,810	78,810	188,116	188,116	188,116
R^2	0.268	0.078	0.091	0.268	0.076	0.082
joint F-statistic	*164.07*	*118.42*	*44.64*	*375.53*	*278.23*	*96.24*
Hansen J-statistic p-value		*0.8328*	*0.5693*		*0.4837*	*0.2000*
Cragg-Donald Wald F-statistic:		*163.65*	*73.01*		*325.04*	*171.30*
Kleibergen-Paap Wald F-statistic:		*93.70*	*42.27*		*192.81*	*95.36*

Note: this table shows the first-stage results of D_i and RR_i/Q_i, where D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). RR_i is the continuous measure of the child's relative age rank in the class, measured in percentiles, so 100 means the child is the oldest in class. Q_i is the binary measure of the child's relative age rank in the class (1: child is in the highest quartile of the age distribution in class). $1\{X_i \geq x_d\}$ is a binary variable denoting quarter of birth of child i (1: child was born on/after 1st June 1st, 0: child was born before June 1st). $peers\_summer_i$ denotes the fraction of summer-born children in child i's class. Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Estimation restricted to children who were born in the three-months window around the cutoff date June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HNABC (grade 6) 2008-2014, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding 2nd-stage equation is (3.4). Non-sorting schools "A" are the ones that do not sort children systematically into classes based on prior student achievement, while non-sorting schools "B" are the ones that do not sort children systematically into classes based on month of birth.

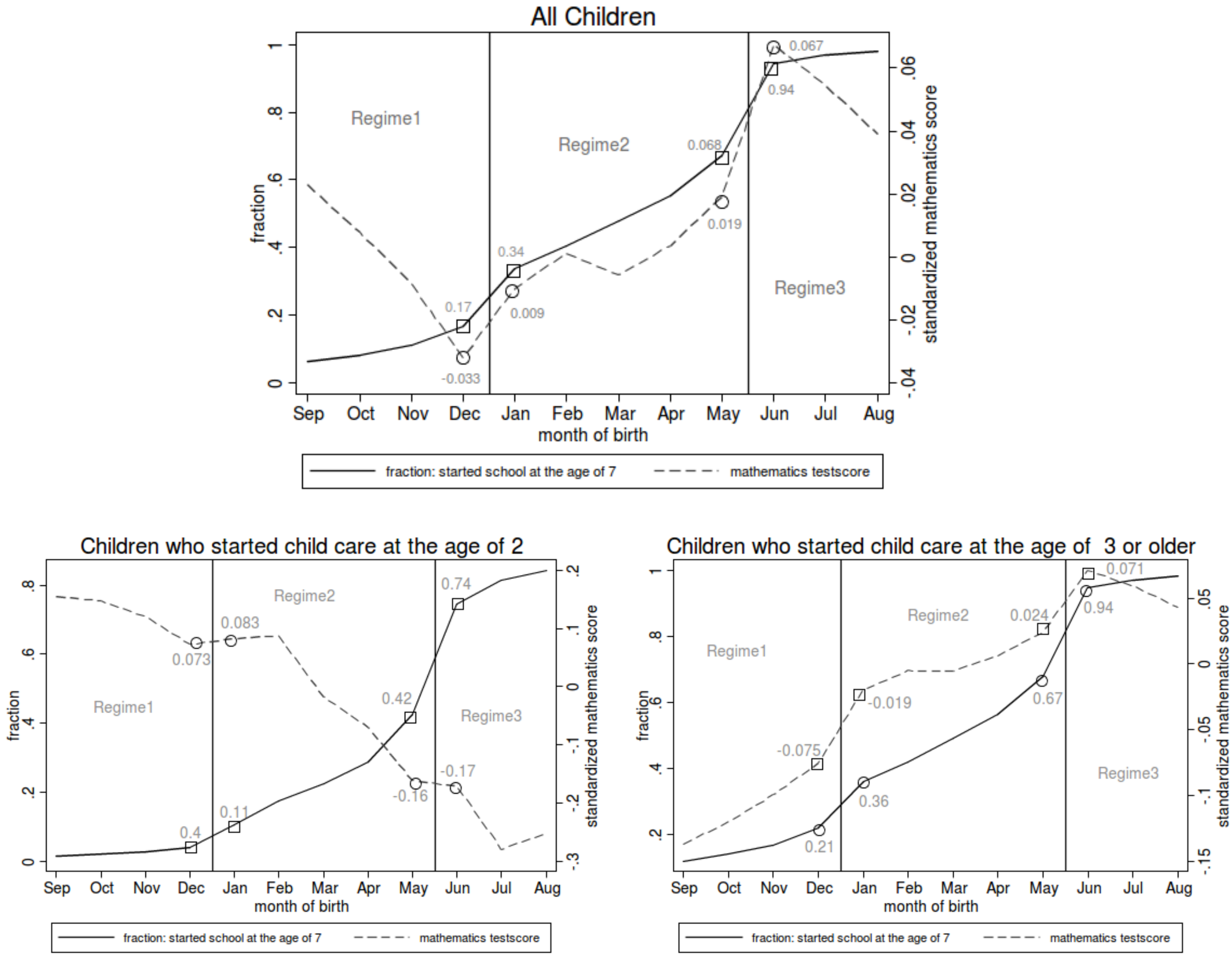
Table 3.13: OLS and IV Estimates of the Effect of Involuntary School Entry Delay (Starting School at 7) and the Relative Rank in Class on Mathematics Score, Administrative Data Grade 6

Panel A: including sorting schools									
	OLS	OLS	OLS	OLS	OLS	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
D_i	-0.0462***			0.0710***	-0.0248***	0.2150***		0.2700**	0.2071***
	[0.005]			[0.011]	[0.006]	[0.046]		[0.121]	[0.047]
RR_i		-0.0013***		-0.0022***			0.0038***	-0.0009	
		[0.000]		[0.000]			[0.001]	[0.002]	
Q_i			-0.0576***		-0.0496***				0.1312
			[0.005]		[0.005]				[0.120]
N	223,364	223,364	223,364	223,364	223,364	223,364	223,364	223,364	223,364
R^2	0.226	0.227	0.226	0.227	0.226	0.214	0.207	0.215	0.203
Panel B: without sorting schools A									
	OLS	OLS	OLS	OLS	OLS	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
D_i	-0.0564***			0.1140***	-0.0249***	0.2730***		0.1771	0.2725***
	[0.008]			[0.018]	[0.009]	[0.078]		[0.185]	[0.078]
RR_i		-0.0017***		-0.0031***			0.0048***	0.0017	
		[0.000]		[0.000]			[0.001]	[0.003]	
Q_i			-0.0802***		-0.0722***				-0.0143
			[0.007]		[0.008]				[0.180]
N	78,797	78,797	78,797	78,797	78,797	78,797	78,797	78,797	78,797
R^2	0.240	0.241	0.241	0.242	0.241	0.222	0.208	0.218	0.223
Panel C: without sorting schools B									
	OLS	OLS	OLS	OLS	OLS	IV	IV	IV	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
D_i	-0.0497***			0.0647***	-0.0297***	0.2053***		0.3506***	0.2036***
	[0.006]			[0.012]	[0.006]	[0.052]		[0.133]	[0.052]
RR_i		-0.0013***		-0.0021***			0.0037***	-0.0025	
		[0.000]		[0.000]			[0.001]	[0.002]	
Q_i			-0.0561***		-0.0468***				0.0541
			[0.005]		[0.006]				[0.126]
N	188,076	188,076	188,076	188,076	188,076	188,076	188,076	188,076	188,076
R^2	0.229	0.230	0.229	0.230	0.229	0.218	0.211	0.220	0.214

Note: this table shows the estimated OLS and IV effects of D_i and RR_i on standardized mathematics testscore in grade 6, where D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). RR_i is the continuous measure of the child's relative age rank in the class, measured in percentiles, so 100 means the child is the oldest in class. Q_i is the binary measure of the child's relative age rank in the class (1: child is in the highest quartile of the age distribution in class). The first-stage results corresponding to the IV strategy can be seen in Table3.12. Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Estimation restricted to children who were born in the three-months window around the cutoff date June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HNABC (grade 6) 2008-2014, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding 2nd-stage equation is (3.4). Non-sorting schools "A" are the ones that do not sort children systematically into classes based on prior student achievement, while non-sorting schools "B" are the ones that do not sort children systematically into classes based on month of birth.

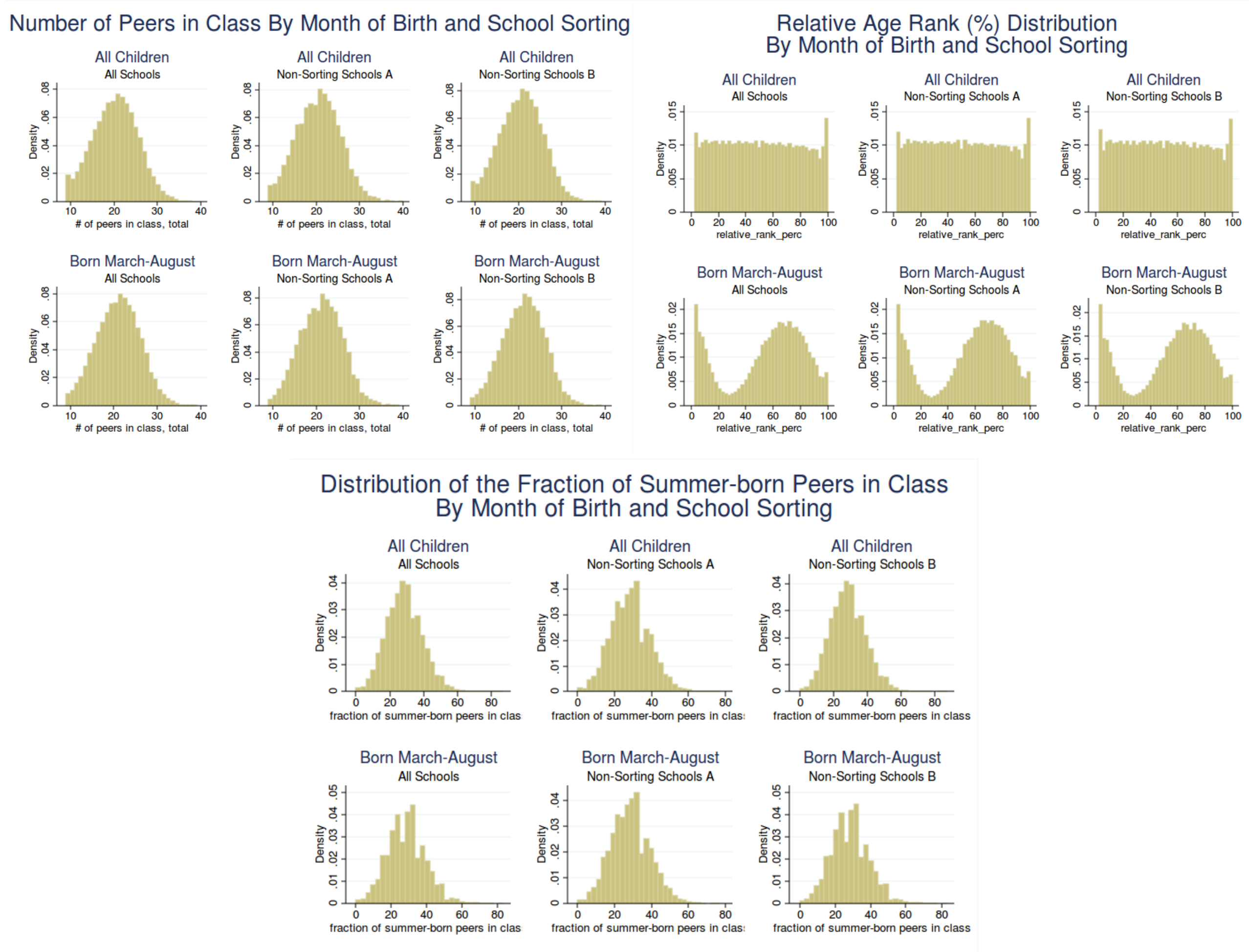
3.8 Figures

Figure 3.1: Fraction of Children who Started School at the Age of 7 and Average Mathematics Testscores, by Month of Birth and child care Starting Age



Note: source of data: HNABC (grade 6) 2008-2014. See Chapter 4 Data and Measurement for details. Average testscores, shown on the right y-axis, standardized to have 0 mean and 1 standard deviation.

Figure 3.2: Histogram of Number of Peers in Class, of Relative Age Rank (%) in Class, and of Fraction of Summer-born Peers in Class; Administrative Data Grade 6



Note: source of data: HNABC (grade 6) 2008-2014. Non-sorting schools “A” are the ones that do not sort children systematically into classes based on prior student achievement, while non-sorting schools “B” are the ones that do not sort children systematically into classes based on month of birth.

Chapter 4

Do Health Insurers Innovate? Evidence from the Anatomy of Physician Payments

4.1 Introduction

Health insurers have a powerful ability to shape the efficiency of health care delivery. Insurers straddle the relationship between patients and medical providers, and enter into contracts with both sides of the market. Consumers or employers purchase insurance plans whose copayments and deductibles influence subsequent demand for care. However, contracts with physicians and hospitals govern how providers will be compensated for treating insured patients, and hence the caregivers' financial incentives.

The literature on optimal consumer cost-sharing is long and well-developed (Feldstein, 1973; Besley, 1988). Only recently, however, has an empirical literature begun to explore how private insurers set copayments in practice. Einav et al. (2016) show that private insurers provide more risk-protection for drugs subject to less moral hazard. They contrast this with public insurance plans, which offer relatively uniform coverage with regards to cost-sharing. Starc and Town (2015) show that insurers responsible for patients' non-pharmaceutical spending provide more generous coverage for drugs that can keep people out of the hospital.

We investigate whether insurers apply a similar logic to the payments they negotiate with providers. To what extent do private insurance carriers adopt Medicare's cost-based approach to physician payment? Looking at the flip side of this question, we analyze how much private insurers customize their physician reimbursements relative to Medicare's industry standard.

Despite recent high-level changes in the U.S. health insurance market, the incentive structure through which physicians are paid remains predominantly "fee for service."[109] A physician's income depends on the quantity and intensity of the treatment she provides—even when part of a larger managed care plan or "accountable care organization" (Zuvekas and Cohen, 2016).[110] A growing body of evidence finds that these high-powered incentives help drive the level and composition of medical spending (McClellan, 2011).

The structure of physician payments is a potentially powerful tool for insurers to encourage more efficient care. Relatively little is known, however, regarding the extent to which private insurers customize their fee-for-service payments for this purpose. Clemens and Gottlieb (2017) find that

[109]Data from the 2004-2005 Community Tracking Study (CSHSC 2006, 46) show that 52 percent of physicians earn zero revenue from capitated contracts, and 79 percent earn less than a quarter of their revenue from such contracts.

[110]Among those same physicians referenced in footnote 109, who earn little from capitated contracts, 65 percent earn more than one quarter of their revenue from managed care (CSHSC 2006, 47). CSHSC (1999) reports similar estimates from 1996-1997.

private payments rise and decline quite strongly with Medicare's payments, which raises the question of whether private insurers' payments are meaningfully different from Medicare's rate schedule. In this chapter, we shed light on this question using detailed data on physician payments.

Medicare compensates physicians and outpatient providers through a detailed fee-for-service pricing system. Physicians submit bills for each instance in which they provide any of 13,000 recognized services. The system assigns each service a certain number of Relative Value Units, determining payment. These relative values aim to measure average cost but not medical value. This procurement model thus has little capacity to steer treatment towards effective—let alone cost-effective—care. It has particular difficulty managing the use of capital-intensive diagnostic imaging services, for which average cost payments exceed providers' marginal costs—as they must in order to facilitate entry.

Private reimbursement arrangements are less transparent than Medicare's. To peer into the black box of these business-to-business contracts, we begin by developing a cross-sectional method for systematically assessing whether payments are benchmarked to Medicare's rate structure. Our first approach involves a classification algorithm motivated by the bunching literature.[111] Using the outpatient claims data of Blue Cross Blue Shield of Texas (BCBS-TX), we begin by computing the ratio of each private payment to the applicable Medicare payment. Among the payments to individual physician groups, the distribution of these ratios reveals spikes that are indicative of exceptionally common markups. We use these exceptionally common mark-ups to identify which payments are likely benchmarked to Medicare's relative rate structure.

We complement our cross-sectional method with an analysis of updates to Medicare's structure of relative payments. If the Medicare links we identify are accurate, then payments for Medicare-benchmarked services should update mechanically when Medicare's schedule of "relative value units" is revised. We are able to assess this mechanical pass through at a high frequency by applying institutional knowledge of the exact dates on which BCBS-TX implements Medicare's updates to the relative value scale.[112] We find that the payments associated with 55 percent of in-network, outpatient spending (and around three quarters of services) are linked to Medicare. These estimates are quite similar to the estimates we obtain using our cross-sectional bunching approach.

We continue our analysis with an effort to understand the circumstances under which payments are more and less likely to be benchmarked to Medicare's relative rate structure. Deviations from benchmarking exhibit several distinctive patterns. Looking across physician groups, payments to relatively large firms are less tightly benchmarked to Medicare than payments to small firms. Payments for only ten percent of services provided by the smallest firms, representing 20 percent of their spending, deviate from Medicare's relative values. The same is true of 40 percent of services—and two-thirds of spending—from firms with total billing exceeding $1 million per year.

Looking across service categories, payments are more likely to deviate from Medicare's relative values for capital-intensive services, like diagnostic imaging, than for labor-intensive services like standard office visits. Payments for roughly 45 percent of imaging services, but only 15 percent of

[111] Our setting differs from standard bunching applications in that the bunching we observe is not driven by kinks or notches in budget sets (Kleven, 2016). Instead, it results from clustering around reference points.

[112] The relatively high frequency at which we can conduct our analysis allows us to limit if not eliminate the relevance of potential confounders including active contract renegotiations and payment changes connected to substantive technological advances.

evaluation and management services, deviate from Medicare's menu. Within imaging, Medicare distinguishes between two types of services: a capital-intensive component for taking the image and a labor-intensive component for interpreting the image. Medicare explicitly amortizes the fixed cost of the imaging equipment into the former. We find that private insurers' payments for interpretation are far less likely to deviate from Medicare rates than are its payments for taking the image itself. The directions of these deviations reveal that the adjustments narrow likely gaps between marginal costs and Medicare's average-cost payments. We find that payments for labor-intensive services tend to be adjusted up while payments for capital-intensive services are adjusted down.

We continue in section 4.2 by presenting institutional background on price setting in U.S. physician markets and potential explanations for the phenomenon we examine. Section 4.3 introduces our claims data. In section 4.4 we present the empirical approaches that estimate the share of payments benchmarked to Medicare using updates to Medicare's relative prices. Our main empirical results are in section 4.5. In section 4.6 we show that the deviations from these Medicare-linked rates narrow gaps between prices and marginal costs. Section 4.7 concludes.

4.2 Medical Pricing Institutions

Public and private payments for health care services are set through very different mechanisms. Medicare reimbursements are set based on administrative estimates of the resource costs of providing care, which we describe in section 4.2.1. For patients with private health insurance, providers' reimbursements are determined through negotiations between the insurers and providers, which we describe in section 4.2.2. Section 4.2.3 offers economic rationales for a potential link between reimbursement rates across these two segments of the market.

4.2.1 Medicare Price Determination[113]

In 1992, Congress established a system of centrally administered prices to reimburse physicians and other outpatient providers. This Resource-Based Relative Value Scale (RBRVS) is a national fee schedule that assigns a fixed number of Relative Value Units (RVUs) to each of 13,000 distinct health care services. Legislation specifies that the RVUs for service j are supposed to measure the resources required to provide that service. Since the costs of intermediate inputs differ across the country, RBRVS incorporates local price adjustments, called the Geographic Adjustment Factor (GAF), to compensate providers for these differences. The payment for service j to a provider in geographic region i is approximately:

$$\begin{aligned}\text{Reimbursement rate}_{i,j,t} = {} & \text{Conversion Factor}_t \times \text{Geographic Adjustment Factor}_{i,t} \\ & \times \text{Relative Value Units}_{j,t}. \end{aligned} \tag{4.1}$$

[113] This section draws from Clemens and Gottlieb (2014).

The "reimbursement rate," a term we use interchangeably with "price," is the amount Medicare pays for this service. The Conversion Factor (CF) is a national scaling factor, usually updated annually.

The variation in payments is mainly driven by the number of RVUs assigned to a service. This assignment is constant across areas while varying across services. Medicare regularly updates the RVUs assigned to each service, primarily based on input from the American Medical Association, using the formal federal rule-making process. These updates are intended to account for technological and regulatory changes that alter a service's resource intensity. We exploit these changes in the empirical strategy we introduce in section 4.4.

4.2.2 Private Sector Price Setting

U.S. private sector health care prices are set through negotiations between providers and private insurers.[114] The details of these negotiations are not transparent, and our limited knowledge about private sector prices comes from claims data that reveal the reimbursements paid once care is provided.[115] A common feature of physician contracts, central to both our theoretical and empirical analyses, is a form of benchmarking to Medicare.

Practitioners emphasize that Medicare's administrative pricing menu features prominently in private insurers' contracts. Newsletters that insurers distribute to participating providers frequently draw explicit links between Medicare's maximum allowable charges and the insurer's fee schedule. For example, reimbursement rates might be linked to Medicare by default unless the contract specifies otherwise. But the relative value scale does not determine an absolute price level. As in Medicare, computing private reimbursements requires multiplying RVUs by a dollar scaling factor. Providers and insurers can simplify contracting by negotiating over these constant markups, but sometimes haggle over reimbursements for specific services or bundles (Gesme and Wiseman, 2010; Mertz, 2004).

Our empirical work will examine specifically when and why this benchmarking occurs in practice. We measure how often exceptions apply, and whether they systematically arise in cases when we would expect the cost of Medicare's inefficiencies to be particularly large.

4.2.3 Potential Rationales

Why might contracts between physicians and private insurers use Medicare's relative rate structure as a benchmark? We consider several, broadly complementary explanations. A first explanation is that benchmarking against Medicare's relative rate structure enables insurers and physicians to greatly simplifying their contracts. A uniformly benchmarked contract requires negotiating over a single parameter, namely the mark-up; alternative contract structures could require negotiating payments for hundreds if not thousands of distinct billing codes. Medicare's payment model may serve this

[114] In rare exceptions, such as in Maryland, the state government determines all hospital payment rates.

[115] A growing literature finds that physician concentration significantly affects this bargaining process. Payments are higher in markets where physicians are more concentrated (Dunn and Shapiro, 2014; Baker et al., 2014; Kleiner et al., 2015; Clemens and Gottlieb, 2017).

purpose because it is an "industry standard" with which all parties are familiar. In Appendix C.1, we present a formal model of this idea, which generates several empirical predictions that are consistent with what we find in our empirical analysis.

A second, strongly complementary explanation speaks to the Medicare menu's information content. By design, Medicare's payment model contains substantial information regarding the relative costs of providing physicians' services. If average cost reimbursement is, more or less, what insurers desire to implement, Medicare's payment model provides a natural "information standard" for private insurers to adopt. That is, Medicare's relative cost estimates can be interpreted as a public good. Although they may fail to reflect variations in local cost structures, the expense to insurers of independently calculating these costs may be quite high.

A third possibility is that providing care for Medicare beneficiaries represents physicians' primary outside option when they negotiate with private insurers (Clemens and Gottlieb, 2017). Because Medicare accounts for a large share of the market, its payments inevitably loom large in insurer-physician negotiations. Benchmarking private payments to Medicare's payments may be a straightforward way for contracts to acknowledge and mechanically adjust in response to that option.

A fourth possibility emphasizes insurance regulations. Regulations require insurers to ensure access to "medically necessary" services. Benchmarking payments to Medicare's rate structure may be the easiest approach to satisfying this requirement. Private payments are almost universally "marked-up" rather than "marked-down" relative to Medicare's rates. Consequently, this payment structure ensures that private insurers are paying sufficiently high rates to generate at least as much care access as Medicare beneficiaries enjoy.

4.3 Medical Pricing Data

Our main analysis considers firm-to-firm pricing in the context of medical claims processed by one large insurer, Blue Cross Blue Shield of Texas (BCBS). Our main database covers the universe of BCBS's payments for outpatient care in 2010; we expand our sample to cover 2008–2011 for one analysis.[116] For each claim, the database provides information on the service provided, location, physician, physician group, and BCBS's payment to that group. We restrict this universe along several dimensions. The full 2010 dataset contains 57,613,494 claim lines and $4.29 billion in spending, which we restrict along several dimensions. We clean the data as described in Appendix C.2.1, which initially leaves us with 44,055,829 service lines and $2.63 billion of spending. This initial cut eliminates payments made to out-of-network physicians, who have not reached a negotiated agreement with BCBS on reimbursement rates. We will subsequently examine this segment of the data separately.

In order for private insurers to benchmark prices to Medicare, at a minimum they would need to use Medicare's billing codes. We thus merge the remaining claims with Medicare billing codes, which provides an upper bound on the potential benchmarking. This merge only loses notable portions of one

[116]Our empirical results for other years are very similar to those for 2010; we focus on this one year for brevity and show other years' results in the appendix.

broad spending category, namely laboratory tests, for which both Medicare and BCBS frequently base payments on non-standard codes. We retain over 97 percent of claims for evaluation and management, diagnostic imaging, and surgical services. The final analysis sample includes 3,681 unique HCPCS codes, which comprise 23,933,577 service lines and $2.05 billion of spending.[117]

The claims data further allow us to describe the provider groups serving BCBS beneficiaries, at least in terms of the care they provide to that sample. To enable our subsequent investigation of heterogeneity in Medicare benchmarking, we measure the total value of the care each group provides to BCBS patients in a given year. Our final dataset includes care provided by over 80,000 physician groups, identified by their billing identification number.[118] 15,000 of these groups bill more than $10,000 annually, and account for 97 percent of BCBS spending. Table 4.1 presents summary statistics on the physician groups in our final sample.

4.4 Empirical Approach

Our first empirical goal is to estimate how often private reimbursement rates rely directly on Medicare's. We begin by presenting visually striking evidence of bunching in the ratios of physician groups' payments relative to Medicare's payments. We formalize this visual evidence and then present an empirical approach for exploiting policy-driven changes to Medicare's Relative Value Units (RVUs). Finally, we discuss the complementarity of these approaches.

4.4.1 Measuring Medicare Benchmarking with Bunching

To measure the relationship between private and Medicare pricing in the cross-section, we exploit a straightforward insight. For each claim, we divide the private payment by the corresponding Medicare price, yielding the private markup applying to that payment. When the markups for many payments to a given physician group exhibit bunching at a common level, we infer that these payments reflect a contractual link to Medicare's relative rates.

We begin by simply dividing BCBS's payment to group g for service j at time t ($P_{g,j,t}$) by Medicare's allocation of RVUs. This defines an "Implied Conversion Factor" (ICF) as:

$$ICF_{g,j,t} = \frac{P_{g,j,t}}{RVU_{j,t}}. \tag{4.2}$$

An ICF is defined for every claim. But simply computing an ICF does not tell us whether any given claim was actually priced according to Medicare's RVUs. To gauge the relevance of such pricing schemes, we ask how often a particular group's payments reflect *the same* ICF. Figure 4.1 provides concrete illustrations. Panel A shows payment rates for the services provided regularly by a single

[117] Appendix Table C.1 shows the exact data loss resulting from each step of cleaning. The key conclusion from this table is that, once we restrict ourselves to the relevant universe of data, additional losses from merging in Medicare codes and eliminating infrequent codes are not substantial.

[118] This is analogous to the commonly used tax ID number in Medicare claims data, but our version is anonymized.

physician group in the 2010 BCBS claims data.[119] Each circle on the graph is a unique payment amount for a unique service code. That is, if the group received two unique payment values for a standard office visit (HCPCS code 99213), say \$45 and \$51, those two amounts would show up as separate circles. The Blue Cross payment amount is on the y-axis and the Medicare payment for the service is on the x-axis, both shown on log scales. Taking logs of equation (4.2) reveals that Medicare-linked pricing implies a 1 for 1 relationship between the log Medicare payment and the log private reimbursement.[120] So we draw a solid line in Panel A with a slope of 1, coinciding with the group's most common ICF.[121]

Panel A shows the data from a mid-sized group (billing BCBS between \$200,000 and \$1 million in 2010) for which a single ICF dominates the payment picture. The most natural interpretation of this graph is that those services on the solid line are priced according to Medicare's fee schedule with a common ICF, while the remaining services are priced separately. Several of the circles below the solid line plausibly involve instances of a less common, but still contractually specified, ICF for this group. A conservative estimate would view these and other circles off the solid line as deviations from Medicare-linked pricing.

Panel B shows the full distribution of this group's markups over Medicare rates. To calculate these markups, we simply divide the y-value of each dot in Panel A by its x-value.[122] Panel B shows a clear spike in the distribution of these ratios at around 1.4, indicating that most claims were paid based on a 40 percent markup over Medicare. This spike includes all of the services along the red line in Panel A. Other scattered values reflect the deviations away from that line.

Panels C through F show graphs constructed analogously, but for two larger groups that provide more unique services at more distinct prices. The group shown in Panels C and D exhibits two clear spikes in the ICF frequency distribution, with a smattering of other values. The one shown in Panels E and F has a range of ICFs, none of which visually dominates the payment picture. These plots indicate a remarkably complicated contract with BCBS.

Estimating the pervasiveness of "common" ICFs requires a definition of "common." The estimates presented in section 4.6 thus explore sensitivity to the threshold we impose for the regularity with which an ICF must appear in a group's payments. This further requires an assumption on our rounding of the ratio of private to public payments. We explore sensitivity to the choice of rounding as well.

After defining and computing the ICFs, we run descriptive regressions to understand how they

[119] The figures exclude any code-by-payment combination that appears ten times or fewer in the data for the relevant physician group. The more systematic analysis presented below has no such exclusion. Throughout this analysis, we restrict to data from the period before BCBS implemented the RVU updates (January 1—June 30, 2010). This way our calculations are not confounded by RVU changes.

[120] Rearranging (4.2) and then taking logs yields:

$$\ln(P_{g,j,t}) = \ln(ICF_{g,j,t}) + \ln(RVU_{j,t}), \tag{4.3}$$

which has an implied coefficient of 1 on $\ln(RVU_{j,t})$.

[121] As equation (4.3) shows, the y-intercept (in logs) is the log of the ICF.

[122] This distribution has the same sample restrictions as in Panel A; see footnote 119 for details. Note that each observation from Panel A has equal weight in the distribution in Panel B, so the distributions in Panels B, D, and F are not weighted to reflect the frequency with which we observe each markup. A weighted version would increase the relative heights of the highest bars, since the common ICFs are, by definition, more common than other markups.

vary. We estimate models of the form

$$\ln(ICF_{g,j}) = \mathbf{X_{g,j}}\gamma + e_{g,j} \tag{4.4}$$

where $\mathbf{X_{g,j}}$ contains characteristics of the physician group or local market, such as firm size or concentration. We measure firm size as log total billings to the insurer. We compute firm market share within a local health care market (hospital service area) and specific service, and we measure the degree of concentration across all physician practices within that market (using the HHI at the service-by-area level).

4.4.2 Framework for Analyzing Benchmarking Using RVU Updates

We next develop an estimation framework based on changes in Medicare's relative value scale. A committee of the American Medical Association, composed of representatives of various physician specialties, recommends RVU updates to Medicare (Government Accountability Office, 2015). These updates come in two main forms: reassessments of the resources required to provide a single service, and revisions to part of the underlying methodology. For example, a revision to the method for computing physician effort can change the weights assigned to many service codes. At least one broad update of this sort appears to occur annually over the period we study, as do hundreds of larger service-specific reassessments.

The vast majority of updates to Medicare payments go into effect on January 1 each year. But even when relying on these rates, private insurers have a choice about whether and when to shift from one year's relative value scale to the next year's (Borges, 2003). BCBS informs its providers of the date on which such updates go into effect through its provider newsletter, the *Blue Review*. During our sample, the newsletter announced updates taking place on July 1, 2008, on August 15, 2009, on July 1, 2010, and on September 1, 2011 (BCBS 2008; 2009; 2010; 2011). In all four years, the standard deviations of RVU changes are around 7 percent, generating substantial pricing variation for us to exploit.

Figure 4.2 Panel A shows one example of how these changes impact physician payments in our BCBS data. This graph shows average log payments by day for the most commonly billed service code, a standard office visit with an established patient (code 99213). The average log payment jumps distinctively on July 1, 2010, the day on which BCBS implemented the 2010 relative values. Medicare's log RVUs for this service rose by 0.068 between the 2009 and 2010 fee schedules. BCBS's average log payment rose by just under 0.05. Appendix Figure C.1 shows further examples. To study examples of this sort systematically, we next develop a method for using high frequency RVU changes to infer the share of private reimbursements linked to Medicare.

Our empirical method exploits the institutional details we documented in section 4.2 about how Medicare benchmarking works in practice. When a payment $P_{g,j,t}$ is linked to Medicare's relative values, we can write

$$P_{g,j,t} = \varphi_{g,t} \cdot RVU_{j,t}$$

$$\text{or, in logs,} \qquad \ln(P_{g,j,t}) = \ln(\varphi_{g,t}) + 1 \cdot \ln(RVU_{j,t}), \tag{4.5}$$

where $\varphi_{g,t}$ is the Implied Conversion Factor (ICF) from section 4.1. Equation (4.5) describes a linear relationship between log private insurance payments and log RVUs for a service. It predicts that empirical estimates of this relationship would find a coefficient of 1 on log RVUs. If the markup φ is a constant, it will be reflected in the constant term. If it varies across physician groups, then group fixed effects capture $\ln(\varphi_g)$. If it changes over groups and across time, then group-by-time fixed effects serve the same role.

The institutional details, plus the occasional deviations observed in Figure 4.1, suggest that payments may alternatively be negotiated without reference to RVUs. In this case, we denote the payment by

$$P_{g,j,t} = \rho_{g,j,t} \qquad \Longrightarrow \qquad \ln(P_{g,j,t}) = \ln(\rho_{g,j,t}), \tag{4.6}$$

with no role for $\varphi_{g,t}$ or $RVU_{j,t}$.

When RVUs change, equations (4.5) and (4.6) show how private reimbursements will adjust. Consider two time periods, across which Medicare shifts payments by $\Delta \ln(RVU_{j,t})$. Let $\varepsilon_{g,j,t} = \Delta \ln(\rho_{g,j,t})$ be any change in the alternative non-benchmarked payment. We can now write both types of prices in terms of service fixed effects and changes as follows. For Medicare-linked services, we have:

$$\ln(P_{g,j,t}) = \phi_j 1_j + \phi_g 1_g + \phi_{g,j} 1_g \cdot 1_j + \Delta \ln(RVU_{j,t}) \cdot 1_{\{t=\text{post}\}}. \tag{4.7}$$

For services not linked to Medicare, we have:

$$\ln(P_{g,j,t}) = \phi_j 1_j + \phi_g 1_g + \phi_{g,j} 1_g \cdot 1_j + \varepsilon_{g,j,t} \cdot 1_{\{t=\text{post}\}}. \tag{4.8}$$

In these equations, $1_{\{t=\text{post}\}}$ is an indicator for the second time period. In both types of price setting, the fixed effects capture baseline payments to group g for service j in the first period, while the interaction with $1_{\{t=\text{post}\}}$ captures the change between the two periods.

The linearity of equations (4.7) and (4.8) implies a simple way to measure how many services are linked to Medicare. Equation (4.7) says that a linear regression of log private payments on changes in log Medicare RVUs, for services with prices linked to Medicare, should yield a coefficient of 1 after controlling for appropriate fixed effects. Equation (4.8) shows that the same regression should yield a coefficient of 0 for services not priced based on Medicare, as long as the non-Medicare payment changes ($\varepsilon_{g,j,t}$) are uncorrelated with RVU updates.

More generally, suppose that both types of payments exist, and specifically that a constant share σ of payments are benchmarked to Medicare prices, while $1 - \sigma$ are set independently. (We will subsequently allow for heterogeneity.) The average of log reimbursements is then given by a weighted average of equations (4.7) and (4.8), and the coefficient on log RVU updates can reveal the linked share σ:

$$\ln(P_{g,j,t}) = \phi_j 1_j + \phi_g 1_g + \phi_{g,j} 1_g \cdot 1_j + \sigma \cdot \Delta \ln(RVU_{j,t}) \cdot 1_{\{t=\text{post}\}} + \eta_{g,j,t}, \tag{4.9}$$

where we define $\eta_{g,j,t} = (1-\sigma) \cdot \varepsilon_{g,j,t} \cdot 1_{\{t=\text{post}\}}$. Equation (4.9) suggests that, in a linear regression with appropriate fixed effects, we can infer the Medicare-linked share from the coefficient on log RVU changes. This motivates our baseline specification for estimating σ. We use data at the level of individual claims, indexed by c, to estimate:

$$\ln(P_{c,g,j,t}) = \beta \Delta \ln(RVU_j) \cdot 1_{\{t=\text{post}\}} + \phi_t 1_{\{t=\text{post}\}} + \phi_j 1_j + \phi_g 1_g + \phi_{g,j} 1_g \cdot 1_j + \eta_{c,g,j,t}. \tag{4.10}$$

This is just a claims-level version of equation (4.9) where $\hat{\beta}$ estimates the share of payments based on Medicare rates. It adds a time period fixed effect $1_{\{t=\text{post}\}}$ in case private payments shift broadly across the two time periods. This parametric difference-in-differences specification also incorporates full sets of group (1_g), service (1_j), and group-by-service ($1_g \cdot 1_j$) effects to account for all time-invariant group- and service-specific terms. Thus our estimate of $\hat{\beta}$ is identified only using changes in RVUs across the two time periods. The time effect further limits the identifying variation exclusively to relative changes in RVUs across services.

To obtain the share of spending linked to Medicare, we will primarily estimate equation (4.10) weighted by the average pre-update price of each service.[123] For the estimate of $\hat{\beta}$ in specification (4.10) to equal the true Medicare-linked share σ, we must make several assumptions about active renegotiations of reimbursement rates. Since group and group-by-service fixed effects are intended to capture the level of markup φ, any changes in this markup over time may show up in the error term. In Appendix C.3.2, we discuss the situations in which this challenges our ability to identify the parameter σ. Thanks to the context and data, which allow us to estimate responses at a high frequency, our assumptions are quite plausible.

To describe the timing with which BCBS incorporates Medicare updates into its reimbursements, we also present dynamic estimates from the following parametric event study:

$$\ln(P_{c,g,j,t}) = \sum_{t \neq 0} \beta_t \Delta \ln(RVU_j) \cdot 1_t + \phi_t 1_t + \phi_j 1_j + \phi_g 1_g + \phi_{g,j} 1_g \cdot 1_j + \eta_{c,g,j,t}. \tag{4.11}$$

When estimating equation (4.11), we normalize t such that $t = 1$ is the month in which BCBS has announced that it will implement Medicare's fee updates. We thus expect to see $\hat{\beta}_t = 0$ for periods preceding the updates' incorporation, $t < 0$, while the $\hat{\beta}_t$ for $t > 0$ are our estimates of how often Medicare updates are incorporated into private payments. A flat profile of the post-update $\hat{\beta}_t$ estimates would suggest that all price changes correlated with Medicare changes are implemented instantaneously. An upward trend in these coefficients might suggest that our baseline estimates are affected by ongoing renegotiations between BCBS and firms whose bargaining positions are affected

[123] Since the unweighted regression treats each claim equally, it effectively weights service codes by the frequency with which they are used.

by Medicare updates. We discuss this concern in detail in Appendix C.3.

4.4.3 Relating Our Approaches

The analyses we implement have complementary strengths and weaknesses. A shortcoming of the cross-sectional analysis of "bunching" is that it requires us to observe a constant markup across many services. So it may fail to detect genuine Medicare linkages involving markups that are common across relatively small numbers of services. But these linkages would be detected by the analysis of Medicare payment updates. Regardless of how common the markup is, the associated payments will change when the underlying relative values change.

A shortcoming of the latter approach, on the other hand, is that it could be biased if Medicare updates occur contemporaneously with changes driven by new contract negotiations. Our robustness analysis and our investigation of the precise timing of Medicare-linked changes provide evidence that new contract negotiations are unlikely to underlie our results. Nonetheless, these analyses cannot rule out active contract renegotiations altogether.

The "bunching" and "changes" approaches are thus complementary in that the bunching approach is prone to underestimating the extent of benchmarking while the changes approach is prone to overestimating the extent of benchmarking. We connect these approaches, and demonstrate consistency across the two analyses, by showing that the service-firm pairs we identify as benchmarked are strongly correlated across our approaches. We do this by dividing the data into subsamples according to the benchmarking results we obtain in our bunching analysis. We then estimate equation (4.10) separately on these subsamples.

4.5 Baseline Benchmarking Results

4.5.1 Bunching Estimates

Table 4.2 presents estimates of the share of services linked to Medicare in each year according to the method from section 4.4.1. The estimates explore our approach's sensitivity to two key assumptions. First, we round the value of each $ICF_{c,g,j,t}$ to the nearest 20 cents, 10 cents, or 2 cents to explore sensitivity to rounding error. Second, we define "common ICFs" as those that rationalize a sufficiently large share of the insurer's payments to a single physician group. In Figure 4.1, for example, the red line in Panel A should undoubtedly qualify as common. Other values may also qualify depending on the strictness of the threshold we apply. We consider thresholds ranging from 5 to 20 percent of a group's claims, then calculate the share of the insurer's payments associated with *any* of a group's common ICFs.

The Medicare-benchmarked shares range from 65 to 90 percent depending on the rounding and frequency thresholds; they decrease substantially with the stringency of the definition for a common

ICF, but are not sensitive to the choice of rounding threshold. Appendix Table C.2 shows that alternative measures generate qualitatively similar results.[124]

Going forward, we require as our baseline that common ICFs account for 10 percent of a group's claims, when rounded to the nearest $0.02. The motivation for adopting a stringent rounding threshold is to be conservative in the extent to which our method detects false positives. At the same time, the 10 percent threshold ensures that multiple ICFs can be readily detected. Using this definition, over half of firms have just one common ICF. Fewer than 5 percent have more than 2 common ICFs.

Having identified these ICFs, we use them to describe how the generosity of BCBS reimbursements relates to firm and market characteristics. Table 4.3 presents estimates of equation (4.4), which regresses the ICF values themselves (in logs) against physician group and market characteristics.[125] Columns 1 through 3 reveal that each of firm size, market share, and market concentration is, by itself, positively correlated with the generosity of the firms' payments. Consistent with other work on health care pricing (Dunn and Shapiro, 2014; Baker et al., 2014; Kleiner et al., 2015; Cooper et al., 2015; Clemens and Gottlieb, 2017), payments to large firms in markets with high levels of concentration are more generous than payments to small firms in markets with low levels of concentration. Columns 4 and 5 include all three characteristics together, with column 5 also adding fixed effects for service codes and geographic areas. Firm size remains a strong predictor of the average generosity of a firm's payments, as does overall market concentration. Market share switches signs, likely because of collinearity with log firm size.

4.5.2 Results from Medicare Fee Change Analysis

We next move on from estimating ICFs to exploiting Medicare's RVU changes. Using the method from section 4.4.2, Panel B of Figure 4.2 presents event study estimates of the link between Medicare's relative value scale and BCBS reimbursements. It shows estimates of equation (4.11) for the Medicare payment changes implemented in 2010. BCBS's provider newsletters say that updates to Medicare's payments took effect that year on July 1, 2010.

The estimates reveal substantial—but not universal—links between Medicare updates and the payments providers receive from BCBS. The coefficients imply that $\hat{\sigma} = 55$ percent of spending is linked to Medicare's relative values. The dramatic dynamics in the figure suggest that this reflects a contractual link between Medicare's relative values and BCBS payments. As in the raw data for standard office visits presented in Panel A, we see that payment changes occur when we expect. Importantly, the estimates of σ are both economically and statistically larger than 0 and smaller than 1, implying that payments for a substantial share of services deviate from strict benchmarking to Medicare's relative values. Sections 4.5.3 and 4.6.1 will investigate these deviations in detail.

[124] If we only count the single most common ICF for each group, the estimates are very similar to those reported in Table 4.2 when imposing a 20 percent threshold. Unfortunately, theory does not provide guidance as to which threshold is most appropriate, and the choice of threshold substantially affects our estimate of the linked share. Our changes-based estimation strategy is not sensitive to choices of this sort.

[125] Appendix Table C.3 also shows how these same characteristics relate to the frequency of deviations from Medicare benchmarking, and the value of the deviations when they occur.

Column 1 of Table 4.4 presents our baseline estimates of equation (4.10), which summarizes this result in a single coefficient. The estimate in column 1 of Panel A confirms that roughly 55 percent of BCBS's spending is linked to Medicare's relative values. In Panel B, we weight service codes equally rather than according to baseline payments. The unweighted estimate implies that, on average, roughly three quarters of BCBS's physician claims are paid based on Medicare's relative value scale. The difference in coefficients between Panels A and B implies that payments for relatively expensive services are less likely to be benchmarked to Medicare than are payments for low-cost services.[126]

4.5.3 Robustness and Cross-Validation of the Two Approaches

Table 4.4 probes the robustness of our changes-based estimates to a variety of specification checks. Column 1 reports our baseline specification, which includes a full set of group-by-HCPCS code fixed effects and controls for time effects with a simple post-update indicator. Column 2 drops the group-by-HCPCS code fixed effects in favor of a more parsimonious set of HCPCS code fixed effects. Column 3 augments the baseline specification by controlling for a cubic trend in the day of the year, which we interact with the size of each service's Medicare fee change. Column 4 allows the cubic trend in day to differ between the periods preceding and following the fee schedule update, as in a standard regression discontinuity design. The table shows that these specification changes have essentially no effect on the estimated coefficient $\hat{\beta}$. This reinforces the interpretation that, among services billed using standard HCPCS codes, roughly 55 percent of BCBS's spending is linked to Medicare's relative value scale.

The estimates presented in Figure 4.2 and Table 4.4 may differ from the true Medicare benchmarking parameter σ if changes in other terms of providers' contracts covary with the Medicare changes. Indeed, payment changes that significantly alter physician groups' average Medicare payment can move private payments in subsequent years, due in part to the resulting changes to their bargaining positions (Clemens and Gottlieb, 2017). In Appendix C.3.2, we thus draw on institutional detail and theoretically motivated specification checks to explore how much our estimates might deviate from the true share of payments benchmarked to Medicare's relative values. We find no evidence that renegotiations confound the relationship between BCBS's and Medicare's payments over the time horizons we analyze. Appendix C.3.2 thus bolsters the case for interpreting our estimates of $\hat{\beta}$ as measuring the fraction of services tied directly to Medicare.

To validate that our classification algorithm correctly captures services whose prices are actually linked to Medicare rates, we estimate our baseline changes-based regression separately for the services priced by common ICFs and those that are not. We classify each group-service pair (g, j) as Medicare-linked if all of group g's claims for service j in the pre-update period appear to be linked to Medicare rates, and as non-linked otherwise. We estimate equation (4.10) separately for these two samples.

Panel D of Figure 4.2 shows two binned scatterplots analogous to Panel C, relating log BCBS price changes to log Medicare fee changes separately for the two samples. The linked sample, shown with red triangles, has a slope of 0.9, indicating that BCBS prices for 90 percent of linked services

[126] Appendix Tables C.5 and C.6 replicate Panels A and B, respectively, in other years' data.

update in response to Medicare changes. The non-linked sample, shown with blue circles, has a much smaller slope of 0.3. Although smaller, this slope is significantly positive, indicating that around one-third of the services *not* priced according to common ICFs nevertheless react to Medicare changes.[127]

4.6 How Do Private Payments Deviate from Medicare?

In order to illuminate the economic determinants of benchmarking, we next consider variation in the strength of the link between private payments and Medicare's relative values. We consider the two dimensions that enter into reimbursement contracts: differences across physician groups and types of care.

4.6.1 Deviations from Benchmarking across Physician Groups

One key difference across groups, motivated by theory, is the scale of their business with BCBS. We measure the quantity of care each group provides in our data. To estimate heterogeneity along this dimension, we simply add interactions with practice size to our baseline changes regression, equation (4.10).

Table 4.5 shows the results. The first column reports the baseline, equally weighted regression from Table 4.4. The second column introduces interactions between the Medicare updates and indicators for the size of the physician group providing the care. We define mid-sized firms as those with \$200,000 to \$1,000,000 in annual billing with BCBS, and large firms as those with more than \$1,000,000 in annual billing. Each of these categories comprises one-quarter of the sample, with the remaining half of claims coming from smaller firms. The estimates imply that nearly 90 percent of services provided by firms billing less than \$200,000 are benchmarked to Medicare, while roughly 60 percent of services provided by firms billing more than \$1,000,000 are benchmarked. Columns 3 and 4 present similar, but dollar-weighted, estimates. The results in column 4 suggest that 77 percent of payments to firms billing less than \$200,000 are benchmarked to Medicare, while one-third of payments to firms billing more than \$1,000,000 are benchmarked.[128]

Figure 4.4 shows that we find a similar relationship between the share linked to Medicare and physician group size using our cross-sectional bunching approach. The series in the figure reveal that this is true in both the equally weighted and payment weighted series. It is also true whether or not we adjust for the underlying composition of each group's services, to which we now turn.[129]

[127] One way to interpret this one-third of services is that they still use Medicare's RVUs as a unit of measurement, but then use independent conversion factors. In this sense the RVUs are analogous to a currency; even though these one-third of prices are negotiated separately, they denominated in the "currency" of RVUs. But this appears not to be the case for most of the non-linked services.

[128] Appendix Table C.9 shows similar results in data from other years.

[129] To check whether the relationship between benchmarking and group size is affected by the composition of large and small groups' services, we run a regression that allows group size and service composition to enter simultaneously. We define fixed effects $1_{b(j)}$ at the level of the same 1-digit "Betos" classification we used in section 4.6. To measure the relationship between group size and the Medicare-linked share, we categorize physician groups g according to vigintiles of their aggregate private billing in a year, using $1_{s(g)}$ to denote vigintile fixed effects. We then estimate the following

4.6.2 Which Services Deviate from the Medicare Benchmark?

The value of improving on Medicare's menu depends on the severity of that menu's inefficiencies. Because it is difficult to systematically quantify Medicare's inefficiencies across a large range of individual services, we focus on one of the Medicare fee schedule's most salient problems. Medicare rates are computed based on average-cost reimbursement, so its reimbursements will hew closer to marginal costs for labor-intensive services than for capital-intensive services. Standard optimal payment models suggest that the latter would be more appropriately reimbursed through combinations of up-front financing of fixed costs and incremental reimbursements closer to marginal cost (Ellis and McGuire, 1986). We can proxy for services' capital and labor intensity by comparing the frequency of benchmarking across categories of care produced with different inputs, such as labor-intensive Evaluation & Management services versus capital-intensive Imaging.[130]

Table 4.6 estimates equation (4.10)—the relationship between private prices and changes in Medicare's relative values—separately across broad categories of services. The estimates imply that nearly 30 percent more of the payments for Evaluation & Management services are linked directly to Medicare's relative values than for Imaging services.[131]

Second, we divide Imaging codes into subcomponents with high capital and high labor content. Providers often bill separately for taking an image (the capital-intensive part, since it requires an imaging machine) and interpreting it (the labor-intensive part). When the same group supplies both components, it submits the bill as a "Global" service. The results in columns 5 through 7 show that payments for the labor-intensive Professional Component are more tightly linked to Medicare's relative values than are the payments for the capital-intensive Technical Component. These patterns support the hypothesis that physicians and insurers are more likely to contract away from Medicare's menu for capital-intensive services than for labor-intensive ones.

Table 4.7 shows that we find a similar relationship between the share linked to Medicare and service categories using our cross-sectional approach. Using our cross-sectional approach, we find that benchmarking is 30–50 percent less frequent for Imaging, Procedures, and Tests than for Evaluation & Management services. The results across columns reveal that we find similarly substantial differentials whether or not we control for firm size and whether services are weighted according to the spending they represent.

These results suggest that private contracts deviate when Medicare's rates are most problematic from an efficiency perspective. One way to interpret this is in light of negotiation and adjustment costs. Private bargaining can overcome these frictions more easily when Medicare's rates are farther

regression at the group-code level:

$$\text{Medicare-Linked Share}_{j,g} = \nu_b 1_{b(j)} + \zeta_s 1_{s(g)} + \upsilon_{j,g}. \quad (4.12)$$

The orange diamonds in Figure 4.4 show the estimates of $\hat{\zeta}_s$. These illustrate the relationship between Medicare links and group size, adjusted for service composition. The composition-adjusted relationship remains strongly negative. The remaining measures in Figure 4.4 show similar results in terms of dollars spent, rather than number of services.

[130] We split all of the medical services in our data into the "Betos" categories defined by Berenson and Holahan (1990). This hierarchical classification system goes from the broad categories we use here (such as Evaluation & Management and Imaging) to 2-digit (e.g. Advanced Imaging [MRIs and CAT scans]) and 3-digit classifications (e.g. CAT Scan: Head).

[131] Appendix Table C.8 replicates this analysis in other years' data.

from the efficient or equilibrium level that would obtain under unconstrained negotiations.

4.6.3 How Do Deviations Change Incentives Relative to Medicare?

What are physicians and insurers aiming to achieve when they negotiate reimbursements that deviate from Medicare's relative prices? In this section, we present evidence on the direction of deviations from strictly Medicare-benchmarked rates to investigate what services BCBS rewards through upward adjustments and discourages through downward adjustments. We do so by describing residuals from the following regression:

$$\ln(P_{g,j}) = \psi \ln(RVU_j) + \mu_g + e_{g,j}. \tag{4.13}$$

In a world of perfect benchmarking, we would find $\hat{\psi} = 1$ and $e_{g,j}$ uniformly equal to 0. So the empirical prediction errors $\hat{e}_{g,j}$ contain information about the direction of deviations from strict Medicare benchmarking. We examine heterogeneity in this prediction error across categories of services.[132]

Table 4.8 presents means of $\hat{e}_{g,j}$ from equation (4.13) across Betos categories.[133] The table shows that payments for Evaluation & Management and Testing services generally have positive residuals while payments for services in Imaging and Procedures have negative residuals. Figure 4.5 Panel A plots the distributions of these residuals by service category. The distribution for Imaging shows far more density of negative residuals than those for other services. Testing has more positive residuals, although that is largely driven by one outlier code.[134] Compared to the relative payments implied by Medicare's relative values, BCBS systematically adjusts its contracts to discourage imaging services. This coincides with the conventional wisdom that Medicare's relative values underpay for labor-intensive services relative to other services, and suggests that BCBS aims to partly rectify that mispricing.

Differences in BCBS's adjustments for labor- and capital-intensive services are particularly sharp across the subcategories of diagnostic imaging. Payment adjustments for the labor-intensive Professional Component of these services are substantially positive, at around 0.07 in logs (approximately 7 percent). Payment adjustments for the capital-intensive Technical Component of these services are substantially negative, averaging -0.12 in logs. Figure 4.5 Panel B shows that this pattern holds throughout the distribution. While it is clear that BCBS reimbursements lean heavily on Medicare's relative values for their basic payment structure, these results suggest that BCBS adjusts its con-

[132] A subtle but important point is that this approach captures deviations from Medicare's relative prices that come through the introduction of multiple Medicare-benchmarked conversion factors. If an insurer thinks the Medicare menu's primary inefficiency is that it uniformly overpays for diagnostic imaging services relative to other services, for example, its preferred contract may simply set a low conversion factor for imaging services and a high conversion factor for other services. Our previous analyses would describe such a contract as being fully linked to Medicare. The analysis in the current section will capture the fact that this is structured to discourage the use of imaging services relative to other services.

[133] To be precise, these means are $\overline{\hat{e}_{g,j}} = \frac{1}{N_b}\sum_{j\in b}\hat{e}_{g,j}$, where each Betos group b comprises N_b claims for all services $j \in b$ in that group.

[134] In the Testing category the vast majority of residuals are negative, with the exception of one of the more common tests, which has a large and positive average residual. Recall from section 4.3, however, that Testing is the one category with significant missing data problems.

tracts to increase the generosity of payments for labor-intensive services and decrease its payments for capital-intensive services. This is consistent with deviating from Medicare with an eye towards more closely targeting either marginal costs, or medical value.

4.7 Conclusion

This chapter uses physician payments from a large private insurer as a window into how private firms contract for services in complex environments. Using two empirical strategies, we show that they benchmark to Medicare's schedule of relative prices to significantly simplify this problem, and estimate that roughly 75 percent of services and 55 percent of spending are directly linked to Medicare.

We find evidence that the one quarter of services and nearly half of payments that deviate from Medicare's relative rate structure involve an effort to improve the payment structure. Insurers tend to deviate when the value of doing so appears to be highest. Deviations occur disproportionately in contracts with large physician groups, where significant mutual gains can be achieved. They significantly reduce payments for diagnostic imaging services, a category of care for which many academics and policy makers believe marginal benefits are low relative to costs (Winter and Ray, 2008; MedPAC, 2011). But they hew closely to Medicare in payments for services where average-cost reimbursements will be most aligned with marginal costs, such as labor-intensive primary care services. When contracts deviate from Medicare, the direction of payment adjustments would tend to encourage the provision of primary care and discourage care for which over-utilization is a more widespread concern.

4.8 Tables

Table 4.1: Summary Statistics by Physician Group

Panel A: All Groups					
(N=80,675)	**Mean**	**Median**	**Std. Dev.**	**Min.**	**Max.**
Number of unique services	9.70	3	27.23	1	~1,700
Number of patients	87.59	2	698.85	1	~61,930
Number of doctors	1.73	1	7.93	1	~1,100
Number of claims	201	3	1,763	1	~163,360
Mean allowed amount	108.91	84.43	125.16	0.64	~7,680
Total BCBS revenue	25,457	383	274700.3	0.64	~43,000,000
Panel B: Groups with Billings $> \$10,000$					
(N=15,235)	**Mean**	**Median**	**Std. Dev.**	**Min.**	**Max.**
Number of unique services	35.99	24	53.12	1	~1,700
Number of patients	424.35	151	1,523	1	~61,930
Number of doctors	4.14	2	17.56	1	~1,100
Number of claims	981.13	386	3860	1	~163,360
Mean allowed amount	105.52	84.65	136.3	10.75	~7,680
Total BCBS revenue	124,687	44392	606,644	10000	~43,000,000

Note: this table shows summary statistics for data by physician group. Source: Authors' calculations using claims data from BCBS.

Table 4.2: Services Priced According to Common Implied Conversion Factors

Panel A: Dollar-Weighted			
	Frequency Threshold:		
	5%	*10%*	*20%*
Rounding for ICFs:			
\$0.02	83%	76%	66%
\$0.10	86%	80%	71%
\$0.20	87%	80%	71%
Panel B: Service-Weighted			
	Frequency Threshold:		
	5%	*10%*	*20%*
Rounding for ICFs:			
\$0.02	87%	81%	70%
\$0.10	89%	84%	75%
\$0.20	89%	85%	75%

Note: each cell shows the share of services for which payments are associated with a common Implied Conversion Factor (cICF), as defined in the main text. Data are from January 1—June 30, 2010, over which time BCBS used the 2009 version of Medicare's Resource Based Relative Value Scale. The cells within each panel show how the linked share varies as we apply different thresholds for the frequency required to quality as a cICF. The column labeled "Rounding" indicates the rounding applied to each estimated ICF. An ICF is defined as "common" for the payments to a physician group if it accounts for at least the fraction of services associated with the specified Frequency Threshold. Source: Authors' calculations using claims data from BCBS.

Table 4.3: Firm Size and Implied Conversion Factors

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	Log implied conversion factor (ICF)				
Firm Size (Log Spending)	0.058**			0.058**	0.040**
	(0.004)			(0.005)	(0.006)
Firm Market Share		0.241**		-0.158**	-0.092**
		(0.015)		(0.037)	(0.029)
Market Concentration			0.238**	0.318**	0.159**
			(0.020)	(0.036)	(0.028)
N	20,736,449	20,736,449	20,736,449	20,736,449	20,736,449
No. of Clusters	23,098	23,098	23,098	23,098	23,098
Code Effects	No	No	No	No	Yes
HSA Fixed Effects	No	No	No	No	Yes

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. This table shows estimates of relationship between features of physicians' contracts and measures of firm size and/or market concentration. The construction of all variables is discussed in the main text. Source: Authors' calculations using claims data from BCBS.

Table 4.4: Estimating Medicare Benchmarking Using RVU Changes

	(1)	(2)	(3)	(4)
Dependent variable:	*Log private reimbursement rate*			
	Panel A: Weighted by Price			
Log RVU Change × Post	0.539**	0.544**	0.568**	0.538**
	(0.061)	(0.061)	(0.060)	(0.061)
N	23,933,577	23,933,577	23,933,577	23,933,577
No. of Clusters	3,681	3,681	3,681	3,681
	Panel B: Unweighted			
Log RVU Change × Post	0.750**	0.748**	0.765**	0.749**
	(0.038)	(0.038)	(0.043)	(0.038)
N	23,933,577	23,933,577	23,933,577	23,933,577
No. of Clusters	3,681	3,681	3,681	3,681
Group-by-Code Effects	Yes	No	Yes	Yes
Code Effects	No	Yes	No	No
Cubic Time × RVU Change	No	No	Yes	No
Cubic Time × Post	No	No	No	Yes

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. The table shows the results of OLS specifications of the form described in section 4.4.2. Each column in each panel reports an estimate of $\hat{\beta}$ from equation (4.10). Observations are at the claim-line level and are equally weighted (Panel B), or weighted according to each service's average payment during the baseline period (Panel A). Data are from 2010. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Additional features of each specification are described within the table. The construction of all variables is further described in the main text. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

Table 4.5: Medicare Benchmarking by Firm Size

	(1)	(2)	(3)	(4)
Dependent variable:	*Log private reimbursement rate*			
Log RVU Change	0.750**	0.882**	0.539**	0.775**
× Post-Update	(0.038)	(0.073)	(0.061)	(0.094)
Log RVU Change		-0.074		-0.140*
× Post-Update × Midsize		(0.098)		(0.069)
Log RVU Change		-0.293*		-0.448**
× Post-Update × Large		(0.117)		(0.102)
N	23,933,577	23,933,577	23,933,577	23,933,577
Weighting:	Service	Service	Dollar	Dollar

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. Columns 1 and 3 report the baseline estimates from Table 4.4 Panels A and B respectively. In columns 2 and 4 we augment these specifications to include interactions between firm size indicator variables and both the "Post" indicator and the interaction between the "Log RVU Change" and "Post" indicator. The omitted category is small firms, defined as those with less than $200,000 in billings. Mid-sized firms are those with billings between $200,000 and $1 million, and large firms are those with billings exceeding $1 million. Data are from 2010. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

Table 4.6: Public-Private Payment Links Across Service Categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	*Log private reimbursement rate*						
	Evaluation	Imaging	Procedures	Tests	Imaging Sub-Categories:		
					Global	Technical	Professional
Log RVU Change	0.841**	0.564**	0.720**	1.066**	0.545**	0.387*	0.982**
× Post-Update	(0.036)	(0.084)	(0.081)	(0.066)	(0.109)	(0.152)	(0.066)
N	12,259,186	3,630,019	4,750,313	1,542,254	1,826,666	209,178	1,594,175
No. of Clusters	221	1,085	1,936	408	408	244	433

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. The table shows the results of OLS specifications of the form described in section 4.4.2. The cells in each panel report estimates of $\hat{\beta}$ from equation (4.10), with samples selected to contain the HCPCS codes falling into broad service categories. The name of the relevant service category accompanies each point estimate. Data are from 2010. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). The construction of all variables is further described in the main text. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

Table 4.7: Medicare Benchmarking by Betos Category

	(1)	(2)	(3)	(4)
Dependent variable:	Payments with Common Conversion Factors			
	Spending Share		Service Share	
Imaging	-0.427**	-0.471**	-0.300**	-0.355**
	(0.053)	(0.047)	(0.030)	(0.024)
Procedures	-0.309**	-0.352**	-0.336**	-0.388**
	(0.030)	(0.028)	(0.054)	(0.052)
Tests	-0.383**	-0.415**	-0.258**	-0.297**
	(0.051)	(0.047)	(0.055)	(0.054)
Constant	0.921**	0.828**	0.941**	0.829**
	(0.015)	(0.015)	(0.020)	(0.017)
N	542,207	542,207	542,207	542,207
Omitted Category	Evaluation & Management			
Additional Controls	Group Size	None	Group Size	None

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. This table shows estimates of the η_b coefficients in equation (4.12), namely the relationship between Betos category and the Medicare-linked share of claim lines (columns 1 and 2) or spending (columns 3 and 4). Medicare links are measured using the common Implied Conversion Factors (cICFs) defined in section 4.4.1, using data from January 1 through June 30, 2010. We require that cICFs account for 10 percent of a group's claims, when rounded to the nearest $0.02. Columns 1 and 3 show estimates after controlling for vigintile of group size, as measured with BCBS spending, and columns 2 and 4 show estimates without group size controls. Standard errors are two-way clustered (Cameron et al., 2011) by Betos category and physician group. Sources: Authors' calculations using claims data from BCBS.

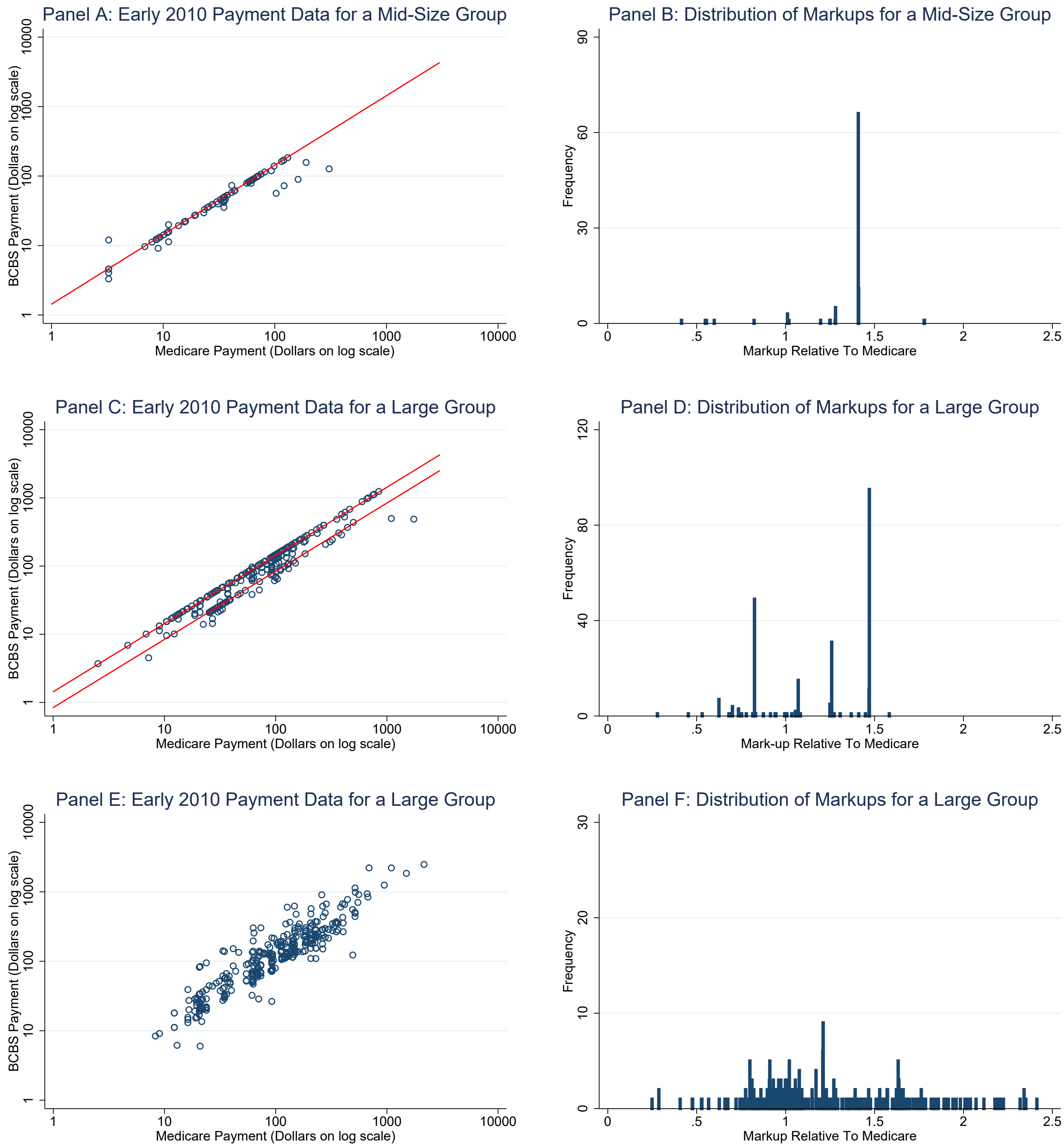
Table 4.8: In What Direction Does BCBS Adjust Its Payments for the Various Service Categories?

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Distributions of Payment Residuals by Betos Categories						
	Evaluation & Management	Imaging	Procedures	Tests	Imaging Sub-Categories: Global	Technical	Professional
Residual Mean	0.0112	-0.0624	0.0107	0.0301	-0.122	-0.124	0.0150
Residaul SD	(0.169)	(0.246)	(0.279)	(0.319)	(0.272)	(0.281)	(0.177)
N	6,010,826	1,743,011	2,312,734	751,726	883,419	102,465	757,127

Note: this table presents means and standard deviations of residuals from estimates of equation (4.13) in data from 2010. That is, we regress the log of BCBS's payments on a set of physician-group fixed effects and the log of each HCPCS code's number of Relative Value Units. This table describes the residuals from that regression. We restrict the sample to the pre-update period (January 1 through June 30, 2010) so that the relative value units are constant for each service throughout the sample. Source: Authors' calculations using claims data from BCBS.

4.9 Figures

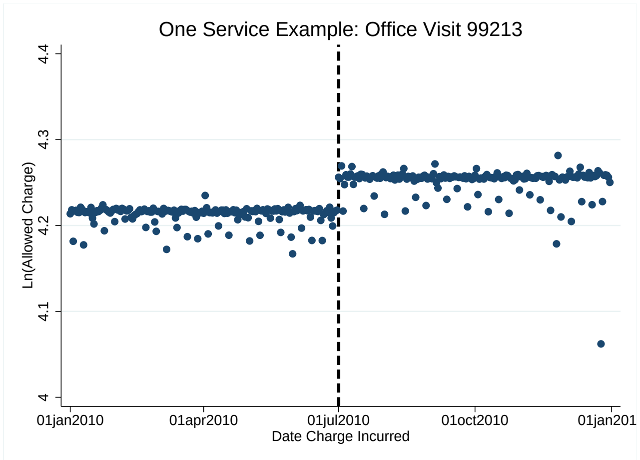
Figure 4.1: Raw Payments For Illustrative Physician Groups



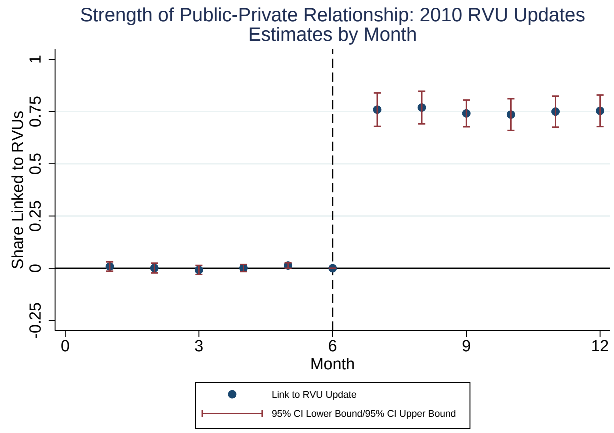
Note: Panels A through F present the raw data on BCBS reimbursement rates, and associated Medicare reimbursement, for 3 different physician groups in 2010. In Panels A, C, and E, each observation is a unique reimbursement paid for a particular service to the group. The lines have a slope of 1 (in logs) and represent the groups' most common Implied Conversion Factors. Panels B, D and F plot the distribution of markups relative to the Medicare rates for all payments each group received. They show clear spikes at the values that we identify as common Implied Conversion Factors in Panels A, C, and E. To comply with confidentiality rules, we omit from these graphs a small share of each group's claims. The share of claims whose observations are suppressed is 14.2% in Panels A and B, 1.94% in Panels C and D, and 2.95% in Panels E and F. Source: Authors' calculations using RVUs from the *Federal Register* and claims data from BCBS.

Figure 4.2: Benchmarking Estimates Based on Price Changes

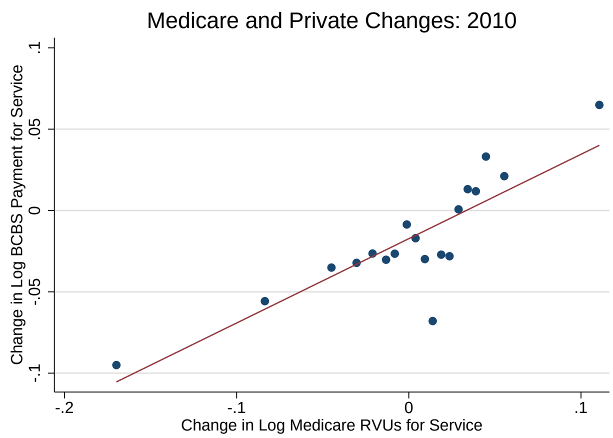
Panel A



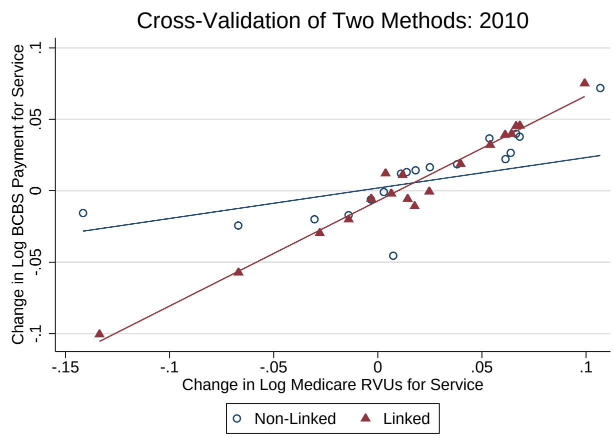
Panel B



Panel C

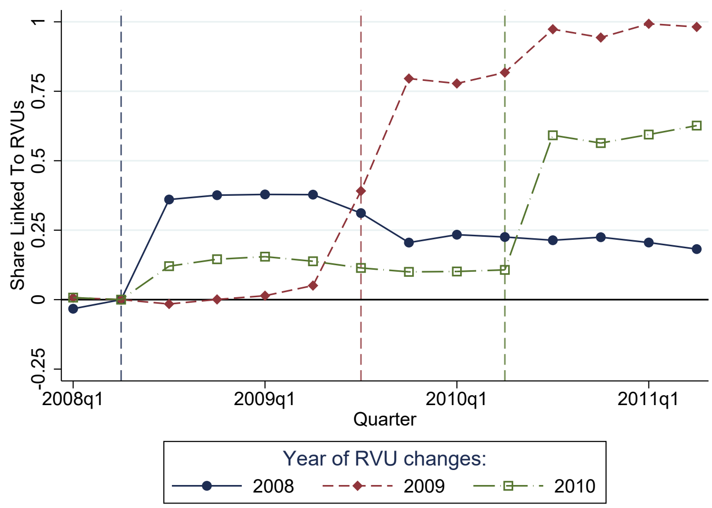


Panel D



Note: all panels use data from calendar year 2010. BCBS implemented its update from the 2009 to 2010 relative value scales on July 1, 2010, as indicated by the vertical dashed line in Panels A and B. Panel A presents daily averages of BCBS's log payment for a standard office visit. Panel B reports estimates of the β_p from estimates of equation (4.11). Panel C presents a binned scatterplot of the relationship between Medicare payment updates (sorted into 20 vigintiles) and changes in private payments. In Panel C, private price changes are computed as the difference between service-level average payments after and before July 1, 2010. Panel D is similar, but with separate data and estimation for services that we identify as being linked to Medicare on the basis of their implicit conversion factors and those we identify as being non-linked. For presentation in the binned scatterplot, observations within each class of services (i.e.,linked or non-linked) are grouped into twenty vigintiles on the basis of the log change in the service code's Medicare RVU allocation. The regression lines shown in Panels C and D are estimated at the underlying service-code level. Sources: Authors' calculations using RBRVS updates from the *Federal Register* and claims data from BCBS.

Figure 4.3: Estimating Multiple Years' RVU Updates Simultaneously

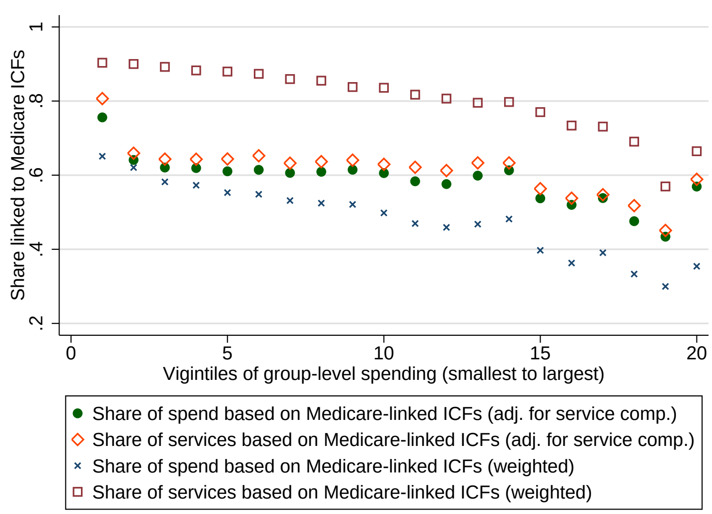


Note: the figure reports estimates of β_t^{08}, β_t^{09} and β_t^{10} from the following modification of equation (4.11):

$$\ln(P_{c,g,j,t}) = \sum_{t\neq 0} \beta_t^{08} \Delta \ln(RVU_{j,08}) \cdot 1_t + \sum_{t\neq 0} \beta_t^{09} \Delta \ln(RVU_{j,09}) \cdot 1_t^{10} + \sum_{t\neq 0} \beta_t \Delta \ln(RVU_{j,10}) \cdot 1_t + \phi_t 1_t + \phi_j 1_j + \phi_g 1_g + \phi_{g,j} 1_g \cdot 1_j + \eta_{c,g,j,t}. \tag{4.14}$$

In this specification, $\Delta \ln(RVU_{j,T})$ refers to the log of Medicare's RVU updates from calendar year $T-1$ to calendar year T. The corresponding coefficients β_t^T indicate what share of the year-T RVU updates were incorporated into BCBS payments during calendar quarter t. BCBS implemented its RVU updates on July 1, 2008, August 15, 2009, and July 1, 2010. The omitted interaction ($t = 0$) is 2008Q2 for all of the RVU update variables. The regression line is estimated at the underlying service-code level and is dollar-weighted. Sources: Authors' calculations using claims data from BCBS.

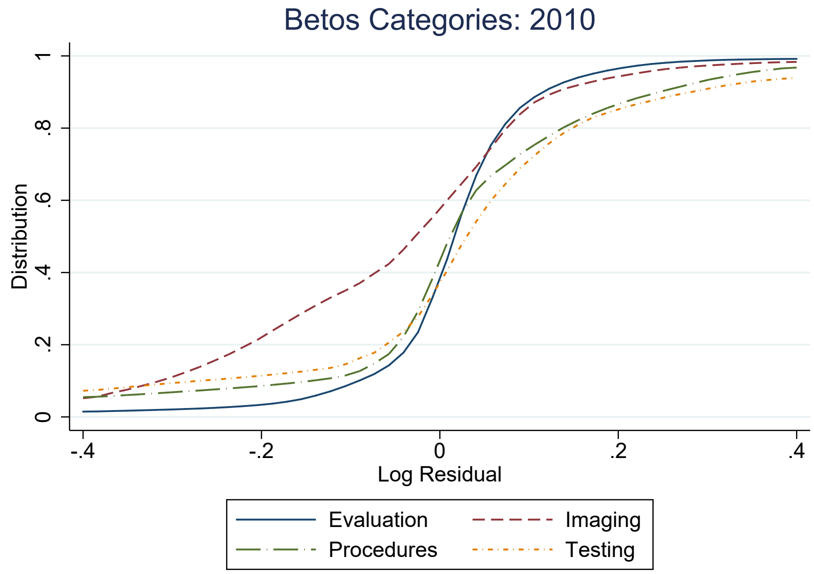
Figure 4.4: Frequency of Benchmarking and Physician Group Size



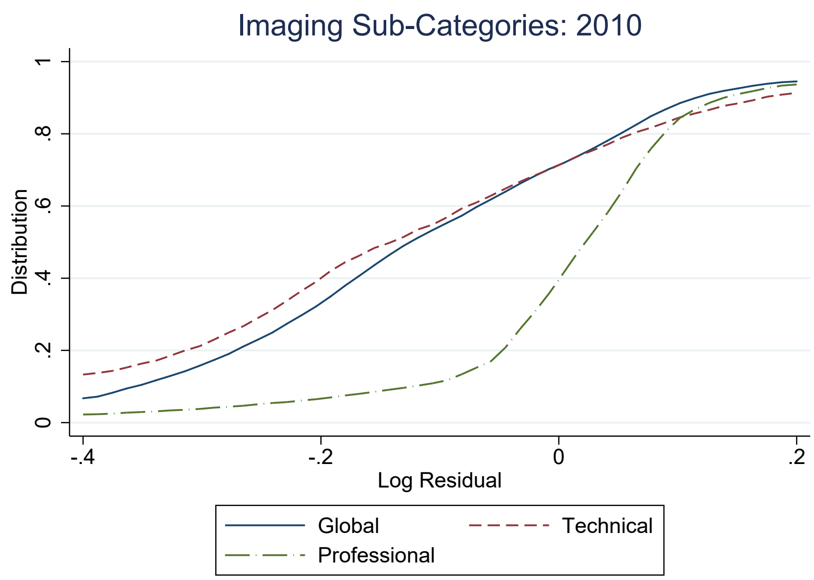
Note: this figure shows the relationship between a group's Medicare-linked service share and group size. Specifically, it plots variation in the share of services priced according to common Implied Conversion Factors (cICFs), as defined in section 4.4.1, according to physician group size. We measure group size by forming 20 vigintiles based on the group's BCBS billing. We require that cICFs account for 10 percent of a group's claims, when rounded to the nearest \$0.02. The green dots and orange diamonds show estimates of ζ_b from equation (4.12), which adjust for the composition of each group's services. The blue $\times$'s and red squares are unadjusted, but weighted to measure the Medicare-linked share of spending in dollar terms as opposed to the share of services. All data are from 2010. Sources: Authors' calculations using claims data from BCBS.

Figure 4.5: Deviations from Medicare Benchmark by Service Category

Panel A



Panel B



Note: the figure presents residuals $\epsilon_{g,j}$ from estimates of equation (4.13). The distribution of residuals is shown within either broad Betos categories (Panel A), or within the subcategories of Imaging (Panel B). The distributions are smoothed using a local linear regression, with an Epanechnikov kernel and a bandwidth of 0.01. Source: Authors' calculations using claims data from BCBS.

Chapter 5

Conclusion

This thesis consists of three chapters in family economics, early childhood development, economics of education and health economics. In all chapters I examine how a unique institutional feature of a major economic sector—a universal daycare price change, the opportunity of academic redshirting and Medicare price change—shapes families' and firms' incentives and impacts their behavior.

First I study parents' quality time allocation to their children, and show that universal daycare price can be an effective policy instrument to influence parental quality time. I focus on differences across education groups, where the motivation stems from the parental time-education puzzle. To resolve it, I examine parents' responses to a shock to the opportunity costs of their time. For identification, I exploit a drastic decrease in universal daycare prices in Quebec in 1997, leading mothers to work more and households to consume more of typical home production and child market goods. Strikingly, I find that higher-educated parents do not only they spend more time with their children in the cross-section, but they increase child time even more, after daycare price falls. The child time increase and home production time decrease is driven by higher-educated mothers drawn (back) into the labor market—for whom mental health and parenting practices improved, impacting positively their child's hyperactivity/aggression/anxiety scores. The mechanism I confirm is that higher-educated parents' time has a larger marginal return in non-market activities, outweighing their higher wage. My findings uncover the pivotal role of substitutability between time and market goods, and suggest: parents substitute their time from activities where time is substitutable with market goods (home production) to activities where time is complementary to market goods (child human capital production).

One immediate policy implication of my findings is that since both higher- and lower-educated parents increase parental time and daycare time after daycare price falls, the level of child human capital–if produced solely from these two inputs–is expected to increase for both; whether the childhood skill gap shrinks or expands depends on the concavity of the human capital production function, and the degree of time efficiency advantage of higher-educated parents. Compared to the counterfactual scenario of no policy for ages 3-4, the subsidized daycare policy actually increases the level of child human capital by 17 and 14 percentages for higher- and lower-educated parents' children, respectively; resulting in a 30 percent larger skill gap. If the policy had not been universal but targeted only at low-educated families, approximately half of this (increased) gap could be eliminated. Thus, there is a trade-off between increasing the level of child human capital *versus* decreasing inequality, and these competing policy objectives need to be weighed when designing a universal or a targeted daycare policy. By quantifying the importance of the time efficiency mechanism I find that (1) roughly 36 percent of the actual increase in the level of child human capital comes from high-educated parents' larger increase in inputs, while the rest is due to their time efficiency advantage, and (2) if lower-educated parents had no time efficiency disadvantage, the actual skill gap could be closed by 27 percent.

In the second chapter I present the causal impact of academic redshirting–the practice of postponing school entry of an age-eligible child. I show that redshirting can be a beneficial flexible opportunity within the educational system primarily for disadvantaged boys; this is an important policy-implication from the aspect of equality of opportunity, since disadvantaged boys are the most likely to suffer from school-readiness problems, such as hyperactivity or attention disorders. My results suggest that mental health is one of the mechanisms behind disadvantaged boys benefiting the most from delayed school entry: boosted human capital in the extra year before school leads to better mental health–such as lower anxiety–, associated with better student achievement. This presumed mechanism based on human capital is also consistent with the finding that entering primary school a year later matters because it makes the child older in absolute terms, rather than making him/her older relative to classmates.

In the third chapter with co-authors Jeffrey Clemens and Joshua Gottlieb we investigate how frequently privately negotiated physician payments deviate from the public sector (Medicare) benchmark, and find that prices for 25 percent of physician services, representing 45 percent of spending, do deviate. We use administrative outpatient claims data from a large private insurer Blue Cross Blue Shield of Texas (BCBS), and exploit the institutional detail of the exact day on which BCBS implemented the annual Medicare payment updates (e.g. July 1st in 2010). We show that the Medicare-benchmarked share is high for services provided by small physician groups and low for capital-intensive services, for which Medicare's average-cost reimbursement schedule deviates most from the marginal cost. Our results suggest that providers and private insurers coordinate around Medicare's menu of relative payments for simplicity but—when the value at stake is sufficient—do indeed innovate.

A number of different factors could contribute to this behavior, and we hope that future research will disentangle them. Benchmarking could be the market's way of simplifying a complex contracting problem, analogous to bounded rationality for individuals. Relatedly, it could reflect insurers' desire to free-ride off of Medicare since computing optimal prices for every service may be excessively costly. Alternatively, it could purely reflect a bargaining outcome in which Medicare is the outside option for many, but not all, services. However, regardless of the explanation, the use of Medicare as a pricing backstop implies that many inefficiencies in Medicare's reimbursements spill over into private fee schedules. By extension, the value of improvements to public payment systems may ripple through private contracts in addition to improving the performance of Medicare itself. At the same time, we find that the insurers adjust their payments to curb what policy analysts regard as Medicare's greatest inefficiencies. Both public and private players thus appear to have important roles in the process of fee schedule improvement and payment system reform.

Bibliography

Aguiar, M. and Hurst, E. (2007). Life-Cycle Prices and Production. *American Economic Review*, 97(5):1533–1559.

Aizer, A. and Cunha, F. (2012). The Production of Human Capital: Endowments, Investments and Fertility. Working Paper No. 18429, National Bureau of Economic Research.

Almond, D. and Doyle, J. J. (2011). After Midnight: A Regression Discontinuity Design in Length of Postpartum Hospital Stays. *American Economic Journal: Economic Policy*, 3(3):1–34.

Alpert, A., Duggan, M., and Hellerstein, J. K. (2013). Perverse Reverse Price Competition: Average Wholesale Prices and Medicaid Pharmaceutical Spending. *Journal of Public Economics*, 103:44–62.

Anderson, E., Jaimovich, N., and Simester, D. (forthcoming). Price Stickiness: Empirical Evidence of the Menu Cost Channel. *Review of Economics and Statistics*.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 1 edition.

Aud, S., Wilkinson-Flicker, S., Kristapovich, P., Rathbun, A., Wang, X., and Zhang, J. (2013). The Condition of Education 2013. Technical Report 37, National Center for Education Statistics.

Baker, L. C., Bundorf, M. K., Royalty, A. B., and Levin, Z. (2014). Physician Practice Competition and Prices Paid by Private Insurers for Office Visits. *Journal of the American Medical Association*, 312(16):1653–1662.

Baker, M. (2011). Innis Lecture: Universal Early Childhood Interventions: What is the Evidence Base? *Canadian Journal of Economics*, 44(4):1069–1105.

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. (2015). Non-Cognitive Deficits and Young Adult Outcomes: The Long-Run Impacts of a Universal Child Care Program. NBER Working Paper 21571, National Bureau of Economic Research.

Baker, M. and Milligan, K. (2016). Boy-Girl Differences in Parental Time Investments: Evidence from Three Countries. *Journal of Human Capital*, 10(4):1–44.

Bassok, D. and Reardon, S. F. (2013). "Academic Redshirting" in Kindergarten Prevalence, Patterns, and Implications. *Educational Evaluation and Policy Analysis*.

Bauernschuster, S. and Schlotter, M. (2015). Public Child Care and Mothers' Labor Supply: Evidence from Two Quasi-Experiments. *Journal of Public Economics*, 123:1–16.

Becker, G. S. (1965). A Theory of the Allocation of Time. *The Economic Journal*, 75(299):493–517.

Bedard, K. and Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *The Quarterly Journal of Economics, MIT Press*, 121(4):1437–1472.

Berenson, R. and Holahan, J. (1990). Using A New Type-of-Service Classification System to Examine the Growth in Medicare Physician Expenditures, 1985–1988. NTIS Publication PB91-188599, Urban Institute Health Policy Center.

Berlinski, S. and Galiani, S. (2007). The Effect of a Large Expansion of Pre-Primary School Facilities on Preschool Attendance and Maternal Employment. *Labour Economics*, 14(3):665–680.

Besley, T. J. (1988). Optimal Reimbursement Health Insurance and the Theory of Ramsey Taxation. *Journal of Health Economics*, 7(4):371–336.

Bils, M. and Klenow, P. J. (2004). Some Evidence on the Importance of Sticky Prices. *Journal of Political Economy*, 112(5):947–985.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2011). Too Young to Leave the Nest? The Effects of School Starting Age. *The Review of Economics and Statistics*, 93(2):455–467.

Blau, D. and Tekin, E. (2007). The Determinants and Consequences of Child Care Subsidies for Single Mothers in the USA. *Journal of Population Economics*, 20(4):719–741.

Blue Cross Blue Shield of Texas (2008). Fee Schedule Update. *Blue Review*, (1):8.

Blue Cross Blue Shield of Texas (2009). Fee Schedule Updates Effective in August. *Blue Review*, (2):5.

Blue Cross Blue Shield of Texas (2010). Fee Schedule Update. *Blue Review*, (3):15.

Blue Cross Blue Shield of Texas (2011). Fee Schedule Update. *Blue Review*, (6):10.

Borges, W. (2003). Mo' Better Blues. *Texas Medicine*, 99(4).

Borghans, L., Duckworth, A. L., Heckman, J. J., and ter Weel, B. (2008). The Economics and Psychology of Personality Traits. *Journal of Human Resources*, 43(4):972–1059.

Bound, J. and Jaeger, D. A. (2001). Do Compulsory School Attendance Laws Alone Explain the Association Between Quarter of Birth and Earnings? *Research in Labor Economics*, 19:83–108.

Brekke, K. R., Holmås, T. H., Monstad, K., and Straume, O. R. (2015). Do Treatment Decisions Depend on Physicians' Financial Incentives? Working Ppaer No. 07/2015, NIPE. Available online at http://www.nipe.eeg.uminho.pt/Uploads/WP_2015/NIPE_WP_07_2015.pdf (accessed December 9, 2015).

Brodeur, A. and Connolly, M. (2013). Do Higher Child Care Subsidies Improve Parental Well-being? Evidence from Quebec's Family Policies. *Journal of Economic Behavior and Organization*, 93:1–16.

Buckles, K. S. and Hungerman, D. M. (2013). Season of Birth and Later Outcomes: Old Questions, New Answers. *The Review of Economics and Statistics*, 95(3):711–724.

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.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Cannon, J. S. and Lipscomb, S. (2008). Changing the Kindergarten Cutoff Date: Effects on California Students and Schools. Occasional papers, Public Policy Institute of California.

Carlsson, M., Dahl, G. B., Okert, B., and Rooth, D.-O. (2015). The Effect of Schooling on Cognitive Skills. *The Review of Economics and Statistics*, 97(3):533–547.

Carneiro, P. and Ginja, R. (2014). Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start. *American Economic Journal: Economic Policy*, 6(4):135–73.

Carneiro, P., Heckman, J. J., and Masterov, D. V. (2005). Labor Market Discrimination and Racial Differences in Premarket Factors. *Journal of Law and Economics*, 48(1):1–39.

Carroll, J. (2007). How Doctors Are Paid Now, And Why It Has to Change. *Managed Care*, (12).

Cascio, E. (2015). The Promises and Pitfalls of Universal Early Education. *IZA World of Labor*.

Cascio, E. and Schanzenbach, D. W. (2013). The Impacts of Expanding Access to High-Quality Preschool Education. *Brookings Papers on Economic Activity*, 47(2):127–192.

Case, A., Lubotsky, D., and Paxson, C. (2002). Economic Status and Health in Childhood: The Origins of the Gradient. *American Economic Review*, 92(5):1308–1334.

Center for Studying Health System Change (1999). Physician Survey Summary File: User's Guide and Codebook. Community Tracking Study Physician Survey, 1996–1997 (ICPSR 2597). Available at http://doi.org/10.3886/ICPSR02597.v3.

Center for Studying Health System Change (2006). 2004–05 Physician Survey Summary File: User's Guide and Codebook. Community Tracking Study Physician Survey, 2004–2005 (ICPSR 4584). Available at http://doi.org/10.3886/ICPSR04584.v2.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., and Yagan, D. (2011). How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics*, 126(4):1593–1660.

Chetty, R., Hendren, N., Kline, P., Saez, E., and Turner, N. (2014). Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review*, 104(5):141–147.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, 113(1):1–45.

Clemens, J. and Gottlieb, J. D. (2013). In the Shadow of a Giant: Medicare's Influence on Private Payment Systems. Working Paper No. 19503, National Bureau of Economic Research.

Clemens, J. and Gottlieb, J. D. (2014). Do Physicians' Financial Incentives Affect Medical Treatment and Patient Health? *American Economic Review*, 104(4):1320–1349.

Clemens, J. and Gottlieb, J. D. (2017). In the Shadow of a Giant: Medicare's Influence on Private Payment Systems. *Journal of Political Economy*, 125(1):1–39.

Coneus, K., Gernandt, J., and Saam, M. (2011). Noncognitive Skills, School Achievements and Educational Dropout. *Schmollers Jahrbuch - Journal of Applied Social Science Studies*, 131:1–22.

Cooper, Z., Craig, S. V., Gaynor, M., and Van Reenen, J. (2015). The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured. Working Paper No. 21815, National Bureau of Economic Research.

Crawford, C., Dearden, L., and Meghir, C. (2007). When You Are Born Matters: The Impact of Date of Birth on Child Cognitive Outcomes in England. CEE Discussion Papers, Centre for the Economics of Education, LSE 0093, Centre for the Economics of Education, LSE.

Cunha, F., Heckman, J. J., and Lochner, L. (2006). *Interpreting the Evidence on Life Cycle Skill Formation*, volume 1 of *Handbook of the Economics of Education*, chapter 12, pages 697–812. Elsevier.

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 Stabile, M. (2003). Socioeconomic Status And Child Health: Why Is The Relationship Stronger For Older Children? *American Economic Review*, 93(5):1813–1823.

Cutler, D. M. and Ly, D. P. (2011). The (Paper)Work of Medicine: Understanding International Medical Costs. *Journal of Economic Perspectives*, 25(2):3–25.

Datar, A., Kilburn, M., and Loughran, D. (2010). Endowments and Parental Investments in Infancy and Early Childhood. *Demography*, 47(1):145–162.

Datta Gupta, N. and Simonsen, M. (2010). Non-Cognitive Child Outcomes and Universal High Quality Child Care. *Journal of Public Economics*, 94(1):30–43.

DeCicca, P. and Smith, J. (2013). The Long-run Impacts of Early Childhood Education: Evidence from a Failed Policy Experiment. *Economics of Education Review*, 36(C):41–59.

Dee, T. S. and Sievertsen, H. H. (2015). The Gift of Time? School Starting Age and Mental Health. Working Paper No. 21610, National Bureau of Economic Research.

Deming, D. and Dynarski, S. (2008). The Lengthening of Childhood. *Journal of Economic Perspectives*, 22(3):71–92.

Deming, D. J. (2015). The Growing Importance of Social Skills in the Labor Market. Working Paper No. 21473, National Bureau of Economic Research.

Dhuey, E. and Lipscomb, S. (2008). What Makes a Leader? Relative Age and High School Leadership. *Economics of Education Review*, 27(2):173–183.

Dhuey, E. and Lipscomb, S. (2010). Disabled or Young? Relative Age and Special Education Diagnoses in Schools. *Economics of Education Review*, 29(5):857–872.

Dickert-Conlin, S. and Elder, T. (2010). Suburban Legend: School Cutoff Dates and the Timing of Births. *Economics of Education Review*, 29(5):826–841.

Duckworth, A. L., Quinn, P. D., and Tsukayama, E. (2012). What No Child Left Behind leaves behind: The Roles of IQ and Self-control in Predicting Standardized Achievement Test Scores and Report Card Grades. *Journal of Educational Psychology*, 104(2).

Duckworth, A. L. and Seligman, M. E. P. (2008). Self-discipline Gives Girls the Edge: Gender in Self-discipline, Grades, and Achievement Test Scores. *Journal of Educational Psychology*, 98(1).

Duggan, M. and Scott Morton, F. (2006). The Distortionary Effects of Government Procurement: Evidence for Medicaid Prescription Drug Purchasing. *Quarterly Journal of Economics*, 121:1–30.

Dunn, A. and Shapiro, A. H. (2014). Do Physicians Possess Market Power? *Journal of Law and Economics*, 57(1):159–193.

Dunn, A. and Shapiro, A. H. (2015). Physician payments under health care reform. *Journal of Health Economics*, 39:89–105.

Dustmann, C., Cornelissen, T., and Trentini, C. (2013a). Early School Exposure, Testscores, and Noncognitive Outcomes. University College London, Department of Economics, mimeo.

Dustmann, C., Raute, A., and Schonberg, U. (2013b). Does Universal Child Care Matter? Evidence from a Large Expansion in Pre-School Education.

Einav, L., Finkelstein, A., and Polyakova, M. (2016). Private Provision of Social Insurance: Drug-Specific Price Elasticities and Cost Sharing in Medicare Part D. Working Paper No. 22277, National Bureau of Economic Research.

Elder, T. E. and Lubotsky, D. H. (2009). Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers. *Journal of Human Resources*, 44(1):641–683.

Ellis, R. P. and McGuire, T. G. (1986). Provider Behavior Under Prospective Reimbursement: Cost Sharing and Supply. *Journal of Health Economics*, 5(2):129–152.

Evans, W. N., Morrill, M. S., and Parente, S. T. (2010). Measuring Inappropriate Medical Diagnosis and Treatment in Survey Data: The Case of ADHD Among School-age Children. *Journal of Health Economics*, 29(5):657–673.

Feldstein, M. S. (1973). The Welfare Loss of Excess Health Insurance. *Journal of Political Economy*, 81(2):251–280.

Fitzpatrick, M. (2010). Preschoolers Enrolled and Mothers at Work? The Effects of Universal Prekindergarten. *Journal of Labor Economics*, 28(1):51–85.

Fortin, N., Chen, K., Oreopoulos, P., and Phipps, S. (2015). Young in Class: Implications for Inattentive/Hyperactive Behaviour of Canadian Boys and Girls. *Canadian Journal of Economics*, 48:1601–1634.

Fredriksson, P. and Ockert, B. (2006). Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labor Market Performance. Working Paper Series, IFAU - Institute for Evaluation of Labour Market and Education Policy 2006:12, IFAU - Institute for Evaluation of Labour Market and Education Policy.

Fryer, R. G. and Levitt, S. D. (2004). Understanding the Black-White Test Score Gap in the First Two Years of School. *The Review of Economics and Statistics*, 86(2):447–464.

Fryer, R. G. and Levitt, S. D. (2006). Testing for Racial Differences in the Mental Ability of Young Children. *American Law and Economics Review*, 8(2):249–281.

Fryer, R. G. and Levitt, S. D. (2013). Testing for Racial Differences in the Mental Ability of Young Children. *American Economic Review*, 103(2):981–1005.

Gathmann, C. and Sass, B. (2012). Taxing Childcare: Effects on Family Labor Supply and Children. Discussion Paper No. 6440, Institute for the Study of Labor (IZA).

Gauthier, A. H., Smeedeng, T. M., and Furstenberg, F. F. (2004). Are Parents Investing Less Time in Children? Trends in Selected Industrialized Countries. *Population and Development Review*, 30(4):647–671.

Gelber, A. M. and Mitchell, J. W. (2012). Taxes and Time Allocation: Evidence from Single Women and Men. *The Review of Economic Studies*, 79(3):863–897.

Geruso, M., Layton, T., and Wallace, J. (2016). Are All Managed Care Plans Created Equal? Evidence from Random Assignment to Medicaid HMOs. UT Austin, mimeo.

Gesme, D. H. and Wiseman, M. (2010). How to Negotiate With Health Care Plans. *Journal of Oncology Practice*, 6(4):220–222.

Government Accountability Office (2015). Medicare Physician Payment Rates: Better Data and Greater Transparency Could Improve Accuracy. Report GAO-15-434, GAO.

Gronau, R. (1977). Leisure, Home Production, and Work–the Theory of the Allocation of Time Revisited. *Journal of Political Economy*, 85(6):1099–1123.

Gruber, J., Kim, J., and Mayzlina, D. (1999). Physician Fees and Procedure Intensity: the Case of Cesarean Delivery. *Journal of Health Economics*, 18(4):473–490.

Guryan, J., Hurst, E., and Kearney, M. (2008). Parental Education and Parental Time with Children. *Journal of Economic Perspectives*, 22(3):23–46.

Hale, L., Berger, L. M., LeBourgeois, M. K., and Brooks-Gunn, J. (2011). A Longitudinal Study of Preschoolers' Language-based Bedtime Routines, Sleep duration, and Well-being. *Journal of Family Psychology*, 25(3):423–433.

Havnes, T. and Mogstad, M. (2011). Money for Nothing? Universal Child Care and Maternal Employment. *Journal of Public Economics*, 95(11):1455–1465.

Havnes, T. and Mogstad, M. (2015). Is Universal Child Care Leveling the Playing Field? *Journal of Public Economics*, 127:100–114.

Heckman, J. J. and Cunha, F. (2007). The Technology of Skill Formation. *American Economic Review*, 97(2):31–47.

Heckman, J. J. and Masterov, D. V. (2007). The Productivity Argument for Investing in Young Children. Working Paper No. 13016, National Bureau of Economic Research.

Heckman, J. J., Stixrud, J., and Urzua, S. (2006). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics*, 24(3):411–482.

Heineck, G. and Anger, S. (2010). The Returns to Cognitive Abilities and Personality Traits in Germany. *Labour Economics*, 17(3):535–546.

Herbst, C. M. and Tekin, E. (2014). Child Care Subsidies, Maternal Health, and Child-Parents Interactions: Evidence from Three Nationally Representative Datasets. *Health Economics*, 23(8):894–916.

Hill, C. R. and Stafford, F. P. (1974). Allocation of Time to Preschool Children and Educational Opportunity. *The Journal of Human Resources*, 9(3):323–341.

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development via Maternal Speech. *Child Development*, 74(5):1368–1378.

Horvath, H. (2015). Classroom Assignment Policies and Implications for Teacher Value-Added Estimation. Technical report, University College London, Institute of Education mimeo.

Hsin, A. (2012). Is Biology Destiny? Birth Weight and Differential Parental Treatment. *Demography*, 49(4):1385–1405.

Jacobson, M., Earle, C. C., Price, M., and Newhouse, J. P. (2010). How Medicare's Payment Cuts For Cancer Chemotherapy Drugs Changed Patterns Of Treatment. *Health Affairs*, 29:1391–1399.

Japel, C., Tremblay, R. E., and Cote, S. (2005). Quality Counts! Assessing the Quality of Daycare Services based on the Quebec Longitudinal Study of Child Development. *IRPP Choices*, 11(5).

Kalb, G. and van Ours, J. C. (2014). Reading to Young Children: A Head-Start in Life? *Economics of Education Review*, 40:1–24.

Kalil, A., Ryan, R., and Corey, M. (2012). Diverging Destinies: Maternal Education and the Developmental Gradient in Time With Children. *Demography*, 49(4):1361–1383.

Kleiner, S. A., White, W. D., and Lyons, S. (2015). Market Power and Provider Consolidation in Physician Markets. *International Journal of Health Economics and Management*, 15(1):99–126.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics*, 8.

Kollo, J. and Hamori, S. (2011). Kinek Hasznal az Evvesztes?. Iskolakezdesi Kor es Tanuloi Teljesitmenyek Magyarorszagon [Who Gains by Postponed Schooling?. Age at Starting School and Achievement of Pupils in Hungary]. *Kozgazdasagi Szemle (Economic Review - monthly of the Hungarian Academy of Sciences)*, 58(2):133–157.

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. (2016). Targeted or Universal Coverage? Assessing Heterogeneity in the Effects of Universal Childcare. NBER Working Paper 22126, National Bureau of Economic Research.

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.

Leibowitz, A. (1974). Home Investments in Children. *Journal of Political Economy*, 82(2, Part 2):S111–S131.

Mahoney, N. (2015). Bankruptcy as Implicit Health Insurance. *American Economic Review*, 105(2):710–746.

McClellan, M. (2011). Reforming payments to healthcare providers: The key to slowing healthcare cost growth while improving quality? *Journal of Economic Perspectives*, 25(2):69–92.

McCrary, J. and Royer, H. (2011). The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth. *American Economic Review*, 101(1):158–95.

McEwan, P. J. and Shapiro, J. S. (2008). The Benefits of Delayed Primary School Enrollment: Discontinuity Estimates Using Exact Birth Dates. *Journal of Human Resources*, 43(1):1–29.

MedPAC (2011). Report to the Congress: Medicare and the Health Care Delivery System. Technical report, Medicare Payment Advisory Commission. Available at http://www.medpac.gov/documents/reports/Jun11_EntireReport.pdf (accessed October 7, 2015).

Mertz, G. J. (2004). Can You Negotiate Better Reimbursement? *Family Practice Management*, 11(9):31–34.

Mueller, G. and Plug, E. (2006). Estimating the Effect of Personality on Male-Female Earnings. *Industrial and Labor Relations Review*, 60(1):3–22.

Muhlenweg, A., Blomeyer, D., Stichnoth, H., and Laucht, M. (2012). Effects of Age at School Entry (ASE) on the Development of Non-cognitive Skills: Evidence from Psychometric Data. *Economics of Education Review*, 31(3):68–76.

Muhlenweg, A. M. and Puhani, P. A. (2010). The Evolution of the School-Entry Age Effect in a School Tracking System. *Journal of Human Resources*, 45(2).

Nakamura, E. and Steinsson, J. (2008). Five Facts About Prices: A Reevaluation of Menu Cost Models. *The Quarterly Journal of Economics*, 123(4):1415–1464.

Olea, J. L. M. and Pflueger, C. (2013). A Robust Test for Weak Instruments. *Journal of Business and Economic Statistics*, 31(3):358–369.

Osborne, M. (2005). How Important is Your Personality? Labor Market Returns to Personality for Women in the US and UK. *Journal of Economic Psychology*, 26(6):827–841.

Osborne, M., Gintis, H., and Bowles, S. (2001). The Determinants of Earnings: A Behavioral Approach. *Journal of Economic Literature*, 39(4):1137–1176.

Price, J. (2008). Parent-Child Quality Time: Does Birth Order Matter? *Journal of Human Resources*, 43(1):240–265.

Price, J. (2012). The Effects of Parental Time Investments: Evidence from Natural Within-Family Variation. *Brigham Young University Working Paper*.

Puhani, P. A. and Weber, A. (2007). Does the Early Bird Catch the Worm? *Empirical Economics*, 32(2-3):359–386.

Rosenzweig, M. R. and Wolpin, K. I. (1988). Heterogeneity, Intrafamily Distribution, and Child Health. *Journal of Human Resources*, 23(4):437–461.

Rubinstein, Y. and Heckman, J. J. (2001). The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *American Economic Review*, 91(2):145–149.

Smets, F. and Wouters, R. (2003). An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association*, 1(5):1123–1175.

Smets, F. and Wouters, R. (2007). Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 97(3):586–606.

Starc, A. and Town, R. J. (2015). Internalizing Behavioral Externalities: Benefit Integration in Health Insurance. Working Paper No. 21783, National Bureau of Economic Research.

Stipek, D. (2002). At What Age Should Children Enter Kindergarten? A Question for Policy Makers and Parents. *Social Policy Report*, 16(2):3–17.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, 20(4):518–529.

Tekin, E. (2007). Childcare Subsidies, Wages, and Employment of Single Mothers. *The Journal of Human Resources*, 42(2):453–487.

United Food and Commercial Workers Union, T. (2002). UFCW 789 Grocery Contract. Available online at http://www.irle.berkeley.edu/library/pdf/0147.pdf.

Vago, I. (2005). Felfele Terjeszkedo Ovodaztatas - Stagnalo Hozzaferes. *Educatio*, 4:742–760.

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) - Scope, Evolution and Enhancements. *Schmollers Jahrbuch : Journal of Applied Social Science Studies / Zeitschrift fur Wirtschafts- und Sozialwissenschaften*, 127(1):139–169.

Weisleder, A. and Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, 24(11):2143–2152.

White, C. (2013). Contrary To Cost-Shift Theory, Lower Medicare Hospital Payment Rates For Inpatient Care Lead to Lower Private Payment Rates. *Health Affairs*, 32(5):935–943.

Winter, A. and Ray, N. (2008). Paying Accurately For Imaging Services In Medicare. *Health Affairs*, 27(6):1479–1490.

Wrohlich, K. (2006). Labor Supply and Child Care Choices in a Rationed Child Care Market. IZA Discussion Paper 2053, Institute for the Study of Labor (IZA).

Zuvekas, S. H. and Cohen, J. W. (2016). Fee-For-Service, While Much Maligned, Remains The Dominant Payment Method For Physician Visits. *Health Affairs*, 35(3):411–414.

Appendix A

Appendix to Chapter 1

A.1 Solving the Utility Maximization Problem

The Lagrangian is the following:

$$\begin{aligned}\mathcal{L} &= \frac{\beta_K}{\rho_K} log\left((\gamma_K T_K)^{\rho_K} + X_K^{\rho_K}\right) + \frac{\beta_H}{\rho_H} log\left((\gamma_H T_H)^{\rho_H} + X_H^{\rho_H}\right) + \frac{\beta_L}{\rho_L} log\left((\gamma_L T_L) + X_L^{\rho_L}\right) \\ &+ \mu\left[(w-m)\bar{T} + I^F - o_K\theta_K T_K - o_H\theta_H T_H - o_L\theta_L T_L - (\alpha X_K + X_H + X_L)\right],\end{aligned}$$

where $o_K = o_H = w - m$ and $o_L = w - m + n$.
Taking the first-order condition with respect to X_K and re-arranging yields

$$(\gamma_K T_K)^{\rho_K} = \frac{\beta_K}{\alpha\mu} X_K^{\rho_K - 1} - X_K^{\rho_K}, \tag{A.1}$$

or

$$T_K = \frac{1}{\gamma_K} X_K \left(\frac{\beta_K}{\alpha\mu} X_K^{-1} - 1\right)^{\frac{1}{\rho_K}}. \tag{A.2}$$

Taking the first-order condition with respect to T_K yields

$$\beta_K \gamma_K^{\rho_K} T_K^{\rho_K - 1} = \mu o_K \theta_K \left[(\gamma_K T_K)^{\rho_K} + X_K^{\rho_K}\right]. \tag{A.3}$$

Substituting (A.1) into (A.3) yields

$$\gamma_K^{\rho_K} T_K^{\rho_K - 1} = \frac{o_K\theta_K}{\alpha} X_K^{\rho_K - 1}. \tag{A.4}$$

Substituting (A.2) into (A.4) yields the expression for X_K solely as a function of the Lagrange multiplier μ and parameters:

$$X_K = \frac{\beta_K/(\alpha\mu)}{1 + \left(\frac{o_K\theta_K}{\gamma_K}\right)^{\frac{\rho_K}{\rho_K - 1}}}. \tag{A.5}$$

Note that from (A.5) $\frac{\beta_K}{\alpha\mu} X_K^{-1} - 1 = \gamma_K^{\frac{\rho_K}{1-\rho_K}} (o_K\theta_K)^{\frac{\rho_K}{\rho_K - 1}}$ in (A.2), so T_K can be also expressed solely as a function of the Lagrange multiplier μ and parameters:

$$T_K = \frac{(\beta_K/\mu)\,(o_K\theta_K)^{\frac{1}{\rho_K - 1}}\,(\alpha\gamma_K)^{\frac{\rho_K}{1-\rho_K}}}{1 + \left(\frac{o_K\theta_K}{\gamma_K}\right)^{\frac{\rho_K}{\rho_K - 1}}}. \tag{A.6}$$

Following the same steps for X_H, T_H and X_L, T_L one arrives at the following general form for optimal demands for goods and time as a function of μ and parameters, for $j = H, L$:

$$X_j = \frac{(\beta_j/\mu)}{1 + \left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j - 1}}} \text{ and } T_j = \frac{(\beta_j/\mu)\,(o_j\theta_j)^{\frac{1}{\rho_j - 1}}\,\gamma_j^{\frac{\rho_j}{1-\rho_j}}}{1 + \left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j - 1}}}. \tag{A.7}$$

Substituting (A.5), (A.6) and (A.7) back into the budget constraint yields an expression of μ only as a function of parameters: $\mu = \frac{1}{\bar{T}(w-m)+I^F}$.
Then, the Marshallian demand for child goods and child time are:

$$X_K^* = \frac{\frac{1}{\alpha}\beta_K I}{1 + \left(\frac{(w-m)\theta_K}{\alpha\gamma_K}\right)^{\frac{\rho_K}{\rho_K - 1}}} \text{ and } T_K^* = \frac{\beta_K I \left((w-m)\theta_K\right)^{\frac{1}{\rho_K - 1}} \left(\alpha\gamma_K\right)^{\frac{\rho_K}{1-\rho_K}}}{1 + \left(\frac{(w-m)\theta_K}{\alpha\gamma_K}\right)^{\frac{\rho_K}{\rho_K - 1}}}, \tag{A.8}$$

while the Marshallian demands for home production H and leisure L are:

$$X_j^* = \frac{\beta_j I}{1 + \left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j - 1}}} \text{ and } T_j^* = \frac{\beta_j I \left(o_j\theta_j\right)^{\frac{1}{\rho_j - 1}} \gamma_j^{\frac{\rho_j}{1-\rho_j}}}{1 + \left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j - 1}}}, \tag{A.9}$$

where $I = \bar{T}(w-m) + I^F$ is the potential income of the household, and the opportunity costs of the mother's home production and leisure time are $o_H = w - m$; $o_L = w - m + n$.
Note that optimal choices depend on schooling through the wage w and time efficiency γ_j. Therefore, total differentiation of optimal choices with respect to schooling leads to a decomposition into two key channels: a wage channel and a time efficiency channel.

A.2 Solving the Expenditure Minimization Problem

The Lagrangian is the following:

$$\begin{aligned} \mathcal{L} \;=\; & o_K\theta_K T_K + o_H\theta_H T_H + o_L\theta_L T_L + (\alpha X_H + X_L + X_K) \\ & -\lambda\left[\frac{\beta_K}{\rho^K} log\left((\gamma_K T_K)^{\rho_K} + X_K^{\rho_K}\right) + \frac{\beta_H}{\rho^H} log\left((\gamma_H T_H)^{\rho_H} + X_H^{\rho_H}\right) + \frac{\beta_L}{\rho^L} log\left((\gamma_L T_L)^{\rho_L} + X_L^{\rho_L}\right) - \bar{U}\right] \end{aligned}$$

where $o_K = o_H = w - m$ and $o_L = w - m + n$.
Taking the first-order condition with respect to X_K and re-arranging yields

$$(\gamma_K T_K)^{\rho_K} = \lambda\beta_K X_K^{\rho_K - 1} - X_K^{\rho_K}, \tag{A.10}$$

or

$$T_K = \frac{1}{\gamma_K} X_K \left(\lambda\beta_K X_K^{-1} - 1\right)^{\frac{1}{\rho_K}}. \tag{A.11}$$

Taking the first-order condition with respect to T_K yields

$$\beta_K \gamma_K^{\rho_K} T_K^{\rho_K - 1} = \lambda^{-1} o_K\theta_K \left[(\gamma_K T_K)^{\rho_K} + X_K^{\rho_K}\right]. \tag{A.12}$$

Substituting (A.10) into (A.12) yields

$$\gamma_K^{\rho_K} T_K^{\rho_K - 1} = o_K\theta_K X_K^{\rho_K - 1}. \tag{A.13}$$

Substituting (A.11) into (A.13) yields the expression for X_K solely as a function of the Lagrange multiplier λ and parameters:

$$X_K = \frac{\lambda\beta_K}{1 + \left(\frac{o_K\theta_K}{\gamma_K}\right)^{\frac{\rho_K}{\rho_K - 1}}}. \tag{A.14}$$

Note that from (A.14) $\lambda\beta_K X_K^{-1} - 1 = \left(\frac{o_K\theta_K}{\gamma_K}\right)^{\frac{\rho_K}{\rho_K - 1}}$ in (A.11), so T_K can be also expressed solely as a function of the Lagrange multiplier λ and parameters:

$$T_K = \frac{\lambda\beta_K \left(o_K\theta_K\right)^{\frac{1}{\rho_K-1}} \gamma_K^{\frac{\rho_K}{1-\rho_K}}}{1+\left(\frac{o_K\theta_K}{\gamma_K}\right)^{\frac{\rho_K}{\rho_K-1}}}. \tag{A.15}$$

Following the same steps for X_H, T_H and X_L, T_L one arrives at the following general form for optimal demands for goods and time as a function of λ and parameters, for $j = K, H, L$:

$$X_j = \frac{\lambda\beta_j}{1+\left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j-1}}} \text{ and } T_j = \frac{\left(\lambda\beta_j\right)\left(o_j\theta_j\right)^{\frac{1}{\rho_j-1}} \gamma_j^{\frac{\rho_j}{1-\rho_j}}}{1+\left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j-1}}}. \tag{A.16}$$

Substituting (A.16) back into the budget constraint yields an expression of λ only as a function of parameters: $\lambda = exp\left\{\bar{U}-\phi_0-\phi_K-\phi_H-\phi_L\right\}$, where $\phi_0 \equiv \beta_K log\beta_K + \beta_H log\beta_H + \beta_L log\beta_L$, $\phi_j \equiv \beta_j \left(\frac{1-\rho_j}{\rho_j}\right) log\left(1+\left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j-1}}\right)$ for $j = K, H, L$.

Thus, general forms of Hicksian demands for goods and time as a function of parameters, for $j = K, H, L$ are:

$$X_j^H = \frac{\beta_j exp\left\{\bar{U}-\phi_0-\phi_K-\phi_H-\phi_L\right\}}{1+\left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j-1}}} \text{ and } T_j^H = \frac{\beta_j exp\left\{\bar{U}-\phi_0-\phi_K-\phi_H-\phi_L\right\}\left(o_j\theta_j\right)^{\frac{1}{\rho_j-1}} \gamma_j^{\frac{\rho_j}{1-\rho_j}}}{1+\left(\frac{o_j\theta_j}{\gamma_j}\right)^{\frac{\rho_j}{\rho_j-1}}}, \tag{A.17}$$

where ϕ-s are defined as above.

Let's look at $\frac{\partial T_K^H}{\partial o_K}$:

$$\frac{\partial T_K^H}{\partial o_K} = \frac{\beta_K\lambda o_K^{\frac{2}{\rho_K-1}}\theta_K^{\frac{1+\rho_K}{\rho_K-1}}\gamma_K^{\frac{2\rho_K}{1-\rho_K}}\left(\beta_K-\frac{\rho_K}{\rho_K-1}\right)}{\left(1+\left(\frac{o_K\theta_K}{\gamma_K}\right)^{\frac{\rho_K}{\rho_K-1}}\right)^2} + \frac{\beta_K\lambda o_K^{\frac{2-\rho_K}{\rho_K-1}}\theta_K^{\frac{1}{\rho_K-1}}\gamma_K^{\frac{\rho_K}{1-\rho_K}}\left(\frac{1}{\rho_K-1}\right)}{1+\left(\frac{o_K\theta_K}{\gamma_K}\right)^{\frac{\rho_K}{\rho_K-1}}}, \tag{A.18}$$

that is negative if (but not only if) $\beta_K - \frac{\rho_K}{\rho_K-1} < 0$.

A.3 Comparative Statics for the Marshallian Demands

A.3.1 How do the Marshallian demands for market goods depend on schooling?

Let us totally differentiate the Marshallian demands in (A.8) with respect to schooling, and decompose $\frac{dX_j^*}{dS}$ into a wage channel $\frac{\partial X_j^*}{\partial w}\frac{dw}{dS}$ and a time efficiency channel $\frac{\partial X_j^*}{\partial \gamma_j}\frac{d\gamma_j}{dS}$ (for $j = K, H$). First, to simplify notation, let $\psi_j \equiv \left(\frac{(w-m)\theta_j}{\alpha_j\gamma_j}\right)^{\frac{\rho_j}{\rho_j-1}}$, $\Phi_j \equiv \left(\left(w-m\right)\theta_j\right)^{\frac{1}{1-\rho_j}}\left(\alpha_j\gamma_j\right)^{\frac{\rho_j}{1-\rho_j}}$ and $\chi \equiv \frac{\partial I^F}{\partial w}$, $\kappa_j \equiv \frac{\partial I^F}{\partial \gamma_j}$. Second, assume that $\frac{dw}{dS} > 0$. Third, splitting the optimal good choices into two parts (one independent of the father's income and the second dependent on the father's income I^F) is helpful to disentangle the channels (derivatives) into a base term ("1"), the assortative matching term ("2") and a remainder term ("3"):

$$X_j^* = \frac{\frac{1}{\alpha_j}\beta_j\left(w-m\right)\bar{T}}{1+\psi_j} + \frac{\frac{1}{\alpha_j}\beta_j I^F}{1+\psi_j} \text{ for } j = K, H.$$

The partial derivative with respect to the wage w is

$$\frac{\partial X_j^*}{\partial w} = \underbrace{\frac{\frac{\beta_j}{\alpha}\bar{T}\left(1+\frac{1}{1-\rho_j}\psi_j\right)}{\left(1+\psi_j\right)^2}}_{\text{base term“1”}} + \underbrace{\frac{\frac{\beta_j}{\alpha}\chi}{1+\psi_j}}_{\text{matching term“2”}} + \underbrace{\frac{\frac{\rho_j}{1-\rho_j}\frac{\beta_j}{\alpha}I^F\frac{1}{w-m}\psi_j}{\left(1+\psi_j\right)^2}}_{\text{remainder term“3”}};$$

The base and matching terms are, unambiguously and independently of the sign of the substitution parameters, positive for both goods; the remainder term is negative for child goods if $\rho_K < 0$ and $\delta_K > 0$ and positive for home production goods if $\rho_H > 0$ and $\delta_H > 0$. If the remainder term for child goods is sufficiently small (that is true if $|\rho_j| \left(\frac{1-\theta_j}{\theta_j}\right)^{\rho_{Tj-1}} < 4$), then comparing two households with the same time efficiency γ for child human capital production, the one with higher wage is able to afford and buy more of the child goods. Higher-wage parents *ceteris paribus* are unambiguously more able to afford and buy home production goods.
The partial derivative with respect to the time efficiency γ is

$$\frac{\partial X_j^*}{\partial \gamma_j} = \underbrace{\frac{\frac{\rho_j}{\rho_j - 1}\frac{\beta_j}{\alpha}\bar{T}\left(\frac{w-m}{\gamma_j}\right)\psi_j}{\left(1+\psi_j\right)^2}}_{\text{base term “1”}} + \underbrace{\frac{\frac{\beta_j}{\alpha}\kappa_j}{1+\psi_j}}_{\text{matching term “2”}} + \underbrace{\frac{\frac{\rho_j}{\rho_j - 1}\frac{\beta_j}{\alpha}\bar{T}\left(\frac{I^F}{\gamma_j}\right)\psi_j}{\left(1+\psi_j\right)^2}}_{\text{remainder term “3”}};$$

The base and the remainder terms are positive for child goods if $\rho_K < 0$ and $\delta_K > 0$ and negative for home production goods if $\rho_H > 0$ and $\delta_H > 0$, while the matching term is unambiguously positive. If the signs are such, comparing two households with the same w, the one with higher time efficiency of child human capital will unambiguously buy more child goods.

A.3.2 How do the Marshallian demands for time depend on schooling?

Similarly to child goods before, let us totally differentiate the Marshallian demands in (A.8) with respect to schooling, and decompose $\frac{dT_j^*}{dS}$ into a wage channel $\frac{\partial T_j^*}{\partial w}\frac{dw}{dS}$ and a time efficiency channel $\frac{\partial T_j^*}{\partial \gamma_j}\frac{d\gamma_j}{dS}$ (for $j = K, H$). Let us keep the notation for $\psi_j, \Phi_j, \chi, \kappa_j$, let us assume that $\frac{dw}{dS} > 0$, and let us split the optimal time choices into two parts (one independent of and the second dependent on the father's income I^F):

$$T_j^* = \frac{\beta_j \bar{T}\left(w-m\right)\Phi_j}{1+\psi_j} + \frac{\beta_j I^F \Phi_j}{1+\psi_j} \text{ for } j = K, H.$$

The partial derivative with respect to the wage w is

$$\frac{\partial T_j^*}{\partial w} = \underbrace{\frac{\frac{\rho_j}{\rho_j - 1}\beta_j \bar{T}\Phi_j}{\left(1+\psi_j\right)^2}}_{\text{base term “1”}} + \underbrace{\frac{\beta_j \chi \Phi_j}{1+\psi_j}}_{\text{matching term “2”}} + \underbrace{\frac{\frac{\beta_j}{\rho_j - 1}I^F \frac{\Phi_j}{w-m}\left(\psi_j(1-\rho_j)+1\right)}{\left(1+\psi_j\right)^2}}_{\text{remainder term “3”}};$$

Consider first child time: if $\rho_K < 0$ and $\delta_K > 0$, the base and matching terms are positive and the remainder term is negative; while the matching term is always positive, the sign of the other two terms is driven by the complementarity between time and market goods in the production of child human capital. Provided that the remainder term is sufficiently small and time efficiency is kept fixed, higher-wage parents will spend more time with child; the intuition is that they buy more of the child goods, and to harvest those, by complementarity, they need to invest more time into their child, too. Regarding home production time, if $\rho_H > 0$ and $\delta_H > 0$, both the base term and the remainder term is negative; these signs are driven by substitutability between time and market goods in home production.
The partial derivative with respect to the time efficiency γ is

$$\frac{\partial T_j^*}{\partial \gamma_j} = \underbrace{\frac{\frac{\rho_j}{1-\rho_j}\beta_j \bar{T}\left(\frac{w-m}{\gamma_j}\right)\Phi_j}{\left(1+\psi_j\right)^2}}_{\text{base term ”1”}} + \underbrace{\frac{\beta_j \kappa_j \Phi_j}{1+\psi_j}}_{\text{matching term ”2”}} + \underbrace{\frac{\frac{\rho_j}{1-\rho_j}\beta_j I^F \frac{1}{\gamma_j}\Phi_j}{\left(1+\psi_j\right)^2}}_{\text{remainder term ”3”}};$$

The base and the remainder terms are negative for child time if $\rho_K < 0$ and $\delta_K > 0$ and positive for home production goods if $\rho_H > 0$ and $\delta_H > 0$, while the matching term is unambiguously positive.
The condition for the 'parental time-education gradient' puzzle to hold is the following:

Condition 1: $\frac{dT_K^M}{dS} = \frac{\partial T_K^M}{\partial w}\frac{dw}{dS} + \frac{\partial T_K^M}{\partial \gamma_K}\frac{d\gamma_j}{dS} > 0$ if

$$\underbrace{\frac{\frac{\rho_K}{\rho_K-1}\beta_K\bar{T}\Phi_K}{(1+\psi_j)^2}\left(\frac{dw}{dS}-\frac{w-m}{\gamma_K}\frac{d\gamma_K}{dS}\right)}_{\text{base term "1"}}+\underbrace{\frac{\beta_K\Phi_K}{1+\psi_j}\left(\chi\frac{dw}{dS}-\kappa_K\frac{d\gamma_K}{dS}\right)}_{\text{matching term "2"}}$$
$$+\underbrace{\frac{\beta_K I^F\Phi_K}{(1+\psi_j)^2}\left(\frac{1+\psi_K(1-\rho_K)}{(\rho_K-1)(w-m)}\frac{dw}{dS}-\frac{\rho_K}{(\rho_K-1)\gamma_K}\frac{d\gamma_K}{dS}\right)}_{\text{remainder term "3"}} > 0.$$

The term inside the bracket in the remainder term is negative if $\rho_K < 0$. Then, even if $\frac{dw}{dS} - \frac{w-m}{\gamma_K}\frac{d\gamma_K}{dS}$ (in the bracket inside the base term) is negative, the condition can hold if $\chi\frac{dw}{dS}$ is sufficiently large; i.e. if the assortative matching on wages, measured in the high-educated mothers' wage advantage, is sufficiently large.

A.3.3 How do Marshallian demands change after a decrease in the price of daycare?

Note that by assumption, the father's income I^F does not depend on m; thus, a marginal decrease in daycare price mirrors the effect of a marginal increase in wage, net of the matching term, and accounting for the dependence of $\alpha_K = \left(1+m^{\frac{1}{\rho_X-1}}\right)^{-1}$ on m. To ease notation, let $\alpha_K = \alpha$, since $\alpha_H = \alpha_L = 1$.

Consider first α and denote $\frac{\partial\left(\frac{1}{\alpha}\right)}{\partial(-m)}$ by a: $\frac{\partial\left(\frac{1}{\alpha}\right)}{\partial(-m)} \equiv a = \left(\frac{1}{1-\rho_X}\right)m^{\frac{2-\rho_X}{\rho_X-1}} > 0$. Then, if m decreases, the share of child books α decreases and the share of daycare $1-\alpha$ increases, as expected from the change in their relative prices. Additionally, denote $\frac{\partial\psi_K}{\partial(-m)}$ by ϱ : $\frac{\partial\psi_K}{\partial(-m)} \equiv \varrho = \frac{\rho_K}{\rho_K-1}\psi_K\left(\frac{1}{w-m}+\alpha a\right)$; and denote $\frac{\partial\Phi_K}{\partial(-m)}$ by υ : $\frac{\partial\Phi_K}{\partial(-m)} \equiv \upsilon = \frac{\psi_K}{(w-m)\theta_K}\left(\frac{\rho_K}{\rho_K-1}\left(\frac{1}{w-m}+\alpha a\right)-\frac{1}{w-m}\right)$.
Consider first optimal child good choices $X_K^* = \frac{\frac{1}{\alpha}\beta_K I}{1+\psi_K}$:

$$\frac{\partial X_K^*}{\partial(-m)} = \frac{\beta_K\frac{1}{\alpha}\left(\bar{T}(1+\psi_K)-I\varrho\right)}{(1+\psi_K)^2}+\frac{\beta_K a I}{1+\psi_K}; \tag{A.19}$$

total child goods increase after a fall in m if (but not only if) *Condition 2* holds:

Condition 2: $\bar{T}(1+\psi_K) - I\varrho > 0$, or if $\frac{1+\psi_K}{w-m+0.25w^F} > \frac{\rho_K}{\rho_K-1}\left(\frac{1}{w-m}+\alpha a\right)$.

If this condition holds, then child books αX_K^* increase, too: $\frac{\partial\alpha X_K^*}{\partial(-m)} = \frac{\beta_K\left(\bar{T}(1+\psi_K)-I\varrho\right)}{(1+\psi_K)^2} > 0$; and clearly daycare use $(1-\alpha)X_K^*$ increases, too; whether it holds or not, does not depend on the sign of ρ_K.
Then, consider optimal child time $T_K^* = \frac{\beta_K I\Phi_K}{1+\psi_K}$:

$$\frac{\partial T_K^*}{\partial(-m)} = \frac{\beta_K\Phi_K\left(\bar{T}(1+\psi_K)-I\varrho\right)}{(1+\psi_K)^2}+\frac{\beta_K\upsilon I}{1+\psi_K}; \tag{A.20}$$

if *2* holds, then the first part in (A.20) is positive; if *3* holds, then υ is positive, making the second part positive:

Condition 3: $\frac{\rho_K}{\rho_K-1}\left(\frac{1}{w-m}+\alpha a\right)-\frac{1}{w-m} \equiv \tilde{a} > 0$, or if $1+\alpha a(w-m) > \frac{\rho_{K-1}}{\rho_K}$.

Whether *3* holds or not, does depend on the sign of ρ_K. Suppose *2* holds; then, a fall in m generates a dominating income effect, so demand for child books and daycare increase; if, in addition, time and goods are complementary in child human capital production, then child time will increase, to harvest investments in X_K.
Consider second optimal home production goods; here, a marginal decrease in daycare price mirrors the effect of a marginal increase in wage, only net of the matching term. Since both the base and the remainder terms are positive, a fall in the price of daycare generates a dominating income effect, so demand for home production

goods unambiguously increases. Note that the base term is positive irrespective of the sign of ρ_H, while the sign of the remainder term hinges on the sign of ρ_H :

$$\frac{\partial X_H^*}{\partial(-m)} = \underbrace{\frac{\beta_H \bar{T}\left(1+\frac{1}{1-\rho_H}\psi_H\right)}{(1+\psi_H)^2}}_{\text{base term "1"}} + \underbrace{\frac{\frac{\rho_H}{1-\rho_H}\beta_H I^F \frac{1}{w-m}\psi_H}{(1+\psi_j)^2}}_{\text{remainder term "3"}};$$

If there is substitutability between the inputs in home production ($\rho_H > 0$), then home production time unambiguously decreases after a daycare price decrease:

$$\frac{\partial T_H^M}{\partial(-m)} = \underbrace{\frac{\frac{\rho_H}{\rho_H-1}\beta_j \bar{T}\Phi_j}{(1+\psi_j)^2}}_{\text{base term "1"}} + \underbrace{\frac{\frac{\beta_H}{\rho_H-1}I^F \frac{\Phi_H}{w-m}\left(\psi_H(1-\rho_H)+1\right)}{(1+\psi_H)^2}}_{\text{remainder term "3"}}.$$

A fall in m generates a dominating income effect, so demand for home production good increases; if, in addition, time and goods are substitutable in home production, then home production time will decrease.

A.3.4 How do good responses after a fall in daycare price depend on schooling?

First, consider the child good response $\frac{\partial X_K^*}{\partial(-m)}$, and decompose its dependence on schooling S to a wage channel $\left(\frac{\partial^2 X_K^*}{\partial(-m)\partial w}\frac{dw}{dS}\right)$ and a time efficiency channel $\left(\frac{\partial^2 X_K^*}{\partial(-m)\partial\gamma_K}\frac{d\gamma_K}{dS}\right)$. For the following arguments note that $\alpha, a, \frac{\rho_K}{\rho_K-1}$ are all less than 1, and ψ_K is relatively large; for the sign-indications, I assume $\rho_K < 0$.
To assess the wage channel, the partial derivative with respect to w is:

$$\frac{\partial^2 X_K^*}{\partial(-m)\,\partial w} = \underbrace{\frac{\beta_K a\left(\bar{T}+\chi\right)}{1+\psi_K}}_{1)\oplus} \underbrace{-\frac{\beta_K\left(aI+\frac{1}{\alpha}\bar{T}\right)\left(\frac{\rho_K}{\rho_K-1}\right)\frac{\psi_K}{w-m}}{(1+\psi_K)^2}}_{2)\ominus} \underbrace{-\frac{\frac{\beta_K}{\alpha}\left(\bar{T}+\chi\right)\varrho}{(1+\psi_K)^2}}_{3)\ominus} \underbrace{-\frac{\frac{\beta_K}{\alpha}\left(\frac{\rho_K}{\rho_K-1}\right)\frac{\psi_K}{w-m}\tilde{a}}{(1+\psi_K)^2}}_{4)\ominus} \underbrace{+\frac{2\frac{\beta_K}{\alpha}I\varrho\left(\frac{\rho_K}{\rho_K-1}\right)\frac{\psi_K}{w-m}}{(1+\psi_K)^3}}_{5)\oplus}.$$

Putting 1) and 3) together yields an expression being negative, if $|\rho_K|$ is sufficiently large:

$$\frac{\beta_K a\left(\bar{T}+\chi\right)}{1+\psi_K} - \frac{\frac{\beta_K}{\alpha}\left(\bar{T}+\chi\right)\varrho}{(1+\psi_K)^2} = \frac{\beta_K\left(\bar{T}+\chi\right)}{1+\psi_K}\left\{a\left(1+\psi_K\right)-\frac{1}{\alpha}\varrho\right\} = a+\psi_K\left(\frac{1}{1-\rho_K}\left(a+\rho_K\frac{1}{\alpha}\frac{1}{w-m}\right)\right).$$

Adding 2) and 5) yields an expression being negative, if $\rho_K < 0$, since likely $\left(aI+\frac{1}{\alpha}\bar{T}\right)(1+\psi_K)-\frac{2}{\alpha}I\varrho > 0$:[135]

$$-\frac{\beta_K\left(aI+\frac{1}{\alpha}\bar{T}\right)\left(\frac{\rho_K}{\rho_K-1}\right)\frac{\psi_K}{w-m}}{(1+\psi_K)^2} + \frac{2\frac{\beta_K}{\alpha}I\varrho\left(\frac{\rho_K}{\rho_K-1}\right)\frac{\psi_K}{w-m}}{(1+\psi_K)^3} = \frac{-\beta_K\left(1+\psi_K\right)\left(\frac{\rho_K}{\rho_K-1}\right)}{(1+\psi_K)^3\left(\frac{\psi_K}{w-m}\right)^{-1}}\left\{\left(aI+\frac{1}{\alpha}\bar{T}\right)(1+\psi_K)-\frac{2}{\alpha}I\varrho\right\}.$$

Thus, $\left(\frac{\partial^2 X_K^*}{\partial(-m)\partial w} < 0\right)$ if $\rho_K < 0$, i.e. if inputs in child human capital production are complementary: holding time efficiency constant, higher-wage parents will increase child goods less than lower-wage parents do.
Consider then the time efficiency channel:

$$\frac{\partial^2 X_K^*}{\partial(-m)\,\partial\gamma_K} = \underbrace{\frac{\beta_K\left(aI+\frac{1}{\alpha}\bar{T}\right)\left(\frac{\rho_K}{\rho_K-1}\right)\psi_K\gamma_K}{(1+\psi_K)^2}}_{1)\oplus} \underbrace{+\frac{\frac{\beta_K}{\alpha}\left(\frac{\rho_K}{\rho_K-1}\right)^2\psi_K\gamma_K I\left(\frac{1}{w-m}+\alpha a\right)}{(1+\psi_K)^2}}_{2)\oplus} \underbrace{+\frac{2\frac{\beta_K}{\alpha}I\varrho\left(\frac{\rho_K}{1-\rho_K}\right)\frac{\psi_K}{w-m}}{(1+\psi_K)^3}}_{3)\ominus}.$$

[135] To see this, consider $Ia\left(1+\psi_K\right)-\frac{1}{\alpha}\varrho+\frac{1}{\alpha}\bar{T}\left(1+\psi_K\right)-\frac{1}{\alpha}I\varrho$. The first term is positive by previous arguments, and the second can be written as $\frac{1}{\alpha}\left(\bar{T}+\psi_K\left(\bar{T}-\frac{\rho_K}{\rho_K-1}\left(\frac{1}{w-m}+\alpha a\right)\right)\right)$.

Adding 2) and 3) yields an expression being positive, if $\rho_K < 0$, since, by previous arguments, $\left(aI + \frac{1}{\alpha}\bar{T}\right)(1+\psi_K) - \frac{2}{\alpha}I\varrho > 0$. Thus, $\left(\frac{\partial^2 X_K^*}{\partial(-m)\partial\gamma_K} > 0\right)$ if $\rho_K < 0$, i.e. if inputs in child human capital production are complementary: holding wage constant, parents with higher time efficiency will increase child goods more.
Second, consider home production goods. The partial derivative of the response with respect to wage is:

$$\frac{\partial^2 X_H^*}{\partial(-m)\,\partial w} = \frac{\left(\frac{\rho_H}{\rho_H-1}\right)\beta_H\bar{T}\Phi_H}{\left(1+o_H\Phi_H\right)^3}\left\{\Phi_H\left(w-m\right)\left(\frac{1}{\rho_H-1}\right)+\frac{1}{1-\rho_H}-2\right\},$$

while the partial derivative with respect to time efficiency γ_H is:

$$\frac{\partial^2 X_H^*}{\partial(-m)\,\partial\gamma_H} = \frac{\left(\frac{\rho_H}{1-\rho_H}\right)\beta_H\bar{T}\Phi_H\frac{w-m}{\gamma_H}}{\left(1+o_H\Phi_H\right)^3}\left\{\left(\frac{1}{1-\rho_H}-2\right)+\left(w-m\right)\Phi_H\left(\frac{1}{\rho_H-1}\right)\right\}.$$

Suppose that $\delta_H > 0, \rho_H > 0$. Then, the wage channel is positive, while the time efficiency channel is negative for if $\rho_H < \frac{1}{2}$; i.e. if inputs are substitutable, then, *ceteris paribus*, higher-wage parents will increase home production time more, while parents with higher time efficiency will increase home production goods less.

A.3.5 How do time responses after a fall in daycare price depend on schooling?

First, consider child time response $\frac{\partial T_K^*}{\partial(-m)}$, and decompose its dependence on schooling S to a wage channel $\left(\frac{\partial^2 T_K^*}{\partial(-m)\partial w}\frac{dw}{dS}\right)$ and a time efficiency channel $\left(\frac{\partial^2 T_K^*}{\partial(-m)\partial\gamma_K}\frac{d\gamma_K}{dS}\right)$.
To assess the wage channel, the partial derivative with respect to w is

$$\begin{aligned}\frac{\partial^2 T_K^*}{\partial(-m)\,\partial w} = {} & \frac{\beta_K\left(\overbrace{\bar{T}\frac{\partial\Phi_K}{\partial w}}^{1)\ominus}+\overbrace{\frac{\partial I}{\partial w}\upsilon}^{2)\oplus}+\overbrace{I\frac{\partial\upsilon}{\partial w}}^{\oplus}\right)}{1+\psi_K} - \overbrace{\frac{\beta_K\left(\bar{T}\Phi_K+I\upsilon\right)\left(\frac{\rho_K}{\rho_K-1}\right)\frac{\psi_K}{w-m}}{\left(1+\psi_K\right)^2}}^{\ominus}\\ & -\frac{\beta_K\left(\overbrace{\frac{\partial I}{\partial w}\Phi_K\varrho}^{3)\oplus}+\overbrace{I\frac{\partial\Phi_K}{\partial w}\varrho}^{4)\ominus}+\overbrace{I\Phi_K\frac{\partial\varrho}{\partial w}}^{\ominus}\right)}{\left(1+\psi_K\right)^2}+\overbrace{\frac{2\beta_K I\Phi_K\varrho\left(\frac{\rho_K}{\rho_K-1}\right)\frac{\psi_K}{w-m}}{\left(1+\psi_K\right)^3}}^{\oplus}.\end{aligned}$$

Adding 1) and 4) yields a negative expression if *2* holds: $\frac{\beta_K\left(\frac{\partial\Phi_K}{\partial w}\right)\left(\bar{T}(1+\psi_K)-I\Phi_K\right)}{(1+\psi_K)^2} < 0$. Adding 2) and 3) yields

$$\frac{\beta_K\left(\bar{T}+\chi\right)\left[c\left(1+\psi_K\right)-\Phi_K\left(\frac{\rho_K}{\rho_K-1}\right)\left(\frac{1}{w-m}+\alpha a\right)\right]}{\left(1+\psi_K\right)^2},\ \text{where } c \equiv \frac{1}{\left(w-m\right)^2\theta_K}\left(\frac{1}{\rho_K-1}+\frac{\rho_K}{\rho_K-1}\alpha a\left(w-m\right)\right).$$

Then, the following expression inside the squared bracket is likely negative:

$$c+\frac{\psi_K}{\left(w-m\right)^2\theta_K}\left(\frac{1}{\rho_K-1}+\left(\frac{\rho_K}{\rho_K-1}\right)\alpha a\left(w-m\right)-\left(w-m\right)^2\theta_K\Phi_K\left(\frac{\rho_K}{\rho_K-1}\right)\left(\frac{1}{w-m}+\alpha a\right)\right).$$

Assuming that $\frac{\beta_K I\frac{\partial\upsilon}{\partial w}}{1+\psi_K}$ and $\frac{2\beta_K I\Phi_K\varrho\left(\frac{\rho_K}{\rho_K-1}\right)\frac{\psi_K}{w-m}}{(1+\psi_K)^3}$ is sufficiently small and $\rho_K < 0$, $\frac{\partial^2 T_K^*}{\partial(-m)\partial w}$ will be negative; i.e. if time and market good inputs in child human capital production are complementary, holding time efficiency constant, parents with higher wage will increase child time less.
To assess the time efficiency channel, the partial derivative with respect to γ_K is

$$\frac{\partial^2 T_K^*}{\partial(-m)\partial\gamma_K} = \underbrace{\frac{\beta_K \bar{T}\left(\frac{\rho_K}{1-\rho_K}\right)\Phi_K\gamma_K}{1+\psi_K}}_{1)\ominus} + \underbrace{\frac{\beta_K I\left(\frac{\rho_K}{1-\rho_K}\right)\psi_K\gamma_K k}{1+\psi_K}}_{2)\ominus} - \underbrace{\frac{\beta_K I\left(\frac{\rho_K}{1-\rho_K}\right)\Phi_K\gamma_K\varrho}{(1+\psi_K)^2}}_{3)\oplus}$$
$$+ \underbrace{\frac{\beta_K\Phi_K\left(\frac{\rho_K}{\rho_K-1}\right)^2\psi_K\gamma_K\left(\frac{1}{w-m}+\alpha a\right)}{(1+\psi_K)^2}}_{4)\oplus} + \underbrace{\frac{2\beta_K I\Phi_K\varrho\left(\frac{\rho_K}{1-\rho_K}\right)\psi_K\gamma_K}{(1+\psi_K)^3}}_{5)\ominus}, \quad \text{where } k \equiv \frac{\tilde{a}}{(w-m)\theta_K}.$$

Adding 2) and 4) yields $\frac{\beta I\left(\frac{\rho_K}{1-\rho_K}\right)\psi_K\gamma_K\left[k(1+\psi_K)-\Phi_K\left(\frac{\rho_K}{\rho_K-1}\right)\left(\frac{1}{w-m}+\alpha a\right)\right]}{(1+\psi_K)^2}$, where the expression in squared bracket is likely negative. Adding 3) and 5) yields $\frac{\beta_K I\left(\frac{\rho_K}{\rho_K-1}\right)\Phi_K\gamma_K\varrho(1-\Phi_K)}{(1+\psi_K)^2} > 0$, excluding extremely low values of γ_K. Thus, assuming that $|\frac{\partial\Phi_K}{\partial\gamma_K}|$ in 1) is sufficiently small and $\rho_K < 0$, $\frac{\partial^2 T_K^*}{\partial(-m)\partial\gamma_K}$ will be positive; i.e. if time and market good inputs in child human capital production are complementary, holding wage constant, parents with higher time efficiency will increase child time more.

Consider now the two parts of the home production time response: $\frac{\partial T_H^*}{\partial(-m)} = \lambda_{1H} + \lambda_{2H}$, where $\lambda_{1j} \equiv \frac{\beta_j \bar{T}\frac{\rho_j}{\rho_j-1}\Phi_j}{(1+\psi_j)^2}$ and $\lambda_{2j} \equiv \frac{\frac{\beta_j}{\rho_j-1}I^F\frac{\Phi_j}{w-m}(\psi_j(1-\rho_j)+1)}{(1+\psi_j)^2}$. Let us totally differentiate the responses with respect to S, and decompose the response into a wage channel and a time efficiency channel: $\left(\frac{\partial\lambda_{1H}}{\partial w}\frac{dw}{dS} + \frac{\partial\lambda_{2H}}{\partial w}\frac{dw}{dS}\right) + \left(\frac{\partial\lambda_{1H}}{\partial\gamma_H}\frac{d\gamma_H}{dS} + \frac{\partial\lambda_{2H}}{\partial\gamma_H}\frac{d\gamma_H}{dS}\right)$.

If $\frac{1}{2} > \rho_H > 0$, the wage channel is positive: $\frac{\partial\lambda_{1H}}{\partial w} = \frac{\left(\frac{\rho_H}{(\rho_H-1)^2}\right)\beta_H\bar{T}\gamma_H^{\frac{\rho_H}{1-\rho_H}}(w-m)^{\frac{2}{\rho_H-1}}}{(1+\psi_H)^3}\left\{(1-2\rho_H)\gamma_H^{\frac{\rho_H}{1-\rho_H}} + \frac{1}{w-m}\right\}$.

Then, consider the time efficiency channel: $\frac{\partial\lambda_1}{\partial\gamma_H} = \frac{\left(\frac{\rho_H}{\rho_H-1}\right)\left(\frac{\rho_H}{1-\rho_H}\right)\beta_H\bar{T}\Phi_H\frac{1}{\gamma_H}}{(1+\psi_H)^3}\{1-o_H\Phi_H\}$. Note that if $\rho_H > 0$, $(w-m)^{\frac{\rho_H}{\rho_H-1}}$ is less than 1, so excluding extremely large values of γ_H leads to $1-o_H\Phi_H$ being positive. Child time λ_2 depends negatively on both the wage and the time efficiency γ_H. Thus,

$$\left(\underbrace{\frac{\partial\lambda_{1K}}{\partial w}}_{\oplus}\frac{dw}{dS} + \underbrace{\frac{\partial\lambda_{2K}}{\partial w}}_{\ominus}\frac{dw}{dS}\right) + \left(\underbrace{\frac{\partial\lambda_{1K}}{\partial\gamma_K}}_{\ominus}\frac{d\gamma_K}{dS} + \underbrace{\frac{\partial\lambda_{2K}}{\partial\gamma_K}}_{\ominus}\frac{d\gamma_K}{dS}\right). \tag{A.21}$$

If the expression in (A.21) is negative, then high-educated parents will decrease their child time more after a decrease in the price of daycare. Note that this expression is more likely to hold with the second part, λ_2 present: its presence essentially makes the wage channel weaker and the time efficiency channel stronger.

A.4 Discussion of Alternative Models

This section discusses alternative models (alternative to heterogeneity in time efficiency, the base model): whether they can generate the same pattern of a child time increase in response to a daycare price decrease.

1. Heterogeneity in utility parameters (or equivalently, discount rates): suppose that the heterogeneity in the model stems from heterogeneity in β_K, where β_K is larger for high-educated parents. Then, since $\frac{\partial T_K}{\partial(-m)}$ depends positively on β_K and negatively on $w-m$, this and the base model give the same prediction for the response of T_K to a decrease in daycare price (increase in$(-m)$). However, their prediction for the response of $\log T_K$ to a change in m is different, that helps testing for this alternative model.

Consider lnT_{Ki} for individual i, where the utility parameter β_K is also indexed by i: $\log T_{Ki} = \log \beta_{Ki} + \log I + \log \Phi_K - \log(1+\psi_K)$. I parametrize β_{Ki} as $\beta_{Ki} = S_i^{\delta_K} \varepsilon_{Ki}$, so $\log \beta_{Ki} = \delta_K \log S_i + \log \varepsilon_{Ki}$. Now, consider the response of $\log T_{Ki}$ to a change in m, that is actually independent of β_K :

$$\frac{\partial \log T_{Ki}}{\partial(-m)} = \frac{\bar{T}}{I} + \frac{1}{w-m}\left\{\frac{1}{\rho_K - 1} - \left(\frac{\rho_K}{\rho_K-1}\right)\left(\frac{\psi_K}{1+\psi_K}\right)\right\}.$$

The response depends on schooling only through the wage, unambiguously negatively:

$$\frac{\partial^2 \log T_K}{\partial(-m)\partial w} = -\left(\frac{1}{w-m+\frac{1}{4}}\right)^2 - \left(\frac{1}{w-m}\right)^2\left\{\frac{1}{\rho_K-1} - \left(\frac{\rho_K}{\rho_K-1}\right)\left(\frac{\psi_K}{1+\psi_K}\right)\right\} - \frac{\left(\frac{\rho_K}{\rho_K-1}\right)^2 \frac{\psi_K}{w-m}}{w-m} < 0.$$

This prediction can be tested by running a Difference-in-Differences model on lnT_K, by regressing lnT_K on the policy variable, time fixed effects and province fixed effects, and, if (i) $E(\log S_i | post_i = 1) = E(\log S_i | post_i = 0)$ for both the treated and control provinces[136], and (ii) $E(\log \varepsilon_{Ki} | post_i = 1) = E(\log \varepsilon_{Ki} | post_i = 0)$ for both the treated and control provinces, then the schooling channel can be differenced out, and the coefficient on the policy variable will give exactly the Difference-in-Difference estimate for $\log I + \log \Phi_K - \log(1+\psi_K)$. However, this estimate (the response of $\log T_K$ to a decrease in m) does not depend negatively on the policy variable.

2. Heterogeneity in the substitution parameters, say ρ_K : suppose that ρ_K is smaller in absolute value for high-educated parents. However, $\frac{\partial^2 X_K}{\partial(-m)\partial\rho_K} > 0$, i.e. low-educated parents with larger ρ_K (meaning more substitutability and less complementarity) will increase child goods more and high-educated parents with smaller ρ_K (meaning less substitutability and more complementarity) will increase child goods less. Also keeping ρ_K constant, child good increase is predicted to be smaller for high-educated parents, so the model unambiguously, through both channels predicts that they increase child goods less. However, Table2.1 and TableA.2 provide evidence that high-educated families increased their daycare use significantly more.

3. Model with Daycare Quality: suppose that the CES production function takes the following form: $X_K = [B_K^{\rho_X} + (QD)^{\rho_X}]^{\frac{1}{\rho_X}}$, where ρ_X is the substitution parameter between books B and daycare D, while Q denotes daycare quality. Denote the fraction of books α, and the fraction of daycare $1-\alpha$. Then, in optimum,

$$\frac{1-\alpha}{\alpha} = m^{\frac{1}{\rho_X - 1}} Q^{\frac{\rho_X}{1-\rho_X}} \Rightarrow \alpha = \left[1 + m^{\frac{1}{\rho_X-1}} Q^{\frac{\rho_X}{1-\rho_X}}\right]^{-1}.$$

Consider the empirically relevant case when $\rho_K < 0$ and how X_K^*, αX_K^* and $(1-\alpha) X_K^*$ change if daycare quality increases. First, α is increasing in Q if $\rho_X < 0$, e.g. if books and daycare are complementary (the more plausible assumption). Then, X_K and B increase, and D decreases if daycare quality increases:

$$\frac{\partial X_K^*}{\partial \alpha} > 0, \frac{\partial \alpha X_K^*}{\partial \alpha} = \frac{\left(\frac{\rho_K}{\rho_K-1}\right)\frac{\beta_K I}{\alpha_K}\psi_K}{(1+\psi_K)^2} > 0 \text{ and } \frac{\partial(1-\alpha)X_K^*}{\partial\alpha} = \frac{-\frac{\beta_K I}{\alpha^2}}{(1+\psi_K)^2}\left\{1+\psi_K\left(1+(1-\alpha)\left(\frac{\rho_K}{1-\rho_K}\right)\right)\right\} < 0.$$

Consider the response of child goods to a fall in m; note that $\frac{\partial\left(\frac{1}{\alpha_K}\right)}{\partial(-m)} = \left(\frac{1}{1-\rho_X}\right) m^{\frac{2-\rho_X}{\rho_X-1}} Q^{\frac{\rho_X}{1-\rho_X}} \equiv \bar{a} > 0$, thus

$$\frac{\partial \psi_K}{\partial(-m)} = \left(\frac{\rho_K}{\rho_K-1}\right)\left(\frac{(w-m)\theta_K}{\alpha_K \gamma_K}\right)^{\frac{1}{\rho_K-1}} \frac{\theta_K}{\gamma_K}\left(\frac{1}{\alpha} + (w-m)\bar{a}\right) \equiv \bar{\varrho} > 0, \text{ if } \rho_K < 0.$$

[136] This is plausible, as the policy had no impact on schooling.

Then, the response of total child goods to a fall in daycare price is:

$$\frac{\partial X_K^*}{\partial(-m)} = \frac{\beta_K}{(1+\psi_K)^2}\left\{I\bar{a}(1+\psi_K) + \frac{1}{\alpha}\left(\bar{T}(1+\psi_K) - I\bar{\varrho}\right)\right\} > 0 \text{ if } \bar{T}(1+\psi_K) - I\bar{\varrho} > 0.$$

If the last condition holds, then not only total child goods, but child books increase, too, in response to $\downarrow m$:

$$\frac{\partial \alpha X_K^*}{\partial(-m)} = \frac{\beta_K}{(1+\psi_K)^2}\left\{\bar{T}(1+\psi_K) - I\bar{\varrho}\right\} > 0 \text{ if } \bar{T}(1+\psi_K) - I\bar{\varrho} > 0.$$

Finally, if (but, again, not only if) that condition holds, then daycare unambiguously increases, too:

$$\frac{\partial(1-\alpha)X_K^*}{\partial(-m)} = \frac{(1-\alpha)\beta_K}{\alpha(1+\psi_K)^2}\left\{\bar{T}(1+\psi_K) - I\bar{\varrho}\right\} + \frac{\bar{a}\frac{\beta_K}{\alpha}I}{1+\psi_K} > 0 \text{ if } \bar{T}(1+\psi_K) - I\bar{\varrho} > 0. \quad \text{(A.22)}$$

Since the response in (A.22) depends positively on quality, if daycare quality decreases at the same time daycare price decreases, then parents will increase daycare quality less. If daycare quality dramatically decreases, and if high-educated parents care about daycare quality more, then they are predicted to increase daycare use considerably less, than low-educated parents increase it. However, this is not consistent with the data.

A.5 Appendix Tables

Table A.1: Effect of a Daycare Price Decrease on Daycare Use (Extensive Margin); Policy Impact for All and by Education

	in institutional care		**care in other's home**		**care in own home**		**in any care**	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	0.184***	0.131***	-0.030***	-0.003	-0.024***	-0.040***	0.131***	0.088***
	(0.0072)	(0.0145)	(0.0037)	(0.0078)	(0.0025)	(0.0072)	(0.0093)	(0.0189)
β_2 : policy		0.070***		-0.034***		0.025***		0.061***
×high-educ.		(0.0144)		(0.0116)		(0.0084)		(0.0201)
β_3 : college	0.048***	0.035***	0.084***	0.097***	0.021***	0.034***	0.153***	0.166***
	(0.014)	(0.012)	(0.012)	(0.015)	(0.003)	(0.007)	(0.011)	(0.017)
β_4 : university	0.076***	0.062***	0.066***	0.079***	0.054***	0.067***	0.195***	0.209***
	(0.011)	(0.010)	(0.008)	(0.010)	(0.007)	(0.009)	(0.008)	(0.013)
R^2	0.123	0.124	0.049	0.049	0.022	0.022	0.117	0.117
N	61,962	61,962	61,962	61,962	61,962	61,962	61,962	61,962
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	0.092	0.184	0.238	-0.030	0.108	-0.024	0.437	0.131
low-educ.	0.053	0.131	0.176	-0.003	0.083	-0.040	0.312	0.088
high-educ.	0.106	0.201	0.261	-0.037	0.117	-0.015	0.484	0.141
Panel C: P-values of Testing Coefficients								
$\beta_1+\beta_2=0$		0.000		0.000		0.000		0.000

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact, for all and by mother's education. The *p-values* in *Panel C* correspond to the test $\beta_1+\beta_2=0$.

Table A.2: Effect of a Daycare Price Decrease on daycare use (Intensive Margin); Policy Impact for All and by Education

	hours in daycare		1: hours in daycare>20	
PanelA: Regression Results				
	(1)	(2)	(3)	(4)
β_1 : policy	5.900***	3.683***	0.160***	0.113***
	(0.2495)	(0.5804)	(0.0080)	(0.0171)
β_2 : policy		3.095***		0.066***
×high-educ.		(0.7146)		(0.0198)
R^2	0.118	0.119	0.105	0.105
N	61,962	61,962	61,962	61,962
PanelB: Means and Policy Impacts				
all	12.026	5.900	0.277	0.160
low-ed.	8.326	3.683	0.186	0.113
high-ed.	13.424	6.778	0.311	0.173
PanelC: P-values of Testing Coefficients				
$\beta_1 + \beta_2 = 0$		0.000		0.000

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: NLSCY waves 1-7 (1994-2006), 0-5 years old children in two-parent families, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.3: Effect of a Daycare Price Decrease on Mother's Working Propensity and daycare use; Policy Impact for All and by Education

	working&daycare		not working&daycare		working&no-daycare		not working&no-daycare	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	0.105***	0.046**	0.030***	0.049***	-0.029***	-0.004	-0.106***	-0.091***
	(0.0087)	(0.0169)	(0.0012)	(0.0047)	(0.0063)	(0.0072)	(0.0074)	(0.0163)
β_2 : policy		0.083***		-0.025***		-0.035***		-0.023
×high-educ.		(0.0185)		(0.0057)		(0.0110)		(0.0179)
β_3 : college	0.161***	0.166***	-0.007	0.000	0.013*	0.010	-0.166***	-0.176***
	(0.014)	(0.018)	(0.005)	(0.003)	(0.007)	(0.010)	(0.009)	(0.014)
β_4 : university	0.198***	0.204***	-0.002	0.006	-0.007	-0.010	-0.189***	-0.200***
	(0.013)	(0.015)	(0.007)	(0.004)	(0.006)	(0.010)	(0.012)	(0.014)
R^2	0.120	0.120	0.021	0.022	0.021	0.021	0.092	0.092
N	61,496	61,496	61,496	61,496	61,496	61,496	61,496	61,496
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	0.395	0.105	0.042	0.030	0.218	-0.029	0.344	-0.106
high-ed.	0.268	0.046	0.044	0.049	0.214	-0.004	0.475	-0.091
low-ed.	0.443	0.129	0.041	0.024	0.220	-0.039	0.295	-0.114
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.000		0.000		0.000		0.000

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact, for all and by mother's education. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.4: Effect of a Daycare Price Decrease on Parents' Labor Supply (Extensive and Intensive Margin); Policy Impact for All and by Education

	mother working		**mother's hours**		**father working**		**father's hours**	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	0.055***	0.031**	1.269***	0.659	0.013***	0.002	-0.529***	-1.933***
	(0.0042)	(0.0110)	(0.1647)	(0.4883)	(0.0022)	(0.0086)	(0.1217)	(0.4679)
β_2 : policy		0.025*		0.633		0.010		1.816**
×high-educ.		(0.0130)		(0.6188)		(0.0111)		(0.6521)
β_3 : college	0.146***	0.151***	3.712***	3.915***	0.057***	0.065***	2.744***	3.661***
	(0.011)	(0.011)	(0.339)	(0.500)	(0.009)	(0.008)	(0.424)	(0.401)
β_4 : university	0.182***	0.188***	4.806***	5.057***	0.063***	0.073***	2.546***	3.546***
	(0.023)	(0.014)	(0.665)	(0.592)	(0.012)	(0.007)	(0.649)	(0.417)
R^2	0.083	0.084	0.117	0.118	0.041	0.042	0.034	0.035
N	698,490	698,490	698,490	698,490	698,490	698,490	698,490	698,490
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	0.622	0.055	17.929	1.269	0.889	0.013	39.975	-0.529
high-educ.	0.494	0.031	14.556	0.659	0.835	0.002	37.596	-1.933
low-educ.	0.683	0.056	19.520	1.292	0.914	0.012	41.097	-0.117
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.000		0.000		0.002		0.619

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: Census (1996,2001,2006), two-parent families with at least one 0-4 years old child, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact, for all and by mother's education. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.5: Effect of a Daycare Price Decrease on Parents' Labor Supply (Extensive Margin); Policy Impact for All and by Education

	mother working		**working part-time**		**father working**		**working part-time**	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	0.034***	0.010	0.006***	0.002	0.006**	-0.006	0.004***	0.002
	(0.0054)	(0.0066)	(0.0019)	(0.0052)	(0.0024)	(0.0053)	(0.0008)	(0.0017)
β_2 : policy		0.028***		0.009		0.015*		0.002
×high-educ.		(0.0094)		(0.0086)		(0.0082)		(0.0024)
β_3 : college	0.173***	0.164***	0.074***	0.080***	0.056***	0.054***	-0.000	-0.000
	(0.010)	(0.010)	(0.005)	(0.004)	(0.009)	(0.007)	(0.001)	(0.001)
β_4 : university	0.201***	0.193***	0.071***	0.078***	0.068***	0.066***	0.009***	0.009***
	(0.018)	(0.013)	(0.007)	(0.005)	(0.012)	(0.008)	(0.002)	(0.001)
R^2	0.077	0.078	0.022	0.022	0.043	0.045	0.002	0.002
N	187,141	187,141	187,141	187,141	187,141	187,141	187,141	187,141
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	0.599	0.034	0.197	0.006	0.870	0.006	0.032	0.004
low-educ.	0.471	0.010	0.154	0.002	0.821	-0.006	0.032	0.002
high-educ.	0.663	0.038	0.218	0.012	0.895	0.009	0.032	0.004
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.000		0.010		0.037		0.020

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: LFS (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact, for all and by mother's education. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.6: Effect of a Daycare Price Decrease on Parents' Labor Supply (Intensive Margin); Policy Impact for All and by Education

	mother's hours		father's hours	
Panel A: Regression Results				
	(1)	(2)	(3)	(4)
β_1 : policy	1.081***	0.414*	0.0001	-0.464
	(0.1716)	(0.2209)	(0.1313)	(0.2885)
β_2 : policy		0.756**		0.704
$\times$high-educ.		(0.3606)		(0.4384)
β_3 : college	5.160***	4.857***	2.318***	2.549***
	(0.396)	(0.357)	(0.429)	(0.367)
β_4 : university	6.506***	6.208***	2.406***	2.661***
	0.071***	0.078***	(0.605)	(0.475)
R^2	0.074	0.074	0.034	0.035
N	186,941	186,941	187,051	187,051
Panel B: Means and Policy Impacts				
	mean	*impact*	*mean*	*impact*
all	19.296	1.081	37.426	0.000
low-educ.	15.097	0.414	35.495	-0.464
high-educ.	21.376	1.170	38.383	0.368
Panel C: P-values of Testing Coefficients				
$\beta_1 + \beta_2 = 0$		0.000		0.236

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province$\times$post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: LFS (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact, for all and by mother's education. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.7: Effect of a Daycare Price Decrease on Mother's Child Time; Policy Impact for All and by Education

	average		0-15 hours		16-30 hours		31-60 hours	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	0.870***	0.591**	-0.021***	-0.012***	-0.003*	-0.0060	0.0176***	0.0080**
	(0.1052)	(0.2385)	(0.0010)	(0.0039)	(0.0016)	(0.0039)	(0.0018)	(0.0029)
β_2 : policy		0.677**		-0.015***		0.001		0.0128***
$\times$high-educ.		(0.2766)		(0.0052)		(0.0043)		(0.0024)
β_3 : college	1.175***	1.809***	-0.037***	-0.044***	0.015***	0.009***	0.026***	0.024***
	(0.374)	(0.143)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)
β_4 : university	-0.313	0.315*	-0.036***	-0.043***	0.042***	0.037***	0.056***	0.054***
	(0.293)	(0.164)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)
R^2	0.067	0.067	0.023	0.023	0.028	0.028	0.016	0.016
N	698,490	698,490	698,490	698,490	698,490	698,490	698,490	698,490
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	44.824	0.870	0.120	-0.021	0.140	-0.003	0.231	0.018
low-educ.	44.409	0.591	0.147	-0.012	0.123	-0.006	0.199	0.008
high-educ.	45.020	1.268	0.107	-0.015	0.149	-0.005	0.246	0.021
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.000		0.000		0.003		0.000

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province$\times$post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: Census (1996,2001,2006), two-parent families with at least one 0-4 years old child, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.8: Effect of a Daycare Price Decrease on Mother's Home Production Time; Policy Impact for All and by Education

	average		0-15 hours		16-30 hours		31-60 hours	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	-2.046***	-2.233***	0.053***	0.051***	-0.009***	-0.001	-0.028***	-0.028***
	(0.1295)	(0.2707)	(0.0021)	(0.0043)	(0.0012)	(0.0029)	(0.0009)	(0.0042)
β_2 : policy		0.528		-0.004		-0.010***		0.006
$\times$high-educ.		(0.3340)		(0.0056)		(0.0033)		(0.0060)
β_3 : college	-1.232***	-0.793***	0.007	-0.011**	0.031***	0.039***	-0.004	0.006*
	(0.345)	(0.262)	(0.008)	(0.004)	(0.001)	(0.004)	(0.005)	(0.003)
β_4 : university	-4.272***	-3.848***	0.053***	0.035***	0.059***	0.068***	-0.030***	-0.019***
	(0.575)	(0.347)	(0.014)	(0.007)	(0.005)	(0.003)	(0.009)	(0.005)
R^2	0.098	0.099	0.056	0.057	0.008	0.008	0.014	0.015
N	698,490	698,490	698,490	698,490	698,490	698,490	698,490	698,490
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	30.523	-2.046	0.296	0.053	0.281	-0.009	0.237	-0.028
low-educ.	32.403	-2.233	0.280	0.051	0.246	-0.001	0.245	-0.028
high-educ.	29.637	-1.805	0.304	0.047	0.297	-0.011	0.234	-0.022
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.000		0.000		0.000		0.000

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: Census (1996,2001,2006), two-parent families with at least one 0-4 years old child, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.9: Effect of a Daycare Price Decrease on Father's Child Time; Policy Impact for All and by Education

	average		0-15 hours		16-30 hours		31-60 hours	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	0.718***	0.442*	-0.027***	-0.015***	0.011***	0.003	0.019***	0.013***
	(0.1345)	(0.2203)	(0.0026)	(0.0036)	(0.0007)	(0.0015)	(0.0014)	(0.0025)
β_2 : policy		0.485*		-0.019***		0.012***		0.008**
$\times$high-educ.		(0.2410)		(0.0040)		(0.0018)		(0.0030)
β_3 : college	1.505***	1.714***	-0.045***	-0.051***	0.025***	0.029***	0.022***	0.019***
	(0.223)	(0.136)	(0.004)	(0.003)	(0.002)	(0.001)	(0.002)	(0.002)
β_4 : university	1.741***	1.947***	-0.071***	-0.077***	0.052***	0.056***	0.040***	0.038***
	(0.108)	(0.248)	(0.005)	(0.006)	(0.003)	(0.001)	(0.004)	(0.004)
R^2	0.019	0.019	0.017	0.017	0.006	0.006	0.007	0.007
N	698,490	698,490	698,490	698,490	698,490	698,490	698,490	698,490
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	24.104	0.718	0.447	-0.027	0.255	0.0110	0.167	0.019
low-educ.	23.203	0.442	0.485	-0.015	0.225	0.0026	0.145	0.013
high-educ.	24.530	0.927	0.429	-0.034	0.269	0.0134	0.177	0.021
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_3 = 0$		0.000		0.000		0.000		0.000

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: Census (1996,2001,2006), two-parent families with at least one 0-4 years old child, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.10: Effect of a Daycare Price Decrease on Father's Home Production Time; Policy Impact for All and by Education

	average		0-15 hours		16-30 hours		31-60 hours	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	-1.084***	-1.215***	0.036***	0.039***	-0.021***	-0.023***	-0.011***	-0.012***
	(0.0592)	(0.1875)	(0.0013)	(0.0043)	(0.0009)	(0.0020)	(0.0009)	(0.0017)
β_2 : policy		0.278		-0.008		0.005*		0.003
×high-educ.		(0.2602)		(0.0059)		(0.0022)		(0.0021)
β_3 : college	0.097	0.106	-0.005	-0.007***	0.015***	0.017***	-0.003	-0.001
	(0.201)	(0.115)	(0.006)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)
β_4 : university	-0.451**	-0.456***	0.011	0.009***	0.019***	0.020***	-0.015***	-0.013***
	(0.199)	(0.119)	(0.006)	(0.003)	(0.003)	(0.002)	(0.003)	(0.001)
R^2	0.016	0.016	0.012	0.012	0.004	0.004	0.005	0.005
N	698,490	698,490	698,490	698,490	698,490	698,490	698,490	698,490
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	14.115	-1.084	0.695	0.036	0.205	-0.021	0.072	-0.011
low-educ.	14.169	-1.215	0.698	0.039	0.189	-0.023	0.077	-0.012
high-educ.	14.089	-0.935	0.695	0.031	0.213	-0.018	0.070	-0.009
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_3 = 0$		0.000		0.000		0.000		0.000

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: Census (1996,2001,2006), two-parent families with at least one 0-4 years old child, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.11: Effect of a Daycare Price Decrease on Father's Time Use; Policy Impact for All and by Education

	work time		child time		home production time		leisure time	
Panel A: Regression Results								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
β_1 : policy	-0.217	6.649**	0.162	-1.828	0.514	3.506***	-0.459	-8.327***
	(1.2490)	(2.8744)	(0.1848)	(1.0935)	(0.3217)	(0.6273)	(0.9612)	(1.6261)
β_2 : policy		-9.382***		3.344**		-4.015***		10.053***
×high-educ.		(2.8045)		(1.3770)		(0.6165)		(1.7698)
β_3 : college	0.450	-2.784	1.152	3.170**	0.526	1.663***	-2.128	-2.049
	(1.930)	(2.718)	(0.984)	(1.219)	(0.524)	(0.325)	(1.398)	(1.635)
β_4 : university	-0.925	-4.277	0.926	2.995***	0.886	2.010***	-0.887	-0.728
	(1.275)	(2.656)	(0.685)	(0.885)	(0.622)	(0.574)	(0.840)	(1.465)
R^2	0.011	0.015	0.039	0.042	0.033	0.036	0.021	0.025
N	1,822	1,822	1,822	1,822	1,822	1,822	1,822	1,822
Panel B: Means and Policy Impacts								
	mean	*impact*	*mean*	*impact*	*mean*	*impact*	*mean*	*impact*
all	26.813	-0.217	6.345	0.162	9.508	0.514	57.333	-0.459
low-educ.	26.093	6.649	5.408	-1.828	8.675	3.506	59.824	-8.237
high-educ.	27.182	-2.733	6.824	1.516	9.935	-0.599	56.059	1.816
Panel C: P-values of Testing Coefficients								
$\beta_1 + \beta_2 = 0$		0.0066		0.0003		0.132		0.097

Note: *Panel A* shows the result of estimating the Difference-in-Differences models (2.8) and (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: GSS (1998,2005,2010), two-parent families with at least one 0-4 years old child, both parents at most 50 years old. In *Panel B,* for each outcome variable, the first column shows the weighted mean in the estimation sample, while the second column shows the estimated policy impact. The *p-values* in *Panel C* correspond to the test $\beta_1 + \beta_2 = 0$.

Table A.12: Effect of a Daycare Price Decrease on Mother's Labor Supply (Both Margins) by the Propensity of Mother's Working in the Absence of the Policy and by Education, Census

	1: mother working				**mother hours**			
education:	*high*	*high*	*low*	*low*	*high*	*high*	*low*	*low*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
γ_1 : policy	0.148***	0.090	0.047*	0.074**	1.077	-4.285	0.210	3.267***
	(0.026)	(0.059)	(0.024)	(0.035)	(2.287)	(2.688)	(0.961)	(1.132)
γ_2 : policy	-0.135***	0.043	-0.039	-0.184	0.097	16.189**	1.334	-15.652***
$\times$propensity	(0.036)	(0.157)	(0.046)	(0.121)	(3.371)	(6.818)	(1.926)	(3.354)
γ_3 : policy		-0.132		0.176		-11.846**		21.541***
$\times$propensity^2		(0.110)		(0.115)		(4.628)		(3.908)
γ_4 : propensity	0.687***	0.004	0.210***	0.363***	9.962***	-105.169***	-3.554**	-28.743***
	(0.107)	(0.267)	(0.044)	(0.088)	(3.446)	(11.373)	(1.561)	(2.963)
γ_5 : propensity^2		0.554*		-0.161		93.449***		26.366***
		(0.283)		(0.110)		(9.993)		(3.149)
R^2	0.046	0.046	0.066	0.066	0.127	0.130	0.090	0.091
N	468,055	468,055	230,432	230,432	464,397	464,397	228,529	228,529

Note: this table shows the result of estimating the Difference-in-Differences model (2.10). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: Census (1996,2001,2006), two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

Table A.13: Effect of a Daycare Price Decrease on Mother's Labor Supply (Both Margins) by the Propensity of Mother's Working in the Absence of the Policy and by Education, LFS

	1: mother working				**mother hours**			
education:	*high*	*high*	*low*	*low*	*high*	*high*	*low*	*low*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
γ_1 : policy	0.053***	-0.067	-0.025**	-0.052**	3.388***	6.933**	-1.967***	1.341
	(0.006)	(0.069)	(0.010)	(0.019)	(0.315)	(3.311)	(0.359)	(0.778)
γ_2 : policy	-0.021**	0.343	0.117***	0.294***	-3.815***	-13.536	7.205***	-12.702***
$\times$propensity	(0.008)	(0.238)	(0.020)	(0.081)	(0.441)	(10.640)	(0.836)	(3.348)
γ_3 : policy		-0.271		-0.252**		6.510		26.896**
$\times$propensity^2		(0.191)		(0.101)		(8.400)		(4.298)
γ_4 : propensity	1.431***	1.812***	0.698***	1.530***	70.292***	41.441***	12.699*	16.753*
	(0.246)	(0.179)	(0.138)	(0.298)	(6.944)	(6.415)	(6.873)	(8.881)
γ_5 : propensity^2		-0.289		-0.689***		22.071***		-4.561
		(0.211)		(0.212)		(5.556)		(4.766)
R^2	0.038	0.038	0.055	0.055	0.043	0.043	0.049	0.049
N	148,141	148,141	77,236	77,236	148,016	148,016	77,135	77,135

Note: this table shows the result of estimating the Difference-in-Differences model (2.10). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: LFS (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old.

Table A.14: Effect of a Daycare Price Decrease on Mother's Labor Supply (Extensive Margin) and Any daycare use; Policy Impact by the Propensity of Mother's Working in the Absence of the Policy and by Education

	1: mother working				**1: any daycare use**			
model:	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*
education:	*high*	*high*	*low*	*low*	*high*	*high*	*low*	*low*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
γ_1 : policy	0.203***	0.583***	-0.040	-0.008	0.213***	0.335***	0.059*	0.124**
	(0.016)	(0.080)	(0.042)	(0.053)	(0.031)	(0.059)	(0.032)	(0.055)
γ_2 : policy	-0.223***	-1.256***	0.159*	0.002	-0.137***	-0.365*	0.046	-0.236
×propensity	(0.023)	(0.266)	(0.084)	(0.220)	(0.041)	(0.192)	(0.059)	(0.231)
γ_3 : policy		0.683***		0.180		0.078		0.277
×propensity^2		(0.214)		(0.287)		(0.152)		(0.270)
γ_4 : propensity	0.804***	0.402	0.833**	0.897*	0.256**	-0.631**	0.392	-0.101
	(0.117)	(0.339)	(0.306)	(0.491)	(0.119)	(0.269)	(0.265)	(0.526)
γ_5 : propensity^2		0.317		-0.064		0.692***		0.440
		(0.249)		(0.299)		(0.200)		(0.331)
R^2	0.072	0.072	0.074	0.074	0.105	0.106	0.074	0.074
N	42,544	42,544	15,819	15,819	42,829	42,829	15,953	15,953

Note: this table shows the result of estimating the Difference-in-Differences model (2.10). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old.

Table A.15: Effect of a Daycare Price Decrease on Reading Propensity and Child Time; Policy Impact by the Propensity of Mother's Working in the Absence of the Policy and by Education

	1: reading often				**child time**			
model:	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*
education:	*high*	*high*	*low*	*low*	*high*	*high*	*low*	*low*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
γ_1 : policy	0.066***	0.059	0.047	-0.182*	10.028***	5.216**	-0.379	-2.521
	(0.014)	(0.121)	(0.046)	(0.097)	(2.158)	(2.396)	(2.258)	(3.715)
γ_2 : policy	-0.108***	-0.097	-0.085	0.981**	-11.381***	6.302	0.230	12.598
×propensity	(0.023)	(0.386)	(0.082)	(0.352)	(3.353)	(5.407)	(5.633)	(15.122)
γ_3 : policy		-0.002		-1.143***		-14.397***		-14.292
×propensity^2		(0.291)		(0.334)		(3.375)		(12.982)
γ_4 : propensity	0.255***	0.317	-0.620**	-0.036	-3.342	8.401	-8.654**	-20.453*
	(0.055)	(0.311)	(0.226)	(0.387)	(3.647)	(8.355)	(3.513)	(10.993)
γ_5 : propensity^2		-0.048		-0.496*		-9.769*		15.611
		(0.247)		(0.283)		(5.455)		(12.706)
R^2	0.040	0.040	0.029	0.029	0.176	0.179	0.148	0.149
N	42,198	42,198	15,638	15,638	1,426	1,426	556	556

Note: this table shows the result of estimating the Difference-in-Differences model (2.10). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, and GSS (1998,2005,2010), two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

Table A.16: Effect of a Daycare Price Decrease on Home Production and Leisure Time; Policy Impact by the Propensity of Mother's Working in the Absence of the Policy and by Education

	home production time				**leisure time**			
model:	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*
education:	*high*	*high*	*low*	*low*	*high*	*high*	*low*	*low*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
γ_1 : policy	-4.657	-5.671	6.206**	5.150*	-5.192*	-0.468	-2.642*	-1.928
	(2.724)	(5.340)	(2.527)	(2.729)	(2.569)	(2.158)	(1.297)	(2.941)
γ_2 : policy	6.323	10.994	-10.762*	-3.450	4.342	-15.116**	13.453***	6.162
$\times$propensity	(3.972)	(16.821)	(5.627)	(8.445)	(3.464)	(6.380)	(2.868)	(15.437)
γ_3 : policy		-4.142		-9.087		16.594**		10.090
$\times$propensity^2		(12.226)		(9.765)		(6.360)		(17.757)
γ_4 : propensity	-4.687	-11.841	-10.276*	-10.099	-2.895	6.934	12.437***	0.728
	(3.124)	(18.517)	(5.275)	(8.742)	(2.898)	(7.147)	(3.176)	(13.671)
γ_5 : propensity^2		6.048		-0.327		-8.391		15.733
		(5.455)		(12.706)		(15.423)		(7.944)
R^2	0.176	0.179	0.148	0.149	0.013	0.013	0.009	0.005
N	1,426	1,426	556	556	1,426	1,426	556	556

Note: this table shows the result of estimating the Difference-in-Differences model (2.10). Standard errors are in parentheses and are clustered at the (province$\times$post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: GSS (1998,2005,2010), two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

Table A.17: Effect of a Daycare Price Decrease on Maternal Mental Health and Parenting Scores; Policy Impact by the Propensity of Mother's Working in the Absence of the Policy and by Education

	depression score				**hostile parenting score**			
model:	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*
education:	*high*	*high*	*low*	*low*	*high*	*high*	*low*	*low*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
γ_1 : policy	-0.424***	-1.252***	0.127	-0.493**	-0.279***	-0.442***	-0.222***	-0.630***
	(0.045)	(0.230)	(0.089)	(0.186)	(0.063)	(0.106)	(0.069)	(0.167)
γ_2 : policy	0.722***	3.336***	0.075	3.155***	0.581***	0.902**	1.034***	2.949***
$\times$propensity	(0.069)	(0.751)	(0.162)	(0.787)	(0.079)	(0.317)	(0.088)	(0.736)
γ_3 : policy		-1.985***		-3.537***		-0.125		-2.075**
$\times$propensity^2		(0.578)		(0.893)		(0.255)		(0.894)
γ_4 : propensity	0.447***	-1.345	-0.920**	-2.246**	0.450	1.526***	-0.582	0.243
	(0.154)	(0.801)	(0.439)	(1.025)	(0.277)	(0.378)	(0.446)	(0.814)
γ_5 : propensity^2		1.388**		1.307		-0.840***		-0.688
		(0.572)		(0.903)		(0.261)		(0.834)
R^2	0.011	0.012	0.015	0.016	0.023	0.023	0.031	0.032
N	42,829	42,829	15,953	15,953	42,829	42,829	15,953	15,953

Note: this table shows the result of estimating the Difference-in-Differences model (2.10). Standard errors are in parentheses and are clustered at the (province$\times$post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old.

Table A.18: Effect of a Daycare Price Decrease on Child Outcomes; Policy Impact by the Propensity of Mother's Working in the Absence of the Policy and by Education

	hyperactivity score				aggression score			
model:	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*	*linear*	*quadratic*
education:	*high*	*high*	*low*	*low*	*high*	*high*	*low*	*low*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
γ_1 : policy	-0.247***	0.701***	0.099	0.300*	-0.218***	0.030	0.253***	-0.318*
	(0.045)	(0.099)	(0.123)	(0.161)	(0.028)	(0.107)	(0.063)	(0.160)
γ_2 : policy	0.509***	-2.033***	-0.211	-1.176*	0.434***	-0.501	-0.188	2.505***
$\times$propensity	(0.062)	(0.318)	(0.229)	(0.566)	(0.040)	(0.313)	(0.129)	(0.681)
γ_3 : policy		1.660***		1.070*		0.804***		-2.931***
$\times$propensity^2		(0.253)		(0.548)		(0.238)		(0.712)
γ_4 : propensity	0.228	-0.993**	-0.323	-0.416	0.544***	2.213***	-0.321	0.655
	(0.182)	(0.360)	(0.471)	(0.781)	(0.124)	(0.311)	(0.454)	(0.677)
γ_5 : propensity^2		0.962***		0.053		-1.299***		-0.800
		(0.192)		(0.491)		(0.269)		(0.518)
R^2	0.017	0.018	0.021	0.021	0.040	0.041	0.042	0.043
N	42,829	42,829	15,953	15,953	42,829	42,829	15,953	15,953

Note: this table shows the result of estimating the Difference-in-Differences model (2.10). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old.

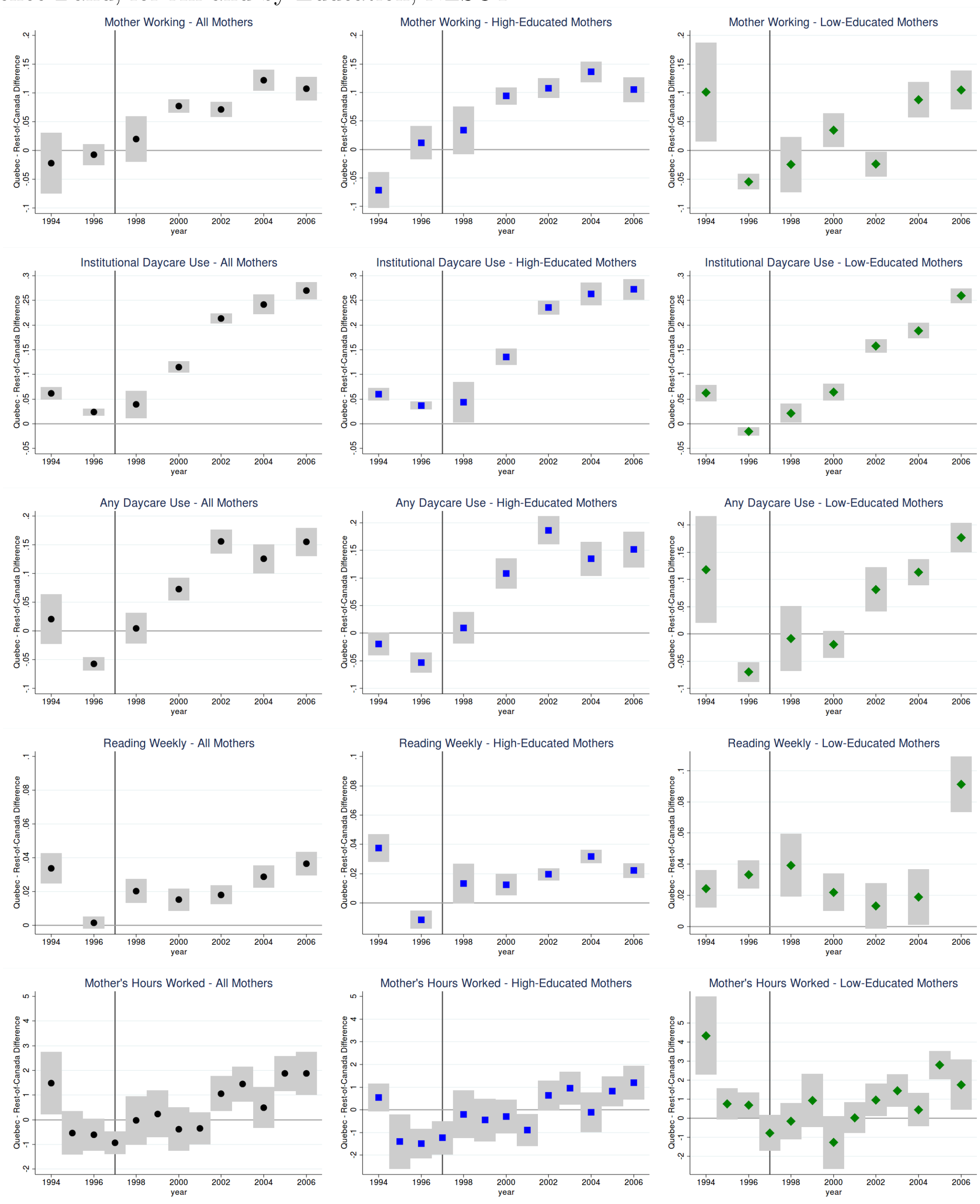
Table A.19: Effect of a Daycare Price Decrease on Household Expenditures; Policy Impact by the Propensity of Mother's Working in the Absence of the Policy

	food(%)		food - restaurant(%)		daycare(%)	
γ_1 : policy	0.817***	0.989***	0.322***	0.654*	-0.126**	1.867***
	(0.114)	(0.313)	(0.057)	(0.319)	(0.047)	(0.217)
γ_2 : policy		-1.046*		-0.175		-3.156***
$\times$propensity		(0.500)		(0.436)		(0.287)
γ_4 : propensity		-1.402		0.625		0.274
		(1.402)		(0.394)		(0.632)
R^2	0.13	0.13	0.04	0.04	0.04	0.04
N	22,725	22,725	22,725	22,725	22,725	22,725

Note: this table shows the result of estimating the Difference-in-Differences model (2.10). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: SHS 1986, 1992, 1996-2009, two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

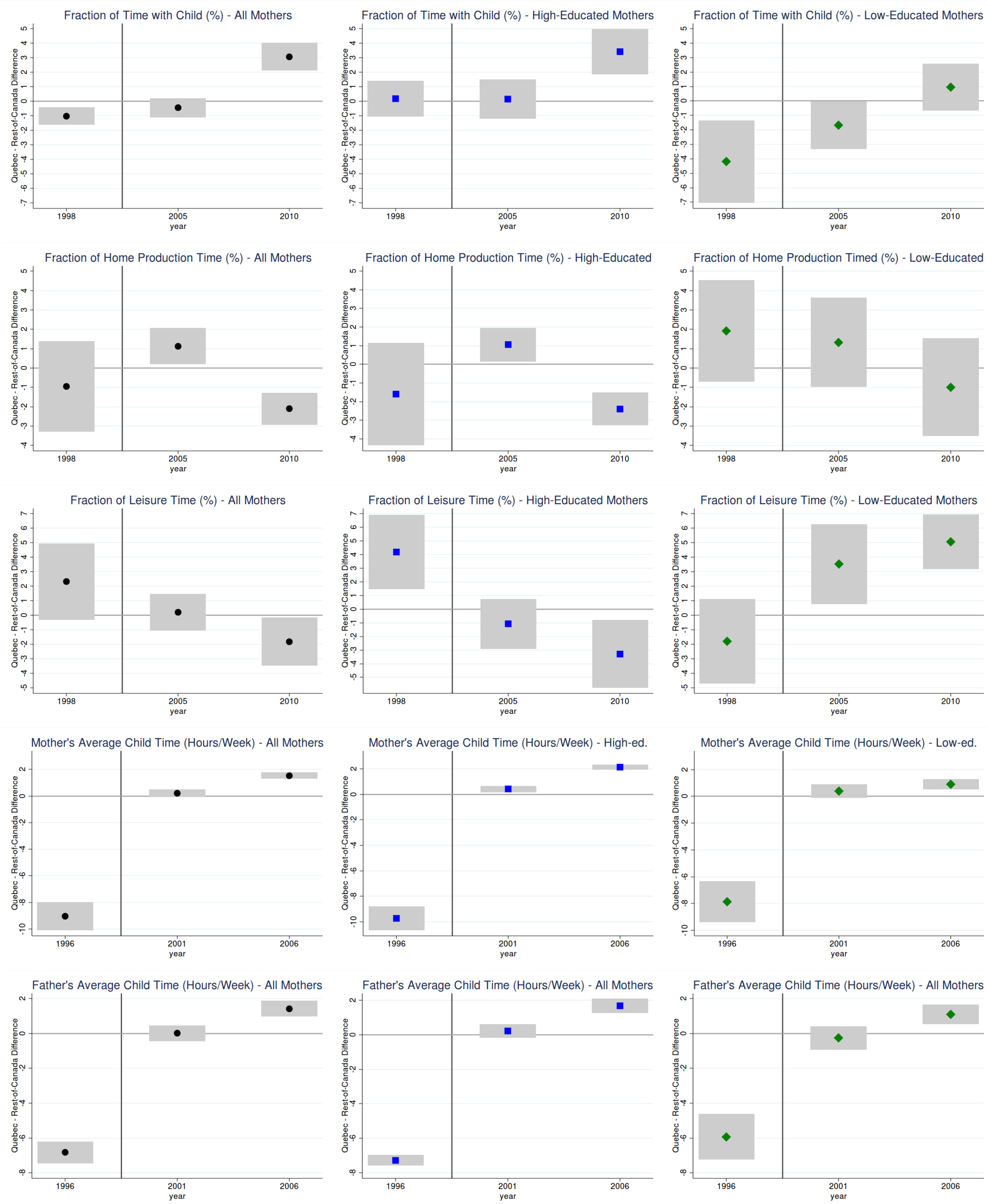
A.6 Appendix Figures

Figure A.1: Estimated Difference between Quebec and the Rest-of-Canada across Years, with a 95% Confidence Band, for All and by Education, NLSCY



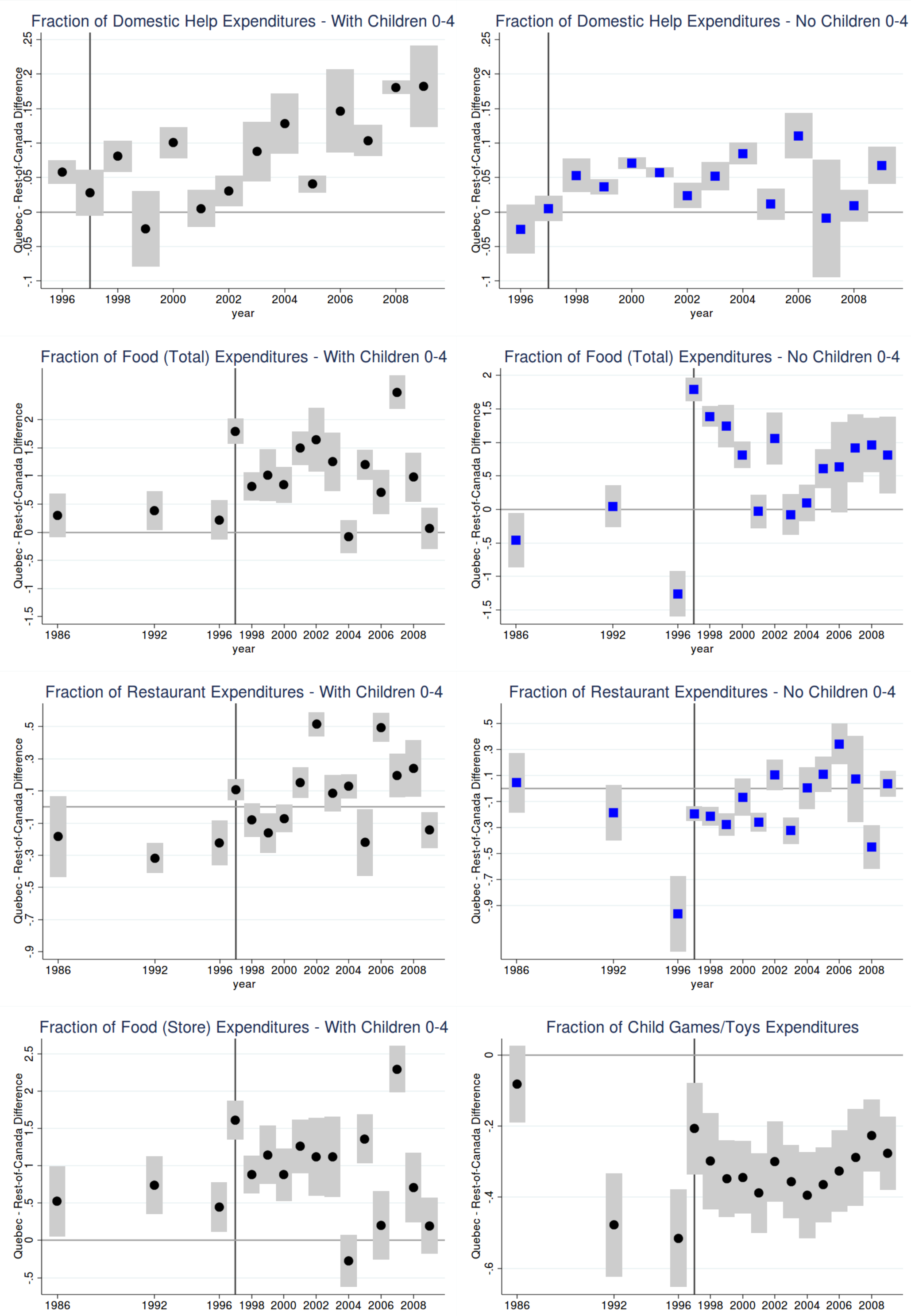
Note: these graphs show the estimated difference between Quebec and the Rest-of-Canada across years for various outcomes, with a 95% confidence band, stemming from a variant of model (2.8). In these models, instead of interacting the eligibility-by-cohort indicator variable with indicator variable for Quebec, year-indicators are interacted with the variable indicating residence in Quebec. Standard errors are clustered at the (province×post)-level. Source of data: NLSCY waves 1-7 (1994-2006) and LFS (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old. The vertical line shows the year of 1997, when the policy was introduced.

Figure A.2: Estimated Difference between Quebec and the Rest-of-Canada across Years, with a 95% Confidence Band, for All and by Education, GSS Time Use Diary and Census



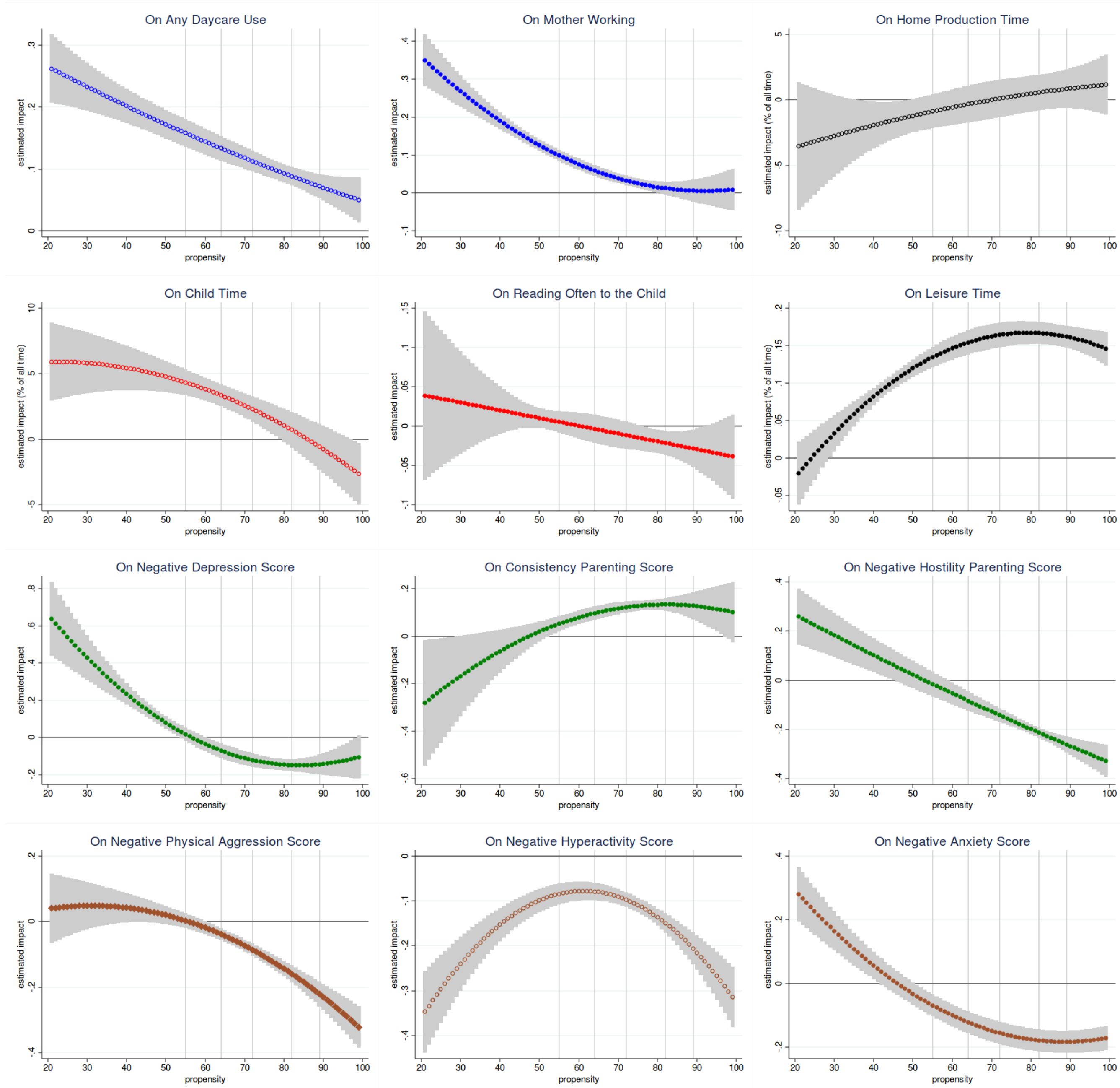
Note: these graphs show the estimated difference between Quebec and the Rest-of-Canada across years for various outcomes, with a 95% confidence band, stemming from a variant of model (2.8). In these models, instead of interacting the eligibility-by-cohort indicator variable with indicator variable for Quebec, year-indicators are interacted with the variable indicating residence in Quebec. Standard errors are clustered at the (province×post)-level. Source of data: Census (1996,2001,2006) and GSS (1998,2005,2010), two-parent families with at least one 0-4 years old child, both parents at most 50 years old. The vertical line shows the year of 1997, when the policy was introduced. In the GSS, 1998 is considered the pre-policy year, due to data limitations.

Figure A.3: Estimated Difference between Quebec and the Rest-of-Canada across Years, with a 95% Confidence Band, for All and by Having Children Aged 0-4, SHS



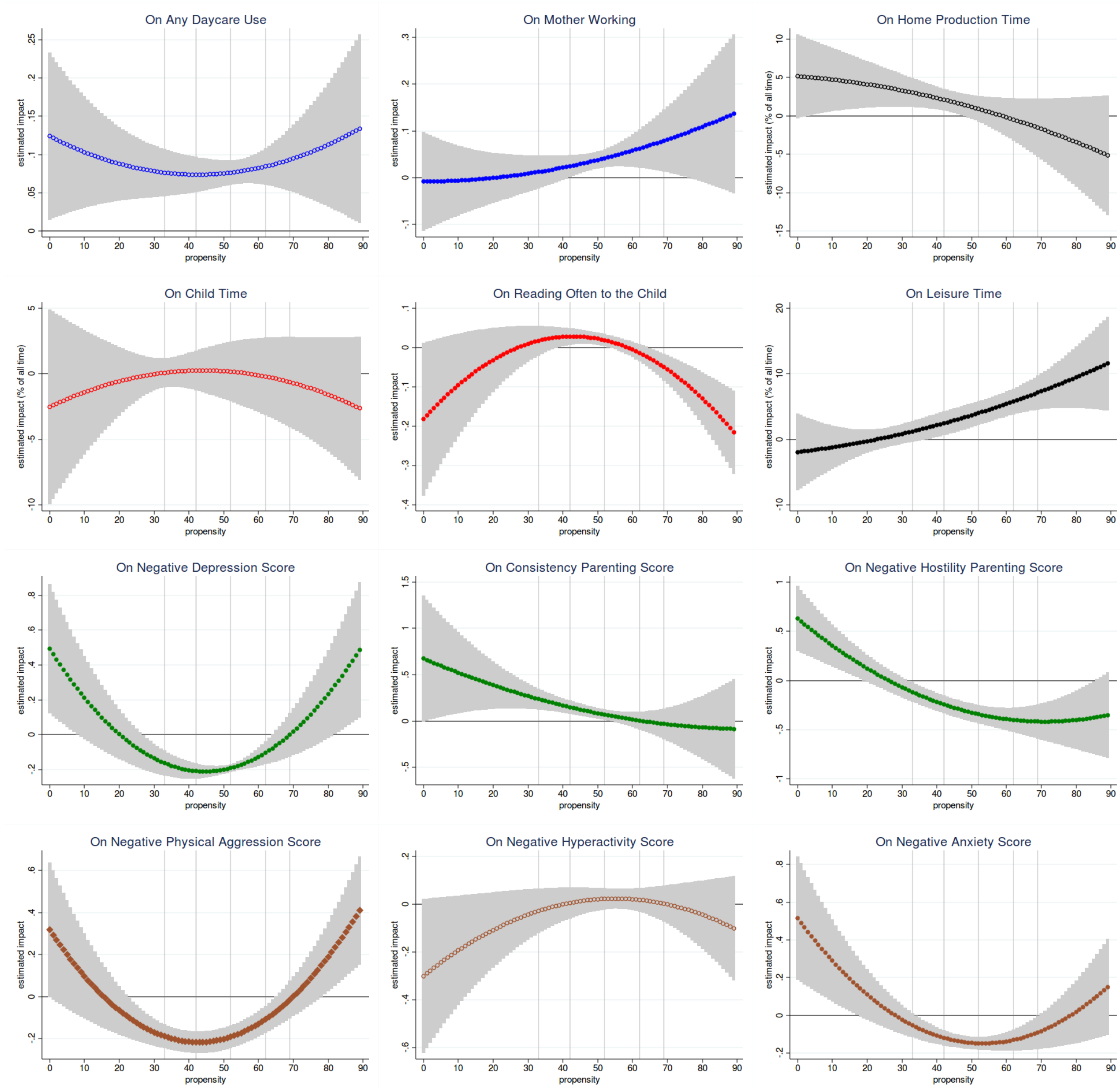
Note*:* these graphs show the estimated difference between Quebec and the Rest-of-Canada across years for various outcomes, with a 95% confidence band, stemming from a variant of model (2.8). In these models, instead of interacting the eligibility-by-cohort indicator variable with indicator variable for Quebec, year-indicators are interacted with the variable indicating residence in Quebec. Standard errors are clustered at the (province×post)-level. Source of data: SHS (1986,1992,1996-2009), two-parent families with and without at least one 0-4 years old child, both parents at most 50 years old. The vertical line shows the year of 1997, when the policy was introduced.

Figure A.4: Estimated Policy Impacts by Propensity with a 95% Confidence Band, for High-Educated Mothers (Quadratic Model)



Note: these figures show the estimated policy impacts for high-educated families by the predicted propensities of the mother working in the absence of the policy, with a 95% confidence band; based on estimates from the quadratic model (2.10), where the underlying coefficients can be seen in Tables A.14-A.18.

Figure A.5: Estimated Policy Impacts by Propensity with a 95% Confidence Band, for Low-Educated Mothers (Quadratic Model)



Note: these figures show the estimated policy impacts for low-educated families by the predicted propensities of the mother working in the absence of the policy, with a 95% confidence band; based on estimates from the quadratic model (2.10), where the underlying coefficients can be seen in Tables A.14-A.18.

A.7 Robustness of Confidence Intervals to Small # of Clusters

The following tables show the robustness of the confidence intervals presented previously ('base') to the Wild-bootstrapping method of Cameron et al. (2008), accounting for small number of clusters.

Table A.20: Alternative Confidence Bounds for Selected Outcome Variables in the NLSCY

1: in institutional care		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.1687	0.1678	0.1669	0.1000	0.1020	0.1028
	upper CI-bound	0.1987	0.1996	0.2004	0.1609	0.1590	0.1560
β_2 : policy×high-educ.	lower CI-bound				0.0398	0.0417	0.0435
	upper CI-bound				0.1001	0.0982	0.0959
1: in any care		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.1110	0.1071	0.1097	0.0486	0.0511	0.0533
	upper CI-bound	0.1500	0.1540	0.1523	0.1276	0.1251	0.1219
β_2 : policy×high-educ.	lower CI-bound				0.0191	0.0218	0.0345
	upper CI-bound				0.1032	0.1006	0.0880
daycare hours		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	5.3781	5.3481	5.3018	2.4682	2.5454	2.7954
	upper CI-bound	6.4225	6.4525	6.4950	4.8980	4.8208	4.5011
β_2 : policy×high-educ.	lower CI-bound				1.5988	1.6939	2.4307
	upper CI-bound				4.5903	4.4953	3.8121
1: mother working		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.0602	0.0552	0.0574	0.0075	0.0097	0.0102
	upper CI-bound	0.0916	0.0967	0.0936	0.0762	0.0740	0.0742
β_2 : policy×high-educ.	lower CI-bound				0.0108	0.0131	0.0180
	upper CI-bound				0.0853	0.0830	0.0764
1: read often		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.0288	0.0292	0.0270	0.0161	0.0177	0.0212
	upper CI-bound	0.0421	0.0417	0.0427	0.0679	0.0662	0.0628
β_2 : policy×high-educ.	lower CI-bound				-0.0530	-0.0503	-0.0403
	upper CI-bound				0.0306	0.0280	0.0188
1: read weekly		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.0158	0.0141	0.0158	-0.0102	-0.0094	-0.0036
	upper CI-bound	0.0229	0.0246	0.0231	0.0160	0.0152	0.0099
β_2 : policy×high-educ.	lower CI-bound				0.0062	0.0072	0.0157
	upper CI-bound				0.0382	0.0372	0.0288

Table A.21: Alternative Confidence Bounds for Selected Outcome Variables in the Census

1: mother working		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.0457	0.0457	0.0454	0.0077	0.0092	0.0127
	upper CI-bound	0.0634	0.0634	0.0638	0.0538	0.0523	0.0484
β_2 : policy×high-educ.	lower CI-bound				-0.0019	-0.0002	0.0103
	upper CI-bound				0.0526	0.0508	0.0405
mother's avg. child time		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.6502	0.6716	0.6230	0.0921	0.1239	0.0786
	upper CI-bound	1.0905	1.0691	1.1378	1.0904	1.0586	1.1017
β_2 : policy×high-educ.	lower CI-bound				0.0978	0.1346	0.3506
	upper CI-bound				1.2556	1.2188	0.9870
mother's avg. home production time		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	-2.3167	-2.3098	-2.3481	-2.7992	-2.7631	-2.5651
	upper CI-bound	-1.7745	-1.7815	-1.7503	-1.6659	-1.7019	-1.9359
β_2 : policy×high-educ.	lower CI-bound				-0.1713	-0.1268	0.3290
	upper CI-bound				1.2269	1.1824	0.7118
father's avg. child time		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.4364	0.4075	0.3876	-0.0191	0.0101	-0.0200
	upper CI-bound	0.9996	1.0285	1.0579	0.9029	0.8736	0.8659
β_2 : policy×high-educ.	lower CI-bound				-0.0199	0.0121	0.2398
	upper CI-bound				0.9889	0.9568	0.7350

Table A.22: Alternative Confidence Bounds for Selected Outcome Variables in the GSS

mother's work time		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	-1.7193	-2.0543	-1.9826	-5.1420	-4.9766	-4.9520
	upper CI-bound	1.8376	2.1727	1.9298	0.1817	0.0163	0.0699
β_2 : policy×high-educ.	lower CI-bound				-0.4793	-0.2407	0.1288
	upper CI-bound				7.2021	6.9635	6.8094
mother's child time		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.6535	0.6765	0.5876	-3.4460	-3.2946	-2.3507
	upper CI-bound	2.3320	2.3090	2.3643	1.4297	1.2783	0.4096
β_2 : policy×high-educ.	lower CI-bound				-0.2383	-0.0192	1.2321
	upper CI-bound				6.8154	6.5963	5.4704
mother's home production time		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	-1.5339	-1.7254	-1.6949	-1.6473	-1.5239	-1.6445
	upper CI-bound	0.2354	0.4269	0.3618	2.3247	2.2014	2.1988
β_2 : policy×high-educ.	lower CI-bound				-3.6127	-3.4686	-2.8004
	upper CI-bound				1.0279	0.8838	0.4120
mother's leisure time		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	-2.2825	-2.5701	-2.3302	0.7494	0.8985	2.0218
	upper CI-bound	0.4772	0.7648	0.5695	5.5498	5.4007	4.3596
β_2 : policy×high-educ.	lower CI-bound				-8.7389	-8.5289	-8.1393
	upper CI-bound				-1.9762	-2.1862	-2.7329

Table A.23: Alternative Confidence Bounds for Selected Outcome Variables in the LFS

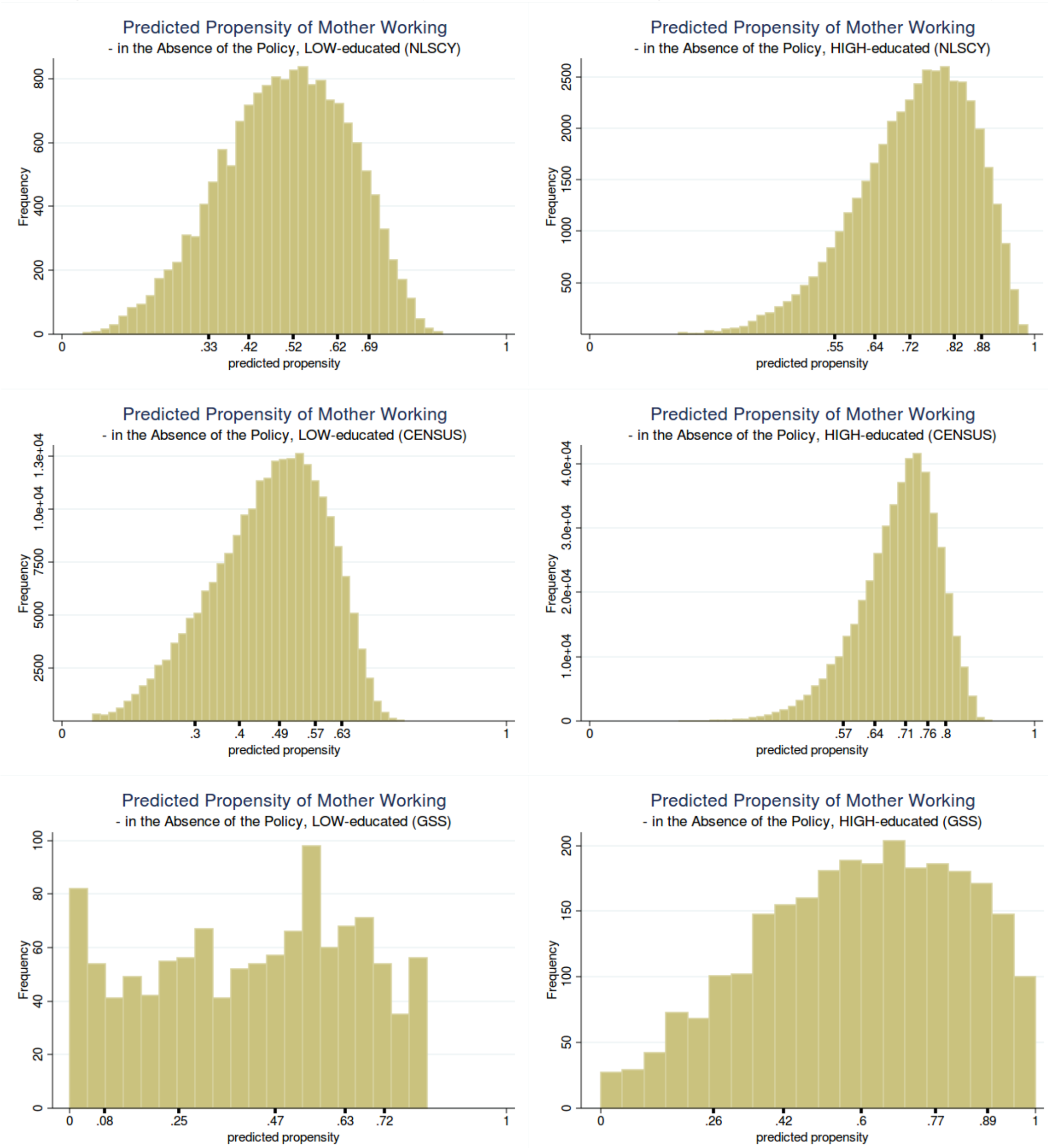
1: mother working		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.0221	0.0215	0.0201	-0.0037	-0.0028	-0.0001
	upper CI-bound	0.0448	0.0454	0.0462	0.0239	0.0230	0.0196
β_2 : policy×high-educ.	lower CI-bound				0.0083	0.0095	0.0156
	upper CI-bound				0.0476	0.0464	0.0397
mother hours		*base*	*CGM*	*wild-boot*	*base*	*CGM*	*wild-boot*
β_1 : policy	lower CI-bound	0.7216	0.7188	0.6948	-0.0482	-0.0188	0.1259
	upper CI-bound	1.4401	1.4429	1.4518	0.8767	0.8473	0.6867
β_2 : policy×high-educ.	lower CI-bound				0.0017	0.0497	0.5146
	upper CI-bound				1.5111	1.4632	1.0029

A.8 Details on Daycare Price Imputation

Neither the GSS, nor the Census dataset contains data on the hourly daycare cost, thus I impute them from the NLSCY according to the following steps:
Step 1. In the NLSCY waves 7 and 8 information on the weekly amount spent at all daycare providers is available (considering only out-of-pocket payments made for regular childcare arrangements, excluding ad-hoc babysitting); also, for Quebec-residents it is known whether the child attends the subsidized program (referred to as $5/$7 per day daycare centre/childcare program, at an Early Childhood Centre (ECC)). I calculate the hourly cost of daycare arrangements as weekly expenses for all daycare arrangements divided by the total number of hours in all daycare arrangements per week, excluding extreme payments (top&bottom 1%).
Step 2. I estimate a model predicting the hourly daycare cost only for daycare users, separately by ECC-attendance, based on variables available both in the GSS/Census and the NLSCY, as parents' education, family structure, parents' age, residence characteristics (province, urban size), and the child's age and gender.
Step 3. For Quebec-residents for years 2005 and 2010 in the GSS, and 2001 and 2006 in the Census, I do an out-of-sample prediction based on the ECC-users and deflate to Canadian dollars in 2002. For Quebec-residents for years 1998 in the GSS and 1996 in the Census, and non-Quebec residents for all years (1998,2005,2010 in the GSS and 1996,2001,2006 in the Census) I do an out-of-sample based on the non-ECC-users and deflate to Canadian dollars in 2002.
A similar procedure in the LFS is used, to impute hourly wages in 1998,2005,2010 in the GSS and 1996,2001,2006 in the Census. Then, I generate imputed hourly net wage as the difference between imputed hourly wage and imputed hourly daycare cost.

A.9 Details on the Propensity Score Estimation

Figure A.6: Predicted Propensity of Mother Working in the Absence of the Policy



Note: these histograms show the predicted propensity of the mother working, predicted by the interaction of parental education of the mother and the father, a full set of year/province/age fixed effects, mother's and father's age and household size (the model was estimated on the pre-policy sample and predicted for the whole sample). Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old. The ticks on the x-axis in each graph correspond to the 10th, 25th, 50th, 75th and 90th percentile of the predicted propensity distribution.

Table A.24: Details on the Propensity Score Estimation, for High- and Low-Educated Families

1: mother working	(1)	(2)	(3)	(4)	(5)	(6)
data & education group:	*NLSCY high*	*NLSCY low*	*GSS high*	*GSS low*	*CENSUS high*	*CENSUS low*
mother: low – father: middle		0.1508***		1.5955***		0.0689***
		(0.049)		(0.418)		(0.010)
mother: low – father: high		0.0609		1.8545***		0.0254
		(0.103)		(0.479)		(0.023)
mother: middle – father: middle	0.0451		-0.8359		0.0011	
	(0.039)		(0.864)		(0.010)	
mother: middle – father: high	-0.1321**		-1.1139		-0.1194***	
	(0.058)		(0.844)		(0.014)	
mother: high – father: low	0.5407***		-1.8016**		0.2718***	
	(0.093)		(0.839)		(0.026)	
mother: high – father: middle	0.3201***		-1.6305*		0.2793***	
	(0.060)		(0.844)		(0.015)	
mother: high – father: high	0.0248		-0.7203		0.1191***	
	(0.052)		(0.877)		(0.013)	
Prince Edward Island	0.2137**	0.4900***	0.2868	0.4478	0.2933***	0.8382***
	(0.090)	(0.138)	(0.589)	(0.887)	(0.060)	(0.084)
Nova Scotia	-0.1125	0.4962***	0.9442*	-0.6742	0.0975***	0.5059***
	(0.075)	(0.131)	(0.563)	(0.819)	(0.034)	(0.049)
New Brunswick	0.0015	0.5599***	-0.0523	0.6854	0.2367***	0.4771***
	(0.073)	(0.116)	(0.532)	(0.775)	(0.036)	(0.049)
Quebec	-0.0902	0.4154***	0.2454	0.2639	0.1857***	0.4784***
	(0.065)	(0.111)	(0.401)	(0.669)	(0.028)	(0.040)
Ontario	0.0347	0.6803***	0.3125	0.2122	0.1667***	0.6212***
	(0.063)	(0.107)	(0.376)	(0.666)	(0.027)	(0.039)
Manitoba	0.1047	0.7335***	-0.0949	-0.1548	0.2603***	0.7151***
	(0.076)	(0.116)	(0.478)	(0.756)	(0.033)	(0.044)
Saskatchewan	0.2096***	0.7111***	0.0758	-0.6118	0.3329***	0.7863***
	(0.072)	(0.117)	(0.480)	(0.813)	(0.033)	(0.045)
Alberta	-0.0267	0.8255***	0.1101	0.2092	0.1348***	0.7439***
	(0.069)	(0.114)	(0.433)	(0.698)	(0.029)	(0.041)
British Columbia	-0.0762	0.5955***	-0.4523	0.6407	0.0867***	0.6303***
	(0.072)	(0.120)	(0.445)	(0.691)	(0.029)	(0.041)
trend	0.1097***	0.0636**				
	(0.018)	(0.029)				
child age = 1	0.2852***	0.2674***	0.5670**	0.8378**	0.2341***	0.1953***
	(0.039)	(0.062)	(0.261)	(0.384)	(0.010)	(0.014)
child age = 2	0.3029***	0.3227***	0.9347***	0.7932*	0.3214***	0.2838***
	(0.046)	(0.073)	(0.292)	(0.412)	(0.012)	(0.016)
child age = 3	0.3788***	0.4230***	0.8062***	0.7247*	0.3500***	0.3364***
	(0.046)	(0.073)	(0.284)	(0.406)	(0.014)	(0.018)
child age = 4	0.3340***	0.4464***	1.0164***	0.7912	0.4238***	0.4299***
	(0.056)	(0.093)	(0.310)	(0.515)	(0.016)	(0.019)
mother's age	0.2286***	0.1921***	-0.2452	0.2953*	0.1971***	0.2309***
	(0.040)	(0.049)	(0.279)	(0.175)	(0.009)	(0.009)
squared of mother's age	-0.0030***	-0.0026***	0.0023	-0.0047*	-0.0026***	-0.0031***
	(0.001)	(0.001)	(0.004)	(0.003)	(0.000)	(0.000)
father's age	0.0544	0.0330	0.4641**	-0.1489***	0.0191**	0.0149*
	(0.037)	(0.049)	(0.222)	(0.051)	(0.008)	(0.008)
squared of father's age	-0.0009*	-0.0007	-0.0054*	0.0023**	-0.0004***	-0.0004***
	(0.001)	(0.001)	(0.003)	(0.001)	(0.000)	(0.000)
constant	-4.8125***	-4.5850***	5.0600	-5.1082*	-3.4245***	-4.7204***
	(0.565)	(0.685)	(4.501)	(2.802)	(0.133)	(0.133)
N	18,889	7,667	313	189	161,860	92,510

Note: this table shows the estimation result of the probit model to predict the propensity of the mother working in the absence of the policy; the model is estimated solely on the pre-policy sample, separately by the mother's education. Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, and GSS (1998,2005,2010) and Census (1996,2001,2006), two-parent families with at least one 0-4 years old child, both parents at most 50 years old.

A.10 Robustness of the Structural Parameter Estimates

Table A.25: Robustness of the Parameter Estimates for Different Values of β_K and ρ_X

implied ρ_K:				
	ρ_X=-0.5	ρ_X=1	ρ_X=-1.5	ρ_X=-2
β_K=0.25	-0.4580	-0.5514	-0.6351	-0.7103
β_K=0.3	-0.3329	-0.3939	-0.4468	-0.4931
β_K=0.35	-0.2814	-0.3306	-0.3728	-0.4091
β_K=0.4	-0.2520	-0.2949	-0.3314	-0.3626
β_K=0.45	-0.2328	-0.2718	-0.3047	-0.3328
implied δ_K:				
	ρ_X=-0.5	ρ_X=1	ρ_X=-1.5	ρ_X=-2
β_K=0.25	0.5243	0.5213	0.5198	0.5188
β_K=0.3	0.3527	0.3697	0.3810	0.3890
β_K=0.35	0.2728	0.2991	0.3163	0.3285
β_K=0.4	0.2213	0.2536	0.2747	0.2896
β_K=0.45	0.1841	0.2207	0.2446	0.2615

Note: this table shows the value of the different implied structural parameters by estimating model (2.11), for different values of β_K and ρ_X, and θ_K taken to be 1 as the mother's child time is used in the estimation. Source of data: Census (1996,2001,2006), two-parent families with at least one 3-4 years old child, both parents at most 50 years old.

Table A.26: Robustness of the Parameter Estimates for Different Values of β_H

	implied ρ_H:	implied δ_H:
β_H=0.05	0.4460	0.6900
β_H=0.10	0.5748	0.4785
β_H=0.15	0.5039	0.4987
β_H=0.20	0.6730	0.3491
β_H=0.25	0.7165	0.2903
β_H=0.30	0.6885	0.3070
β_H=0.35	0.6739	0.3141
β_H=0.40	0.6644	0.3184
β_H=0.45	0.6577	0.3213
β_H=0.50	0.6526	0.3233

Note: this table shows the value of the different implied structural parameters by estimating model (2.12), for different values of β_H and θ_H taken to be 1 as the mother's home production time is used in the estimation. Source of data: Census (1996,2001,2006), two-parent families with at least one 3-4 years old child, both parents at most 50 years old.

A.11 Details on the Policy Simulation

Table A.27: Underlying Means of Daycare and Parental Time for Policy Simulation, No Daycare Policy in Effect

data:	*NLSCY*		*Census*	
ages:	0-2	3-4	0-2	3-4
outcome:	**daycare time**		**parental time**	
high-educ.	**3.544**	**3.856**	**1.511**	**0.741**
	(0.392)	(0.574)	(0.244)	(0.199)
constant	**9.448**	**11.818**	**40.515**	**47.322**
	(1.472)	(2.227)	(2.209)	(3.271)
N	34,946	21,431	412,595	185,970
R^2	0.068	0.0813	0.0356	0.0379

Note: this table shows the conditional mean for daycare time (measured in hours per week) and parental time (measured in hours per week of the average of the mother's and the father's time spent with the child), conditional on full set of province/waves/mother's age-dummies, age of the father and number/age structure/gender of the children in the household. Standard errors are in parentheses and are clustered at the (province×post)-level. Source of data: NLSCY waves 1-7 (1994-2006), 0-4 years old children in two-parent families, both parents at most 50 years old; and Census (1996,2001,2006), two-parent families with at least one 0-4 years old child, both parents at most 50 years old. Estimation restricted for observations with no daycare policy in effect.

Table A.28: Policy Impacts on Daycare and Parental Time for Policy Simulation, Ages 3-4

data:	*NLSCY*	*Census*
ages:	3-4	3-4
outcome:	**daycare time**	**parental time**
β_1 : policy	**2.595**	**0.632**
	(0.756)	(0.222)
β_2 : policy	**5.840**	**0.228**
×high-educ.	(1.008)	(0.184)
N	21,431	185,970
R^2	0.1206	0.0393

Note: this table shows the result of estimating the Difference-in-Differences model (2.9). Standard errors are in parentheses and are clustered at the (province×post)-level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: NLSCY waves 1-7 (1994-2006), 3-4 years old children in two-parent families, both parents at most 50 years old; and Census (1996,2001,2006), two-parent families with at least one 3-4 years old child, both parents at most 50 years old.

Table A.29: Summary Table of No-Policy, Actual and Counterfactual Levels of Child Human Capital for Higher- and Lower-Educated Parents' Children in Policy Simulation

ages:	0-2	3-4	3-4	3-4	3-4
regime:	prior	no policy	policy	counterfactual	
				H no time efficiency advantage	L no time efficiency disadvantage
higher-educated (H)	K^H_{012}=4.2344	$\tilde{K}^H_{34}$=5.0101	K^H_{34}=5.8811	$\check{K}^{H1}_{34}$=5.3196	K^H_{34}=5.8811
lower-educated (L)	K^L_{012}=3.1153	$\tilde{K}^L_{34}$=3.8016	K^{34}_L=4.3264	K^{34}_L=4.3264	$\check{K}^{L2}_{34}$=4.7453
gap ($\triangle K$)	gap_{012}=1.1190	$\tilde{gap}_{34}$=1.2085	gap_{34} =1.5547	$\check{gap}^{H1}_{34}$=0.9932	$\check{gap}^{L2}_{34}$=1.1358

Note: this table shows the level of child human capital K under different scenarios, by feeding in average daycare time and parental time in bold as inputs from Tables A.27 and A.28 into the CES production function (2.2), using the estimated parameters $\hat{\rho}_K = -0.4$ and $\hat{\delta}_K = 0.36$. For instance,

$$K^L_{012}=3.1153=\left[\left(1^{0.36}\times 40.515\right)^{(-0.4)}+9.448^{(-0.4)}\right]^{\left(\frac{1}{-0.4}\right)};$$

$$\tilde{K}^H_{34}=5.0101=\left[\left(2^{0.36}\times(47.322+0.741)\right)^{(-0.4)}+(11.818+3.856)^{(-0.4)}\right]^{\left(\frac{1}{-0.4}\right)};$$

$$K^{34}_L=4.3264=\left[\left(1^{0.36}\times(47.322+0.632)\right)^{(-0.4)}+(11.818+2.595)^{(-0.4)}\right]^{\left(\frac{1}{-0.4}\right)};$$

$$K^H_{34}=5.8811=\left[\left(1^{0.36}\times(47.322+0.632+0.228)\right)^{(-0.4)}+(11.818+2.595+5.840)^{(-0.4)}\right]^{\left(\frac{1}{-0.4}\right)}.$$

The last but one column shows the counterfactual level of K if high-educated parents had no time efficiency advantage; i.e. if $\delta = 0$ or equivalently in this model, $S = 1$. The last column shows the counterfactual level of K if low-educated parents had no time efficiency disadvantage, in the counterfactual case of $S = 2$.

Appendix B

Appendix to Chapter 2

B.1 Additional Institutional Details

According to Vago (2005, pp.744), the flexible system of primary school start in Hungary, allowing for academic redshirting under certain circumstances, came into effect in 1986. The intention was to provide opportunity for the less developed and immature children to spend one additional year in child care, while for the more developed and mature children to start school earlier than prescribed. However, there were strong social and institutional reasons, other than child abilities around school entry, for redshirting. In the middle of the 1980s Hungary started to become open and for many people studying and working abroad became possible; the increasing entrepreneurial opportunities and the very high unemployment rates among the unskilled and young around the transition from socialism incentivized primarily the middle-class parents to provide better quality, rather than fast education for their children. Moreover, good primary schools required students to pass increasingly difficult entrance exams at that time that were (more) easily taken by (more) mature children.

The physical, psychological and social requirements for school-readiness are described in the enactment of the Ministry of Education, 137/1996. (VIII. 28.), VI./2. Children meeting the physical requirements can be characterized by well-proportioned body, more and more harmonious movements, proper audition, developing coordinational skills and the ability of controlling motion, behavior and physical needs. Children meeting the psychological requirements are open to enter primary school with gradually enhancing perceptional and remembering skills, being able to pay attention to others and knowing the primary behavioral requirements, knowing the own body and primary information about herself and her environment. Children meeting the social requirements are ready to accept the school life and primary school teacher, gradually become able to cooperate, to initiate contact and to conform to the rules. The aforementioned ones are generally needed for heightened ability to learn in school. For instance, proper audition and vision is essential to learn how to read and write properly. Knowing the own body is essential, since individuals generally determine different object of the world relative to themselves. There is also some “knowledge stock” children need to have at the age of primary school start, or it considerably helps them later. The developmental experts highlighted the following ones in general: own name, age, parents’ name and occupation, siblings’ name and age, different parts of the body and their role, assessing distance (“near” and “close”), assessing the order of events, different colors, different parts of the day, different seasons, the relationship between weather and clothing, organs and senses, primitive counting abilities (“more” and “less”, “shorter” and longer”, “whole” and “part”), etc.

B.2 Appendix Tables

Table B.1: Details of Sample Selection, Administrative Data

	2008	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**
All [N]	107,654	100,620	96,898	94,047	93,167	93,907	92,544
SAMPLE1	80.83%	80.48%	83.12%	82.94%	84.54%	84.38%	84.15%
SAMPLE2	79.82%	80.43%	82.96%	82.80%	82.72%	84.04%	84.06%
SAMPLE3	79.69%	80.38%	82.73%	82.62%	82.11%	83.77%	83.71%
FINAL SAMPLE	76.54%	77.84%	80.03%	80.97%	81.94%	81.11%	81.29%

Note: source of data: HNABC (grade 6) 2008-2014. Each cell shows % of sample remaining after keeping children who have information on years spent in child care (Sample1),on exact birth date (Sample2), and who have at least one valid testscore (mathematics or reading, Sample3). The final sample shows sample after keeping children who have valid information about parental background (in specific, parental education).

Table B.2: Distribution, Fraction of Children Entering School at age 7 and Average Years Spent in child care, by Month of Birth and child care Entry Age

month of birth	**child care entry age:**	**2**	**3**	**4**	**5**	**average**
September	**frequency**	**27.86%**	**5.08%**	**3.84%**	**4.28%**	**8.744%**
	fraction of children entering school at age 7	*1.60%*	*8.10%*	*22.07%*	*34.45%*	*0.062*
	average years spent in child care	3.995	3.078	2.238	1.520	3.477
October	**frequency**	**23.21%**	**5.69%**	**3.92%**	**4.18%**	**8.325%**
	fraction of children entering school at age 7	*2.09%*	*9.84%*	*27.30%*	*39.76%*	*0.082*
	average years spent in child care	4.003	3.096	2.292	1.585	3.426
November	**frequency**	**19.35%**	**6.16%**	**4.17%**	**4.62%**	**8.013%**
	fraction of children entering school at age 7	*2.78%*	*12.28%*	*31.43%*	*44.11%*	*0.109*
	average years spent in child care	4.009	3.121	2.327	1.606	3.373
December	**frequency**	**14.02%**	**7.68%**	**4.80%**	**5.52%**	**8.180%**
	fraction of children entering school at age 7	*3.92%*	*17.06%*	*40.48%*	*48.05%*	*0.165*
	average years spent in child care	4.011	3.169	2.426	1.663	3.288
January	**frequency**	**4.64%**	**10.07%**	**6.25%**	**7.77%**	**8.386%**
	fraction of children entering school at age 7	*10.50%*	*30.22%*	*61.80%*	*63.98%*	*0.337*
	average years spent in child care	4.052	3.302	2.650	1.840	3.217
February	**frequency**	**2.99%**	**9.41%**	**6.16%**	**7.17%**	**7.652%**
	fraction of children entering school at age 7	*17.38%*	*36.26%*	*66.17%*	*67.17%*	*0.403*
	average years spent in child care	4.099	3.363	2.697	1.880	3.246
March	**frequency**	**2.44%**	**10.30%**	**7.23%**	**8.65%**	**8.357%**
	fraction of children entering school at age 7	*22.56%*	*43.27%*	*72.28%*	*72.66%*	*0.478*
	average years spent in child care	4.151	3.434	2.756	1.936	3.286
April	**frequency**	**1.68%**	**9.72%**	**7.54%**	**8.69%**	**7.924%**
	fraction of children entering school at age 7	*28.86%*	*50.82%*	*76.86%*	*75.61%*	*0.553*
	average years spent in child care	4.159	3.509	2.800	1.986	3.328
May	**frequency**	**1.38%**	**9.84%**	**8.71%**	**9.26%**	**8.170%**
	fraction of children entering school at age 7	*42.49%*	*63.34%*	*83.83%*	*79.58%*	*0.671*
	average years spent in child care	4.283	3.635	2.866	1.984	3.414
June	**frequency**	**0.83%**	**8.68%**	**13.71%**	**12.76%**	**8.378%**
	fraction of children entering school at age 7	*74.49%*	*92.83%*	*98.05%*	*96.77%*	*0.942*
	average years spent in child care	4.541	3.930	3.013	2.168	3.544
July	**frequency**	**0.84%**	**8.97%**	**16.43%**	**13.30%**	**9.032%**
	fraction of children entering school at age 7	*81.46%*	*96.15%*	*98.82%*	*98.63%*	*0.969*
	average years spent in child care	4.568	3.964	3.020	2.197	3.547
August	**frequency**	**0.77%**	**8.41%**	**17.23%**	**13.81%**	**8.837%**
	fraction of children entering school at age 7	*84.33%*	*97.62%*	*99.24%*	*98.51%*	*0.980*
	average years spent in child care	4.605	3.980	3.027	2.226	3.532

Note: source of data: HNABC (grade 6) 2008-2014. The bold percentages in each column add up to 100%.

Table B.3: The Effect of Quarter of Birth - Detailed First-stage Results on School Entry Delay and its Interactions by Gender and Parental Education, Administrative Data, Grades 6/8/10

Grade 6	***Academic Redshirting,*** x_d : *January 1st*				***Involuntary Delay,*** x_d : *June 1st*			
	D_i	$boy_i \times D_i$	$lowp_i \times D_i$	$lowp_i \times boy_i \times D_i$	D_i	$boy_i \times D_i$	$lowp_i \times D_i$	$lowp_i \times boy_i \times D_i$
Z_i	0.0884***	-0.0882***	-0.0582***	-0.0322***	0.2552***	-0.0991***	-0.0680***	-0.0333***
	[0.005]	[0.002]	[0.002]	[0.002]	[0.006]	[0.003]	[0.002]	[0.002]
$boy_i \times Z_i$	0.0987***	0.3417***	0.0002	0.00	-0.1326***	0.3358***	0.0006	0.0003
	[0.004]	[0.004]	[0.000]	[0.000]	[0.004]	[0.004]	[0.000]	[0.000]
$lowp_i \times Z_i$	0.0122**	0.0002	0.2540***	-0.0005	-0.0303***	-0.0017***	0.4400***	-0.0011***
	[0.005]	[0.001]	[0.005]	[0.000]	[0.005]	[0.001]	[0.005]	[0.000]
$lowp_i \times boy_i \times Z_i$	-0.0540***	-0.0423***	0.0444***	0.2993***	0.0239***	-0.0021	-0.1091***	0.3335***
	[0.007]	[0.006]	[0.006]	[0.005]	[0.006]	[0.005]	[0.005]	[0.005]
N	223,924	223,924	223,924	223,924	227,185	227,185	227,185	227,185
R^2	0.179	0.314	0.322	0.389	0.281	0.747	0.761	0.82
joint F − statistic	*394.86*	*2442.23*	*1521.7*	*1066.57*	*688.84*	*2312.57*	*2284.95*	*1284.36*
Grade 8	***Academic Redshirting,*** x_d : *January 1st*				***Involuntary Delay,*** x_d : *June 1st*			
	D_i	$boy_i \times D_i$	$lowp_i \times D_i$	$lowp_i \times boy_i \times D_i$	D_i	$boy_i \times D_i$	$lowp_i \times D_i$	$lowp_i \times boy_i \times D_i$
Z_i	0.0703***	-0.0708***	-0.0502***	-0.0256***	0.2897***	-0.1028***	-0.0676***	-0.0329***
	[0.004]	[0.002]	[0.002]	[0.001]	[0.006]	[0.003]	[0.003]	[0.002]
$boy_i \times Z_i$	0.0850***	0.2839***	-0.0015***	-0.0007***	-0.1326***	0.3740***	0.0005	0.0002
	[0.004]	[0.004]	[0.001]	[0.000]	[0.004]	[0.004]	[0.000]	[0.000]
$lowp_i \times Z_i$	0.0230***	0.0012	0.2181***	0.0002	-0.0351***	-0.0012**	0.4732***	-0.0007***
	[0.005]	[0.001]	[0.004]	[0.000]	[0.005]	[0.001]	[0.005]	[0.000]
$lowp_i \times boy_i \times Z_i$	-0.0335***	-0.0149**	0.0527***	0.2696***	0.0370***	0.0052	-0.0963***	0.3789***
	[0.007]	[0.006]	[0.006]	[0.005]	[0.006]	[0.005]	[0.005]	[0.005]
N	206,499	206,499	206,499	206,499	219,173	219,173	219,173	219,173
R^2	0.183	0.271	0.272	0.324	0.307	0.735	0.751	0.808
joint F − statistic	*352.05*	*1724.46*	*1233.23*	*811.97*	*816.98*	*2771.56*	*2868.59*	*1545.64*
Grade 10	***Academic Redshirting,*** x_d : *January 1st*				***Involuntary Delay,*** x_d : *June 1st*			
	D_i	$boy_i \times D_i$	$lowp_i \times D_i$	$lowp_i \times boy_i \times D_i$	D_i	$boy_i \times D_i$	$lowp_i \times D_i$	$lowp_i \times boy_i \times D_i$
Z_i	0.0730***	-0.0644***	-0.0376***	-0.0196***	0.3086***	-0.1098***	-0.0651***	-0.0301***
	[0.004]	[0.002]	[0.002]	[0.001]	[0.006]	[0.003]	[0.003]	[0.002]
$boy_i \times Z_i$	0.0860***	0.2701***	-0.0003	-0.0002	-0.1187***	0.4205***	0.0003	0.0002
	[0.004]	[0.004]	[0.000]	[0.000]	[0.004]	[0.004]	[0.000]	[0.000]
$lowp_i \times Z_i$	0.0122**	0.0001	0.1964***	0.0003	-0.0088*	-0.0009	0.5316***	-0.0007***
	[0.005]	[0.001]	[0.004]	[0.000]	[0.005]	[0.001]	[0.005]	[0.000]
$lowp_i \times boy_i \times Z_i$	-0.0231***	-0.0116**	0.0623***	0.2582***	0.0286***	0.0223***	-0.0901***	0.4428***
	[0.007]	[0.006]	[0.006]	[0.005]	[0.007]	[0.006]	[0.006]	[0.005]
N	219,901	219,901	219,901	219,901	214,092	214,092	214,092	214,092
R^2	0.13	0.241	0.239	0.299	0.337	0.716	0.741	0.791
joint F − statistic	*361.13*	*1538.4*	*1069.96*	*635.91*	*970.04*	*3442.59*	*4409.38*	*1939.11*

Note: the above tables show the result of separately estimating a LPM of D_i, $boy_i \times D_i$, $lowp_i \times D_i$, $boy_i \times lowp_i \times D_i$ on Z_i, $boy_i \times Z_i$, $lowp_i \times Z_i$, $boy_i \times lowp_i \times Z_i$ and control variables, separately for grade 6, 8 and 10. D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). Z_i is a binary variable denoting quarter of birth of child i (1: child was born on/after January 1st or June 1st, 0: child was born before January 1st or June 1st). Control variables include: linear trend in month of birth and its interaction with a binary variable denoting quarter of birth, cohort fixed effects, family background and child care variables listed in Part 4, and missing response characteristics controls. Estimation restricted to children who were born in the three-months window around the cutoff dates of January 1st and June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HNABC (grades 6,8,10) 2008-2014, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding equation is (3.1).

Table B.4: The Effect of Quarter of Birth - Detailed First-stage Results on School Entry Delay and its Interactions by Gender and Parental Education, Survey Data

	Academic Redshirting, x_d : *January 1st*				***Involuntary Delay,*** x_d : *June 1st*			
	D_i	$boy_i \times D_i$	$lowp_i \times D_i$	$lowp_i \times$ $boy_i \times D_i$	D_i	$boy_i \times D_i$	$lowp_i \times D_i$	$lowp_i \times$ $boy_i \times D_i$
Z_i	0.0990***	-0.0940***	-0.0561***	-0.0269***	0.3583***	-0.1177***	-0.0655***	-0.0301***
	[0.023]	[0.013]	[0.011]	[0.009]	[0.030]	[0.015]	[0.013]	[0.009]
$boy_i \times Z_i$	0.0868***	0.3619***	0.0008	0.0008	-0.0555**	0.5338***	0.0033	0.0017
	[0.023]	[0.016]	[0.003]	[0.002]	[0.028]	[0.019]	[0.003]	[0.002]
$lowp_i \times Z_i$	0.0338	0.0034	0.3070***	0.0001	-0.0647**	0.0026	0.5292***	0.0021
	[0.029]	[0.004]	[0.023]	[0.002]	[0.033]	[0.005]	[0.026]	[0.003]
$lowp_i \times boy_i \times Z_i$	-0.0171	0.0091	0.0649**	0.3696***	-0.0152	-0.0769**	-0.0738**	0.4547***
	[0.040]	[0.029]	[0.033]	[0.024]	[0.044]	[0.030]	[0.035]	[0.023]
N	6,899	6,899	6,899	6,899	6,116	6,116	6,116	6,116
R^2	0.213	0.323	0.326	0.393	0.331	0.611	0.625	0.697
joint $F-statistic$	*20.85*	*188.13*	*117.12*	*66.13*	*42.85*	*323.02*	*205.62*	*100.37*

Note: this table shows the result of separately estimating a LPM of D_i, $boy_i \times D_i$, $lowp_i \times D_i$, $boy_i \times lowp_i \times D_i$ on Z_i, $boy_i \times Z_i$, $lowp_i \times Z_i$, $boy_i \times lowp_i \times Z_i$ and control variables. D_i is a binary variable denoting age at primary school entry of child i (1: child entered primary school at the age of 7, 0: child entered primary school at the age of 6). Z_i is a binary variable denoting quarter of birth of child i (1: child was born on/after January 1st or June 1st, 0: child was born before January 1st or June 1st). Estimation restricted to children who were born in the three-months window around the cutoff dates of January 1st and June 1st. Standard errors clustered at the school level. *** p<0.01, ** p<0.05, * p<0.1. Source of data: HCLS, KIR-Stat/DEM 1998-2006, LGT 1998-2005. The corresponding equation is (3.1).

Table B.5: Descriptive Statistics of Class Size, Relative Rank and Fraction of Summer-Born Children in Class in All Schools and Non-sorting Schools, Children born in a 3-month window around June 1st, Administrative Data Grade 6

descriptive	*All Schools*				*Non-sorting Schools (Based on Score)*			
statistics	*# of peers*	RR_i	Q_i	$peers\_summer_i$	*# of peers*	RR_i	Q_i	$peers\_summer_i$
Mean	20.504	54.754	0.363	28.659	20.699	55.498	0.370	28.584
Std. Dev.	4.923	27.994	0.481	10.514	4.794	27.885	0.483	10.402
Min	9	2.326	0	0	9	2.326	0	0
Max	39	100	1	83.333	39	100	1	76.923

Note: source of data: HNABC (grade 6) 2008-2014.

Table B.6: Descriptive Statistics of the Fraction of Within-Class Sum-of-Squares, All Schools and Non-sorting Schools, Children born in a 3-month window around June 1st, Administrative Data Grade 6

statistics	*all*	*non-sorting A*	*non-sorting B*
5%	0.523	0.711	0.514
10%	0.620	0.801	0.606
25%	0.784	0.899	0.766
50%	0.925	0.965	0.903
75%	0.994	0.995	0.977
90%	1	1	0.999
95%	1	1	1
Mean	0.863	0.925	0.848
Std. Dev.	0.162	0.106	0.162

Note: source of data: HNABC (grade 6) 2008-2014. Non-sorting schools “A” are the ones that do not sort children systematically into classes based on prior student achievement, while non-sorting schools “B” are the ones that do not sort children systematically into classes based on month of birth.

Appendix C

Appendix to Chapter 3

C.1 Conceptual Framework: Contracting Under Complexity

We present a conceptual framework to illuminate when payment rates will and won't be benchmarked perfectly to Medicare. Physicians and insurers can use Medicare's payments as a default relative price schedule, so that reimbursements are simply a markup over Medicare's rates. Adopting this default has costs if Medicare's relative payments are suboptimal, in a sense developed below. It may nonetheless be efficient to rely on this default due to negotiation and coordination costs.[137]

Consider an insurer that purchases N types of medical services indexed by $j \in \{1, \ldots, N\}$, for treating its enrollees. These types could represent individual billing codes—very specific services such as a 20-minute office visit—or broader categories of care, such as all diagnostic imaging. We abstract from the physician-insurer bargaining process and assume that the insurer sets prices with full knowledge of the aggregate supply curve for each type of care. Let r_j denote the reimbursement rate that the insurer pays to physicians for providing service j, and let r_j^M be the corresponding Medicare rate. In aggregate, the physician market supplies care to the insurer's patients according to the supply functions $s_j(r_j)$. If the true price-setting process is not so simple—say, if physicians are not price-takers—the model's main ideas still hold. In that case, physicians and insurers strive to reach a pricing agreement that maximizes joint surplus, which they then split. We would simply view prices as jointly determined and adjustment costs defined below as those incurred by both parties.

We assume that the insurer aims to minimize its medical expenses while keeping patients, or their employers, satisfied with the insurance product. This constraint requires that the insurer purchase enough care to achieve the patients' reservation value $\bar{u}$.[138] We assume the patients have additively separable preferences over types of care, captured by $U(q_1, \ldots, q_N) = \sum_j u_j(q_j)$ where q_j is the quantity of service j supplied to a representative patient.

We will first consider the insurer's optimally chosen reimbursement rates and then consider deviations from that optimum. The insurer incurs costs of $C(r_1, \ldots, r_N) = \sum_j r_j q_j$ for treating patients. The physician supply curves determine quantities as $q_j = s_j(r_j)$. The insurer aims to minimize costs while keeping the patients satisfied, or:

$$\min_{r_1,\ldots,r_N} \sum_j s_j(r_j) r_j \text{ subject to } \sum_j u_j(s_j(r_j)) \geq \bar{u}. \tag{C.1}$$

Let $\theta_j = u_j'(q_j)$ denote the marginal value of care type j, and let $\epsilon_j = \dfrac{s_j'(r_j) r_j}{s_j(r_j)}$ denote the supply elasticity for that service. The first-order conditions for problem (C.1) then satisfy:

$$\frac{r_j^*}{r_k^*} = \frac{\theta_j / \left({\epsilon_j}^{-1} + 1\right)}{\theta_k / \left({\epsilon_k}^{-1} + 1\right)} \text{ for all } j, k. \tag{C.2}$$

Equation (C.2) shows that the insurer would optimally increase the relative payment for a service that is more highly valued (higher θ) or whose supply is more elastic (higher ϵ).

We next consider how much the insurer's costs increase if its reimbursement rates follow a different ratio. Suppose that the insurer is forced to adopt a scaled version of Medicare's pricing scheme, where reimbursements

[137]Providers themselves may find deviating from Medicare's menu costly due to increases in the non-trivial administrative expenses associated with billing (Cutler and Ly, 2011). A private alternative to Medicare's menu would not reduce these costs. It would also have trouble reflecting differences in market conditions which, as the model shows, influence optimal reimbursement rates.

[138]Geruso et al. (2016) take a similar conceptual approach to a health insurer's objective function.

are set exogenously at r_j^M. The insurer scales all of the Medicare rates by φ so that the patient satisfaction constraint $U(q_1,\ldots,q_N)=\overline{u}$ continues to bind with equality. Let $d_j=\varphi r_j^M - r_j^*$ represent the deviation from optimal reimbursements implied by this scheme, and $\delta_j=\frac{d_j}{r_j^*}$ the proportional deviation. Supposing that these deviations are reasonably small, we can approximate the insurer's costs to second order as

$$\begin{aligned}
C(\varphi r_1^M,\ldots,\varphi r_N^M) &= \sum_j s_j(r_j^*+d_j)(r_j^*+d_j) \\
&\approx C(r_1^*,\ldots,r_N^*) + \sum_j \left[s_j'(r_j^*)+\frac{1}{2}s_j''(r_j^*)\right] d_j^2 \qquad \text{(C.3)} \\
&= C(r_1^*,\ldots,r_N^*) + \sum_j [s_j(r_j^*)\epsilon_j\left(\rho_j+\epsilon_j+2r_j^*-1\right) d_j^2 \qquad \text{(C.4)}
\end{aligned}$$

where $\rho_j = \frac{d\epsilon_j}{dr_j^*}\frac{r_j^*}{\epsilon_j}$ denotes the super-elasticity of supply. Note that, since $C(r_1^*,\ldots,r_N^*)$ defines the cost-minimizing reimbursements, the summation terms in equations (C.3) and (C.4) are positive.

By the envelope theorem, small deviations from optimal pricing that continue to satisfy the patient satisfaction constraint don't generate any first-order cost increase. However larger deviations from the insurer's preferred reimbursements r^* generate convex costs, which are increasing in the supply elasticity ϵ_j and super-elasticity ρ_j. This occurs because larger deviations from the insurer's unconditional optimum force the insurer to spend more to achieve the same patient satisfaction. This is especially so with more elastic supply, which causes any given deviation to shift physician behavior farther from the efficient service mix.

Suppose that Medicare's pricing is used as a benchmark and the insurer can adopt the scaled version $(r_j=\varphi r_j^M)$ for free. Alternatively, the insurer can make costly adjustments to this default. Assume that it can pay a cost of $\alpha\theta$ to shrink the magnitude of the deviations relative to optimal prices from δ_j to $\frac{\delta_j}{\theta+1}$. How much of a reduction will it pay for?[139]

Including the adjustment costs, the insurer's total spending in excess of the first-best is now:

$$\triangle\, C(\theta) = \frac{1}{2}\sum_j s_j(r_j^*)\epsilon_j\left(\rho_j+\epsilon_j+2r_j^*-1\right)\left(\frac{\delta_j}{\theta+1}\right)^2 + \alpha\theta. \qquad \text{(C.5)}$$

The insurer will choose θ to minimize $\triangle C(\theta)$. The minimum is achieved at:

$$\theta^* = \max\left\{0, \left[\frac{1}{\alpha}\sum_j s_j\left(r_j^*\right)\epsilon_j\left(\rho_j+\epsilon_j+2r_j^*-1\right)\delta_j^2\right]^{1/3} - 1\right\}. \qquad \text{(C.6)}$$

Larger pricing errors δ_j, higher quantities $s_j(r_j^*)$, more elastic supply, and lower adjustment costs α will all lead the insurer to spend more on mitigating the deviations.

C.2 Additional Detail on Implied Conversion Factors

C.2.1 Data Cleaning

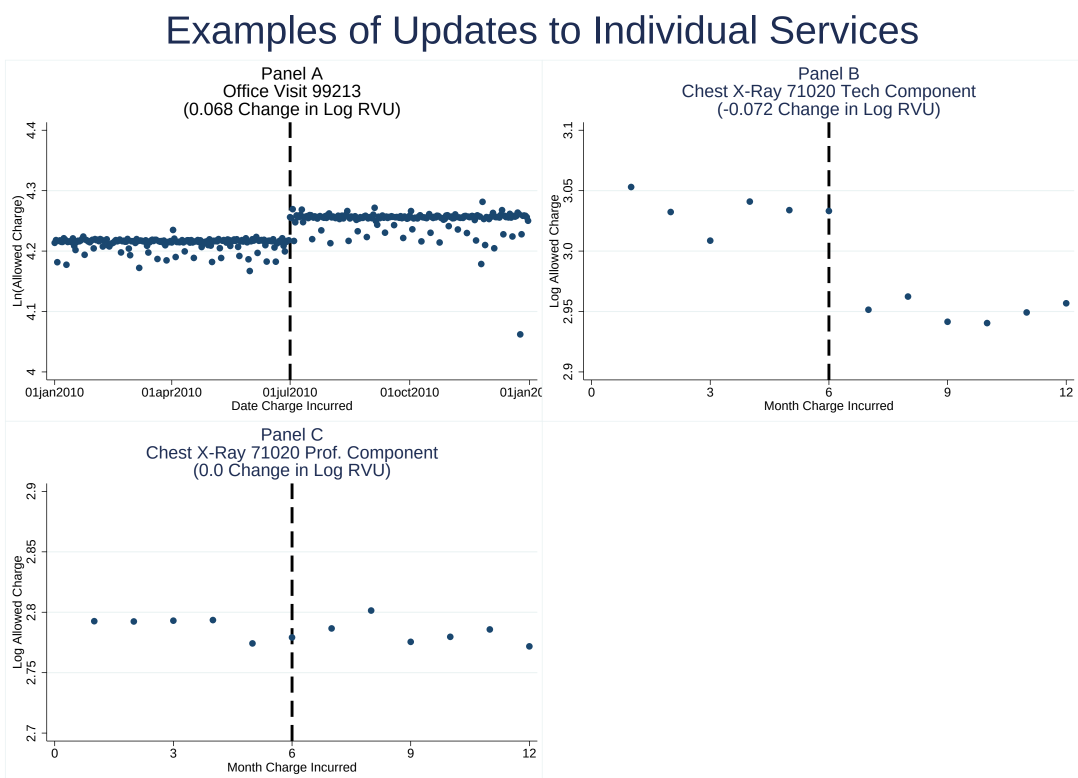
This section describes our process for cleaning and merging the BCBS claims data. Table C.1 shows the data lost as we progress from the raw claims data to the final analysis sample.

[139] We can alternatively think of θ as determining a probability that the reimbursement for each service is switched from exactly the benchmark to exactly the insurer's preferred value. By paying for a higher θ, the insurer can obtain the opportunity to adjust payment for additional randomly chosen services. Following the formulation in the text, a payment of $\alpha\theta$ allows the insurer to adjust a share $1-1/\left(\theta+1\right)^2$ of services, while the remaining share $1/\left(\theta+1\right)^2$ retain the Medicare default φr_j^M.

For concreteness, consider the 2009 claims data. The data for this year start with 54,724,994 claim lines and $4.01 billion in spending (row A). To reduce heterogeneity along several administrative margins, we analyze claim lines for which the payment is non-missing, the service quantity is 1, and the observation is an "original" claim line rather than an adjustment to a past payment.[140] This eliminates 5,090,024 claim lines and leaves us with $3.24 billion in spending (row B). Next, we want to ensure that our analysis focuses on reimbursements for services that are administratively equivalent from a payments perspective, and whose payments have been agreed upon through *ex ante* negotiations. We thus retain only observations that are explicitly coded as being "outpatient" and "in network." These criteria eliminate a total of 8,302,709 claim lines and leave us with $2.45 billion in spending (row C). Next we drop relatively rare service codes for which we have fewer than 10 observations prior to the RVU updates in a given year. In the 2009 data, this eliminates 149,269 claims and leaves us with $2.44 billion in spending (row D). The resulting sample of 41,182,992 service lines and $2.44 billion in spending constitutes the administratively comparable and sufficiently common billing codes we aim to understand.

In order for private insurers to benchmark prices to Medicare, at a minimum they would need to use Medicare's billing codes. On row (E), we thus merge the remaining claims with Medicare billing codes, which provides an upper bound on the potential benchmarking. The final analysis sample in 2009 includes 3,807 unique HCPCS codes, which comprise 21,941,227 service lines and $1.89 billion of spending. The key conclusion from row (E) is that, once we restrict ourselves to the relevant universe of data, additional losses from merging in Medicare codes and eliminating infrequent codes are not substantial. More specifically, this merge only loses notable portions of one broad spending category, namely laboratory tests, for which both Medicare and BCBS frequently base payments on non-standard codes. We retain over 97 percent of claims for evaluation and management, diagnostic imaging, and surgical services.

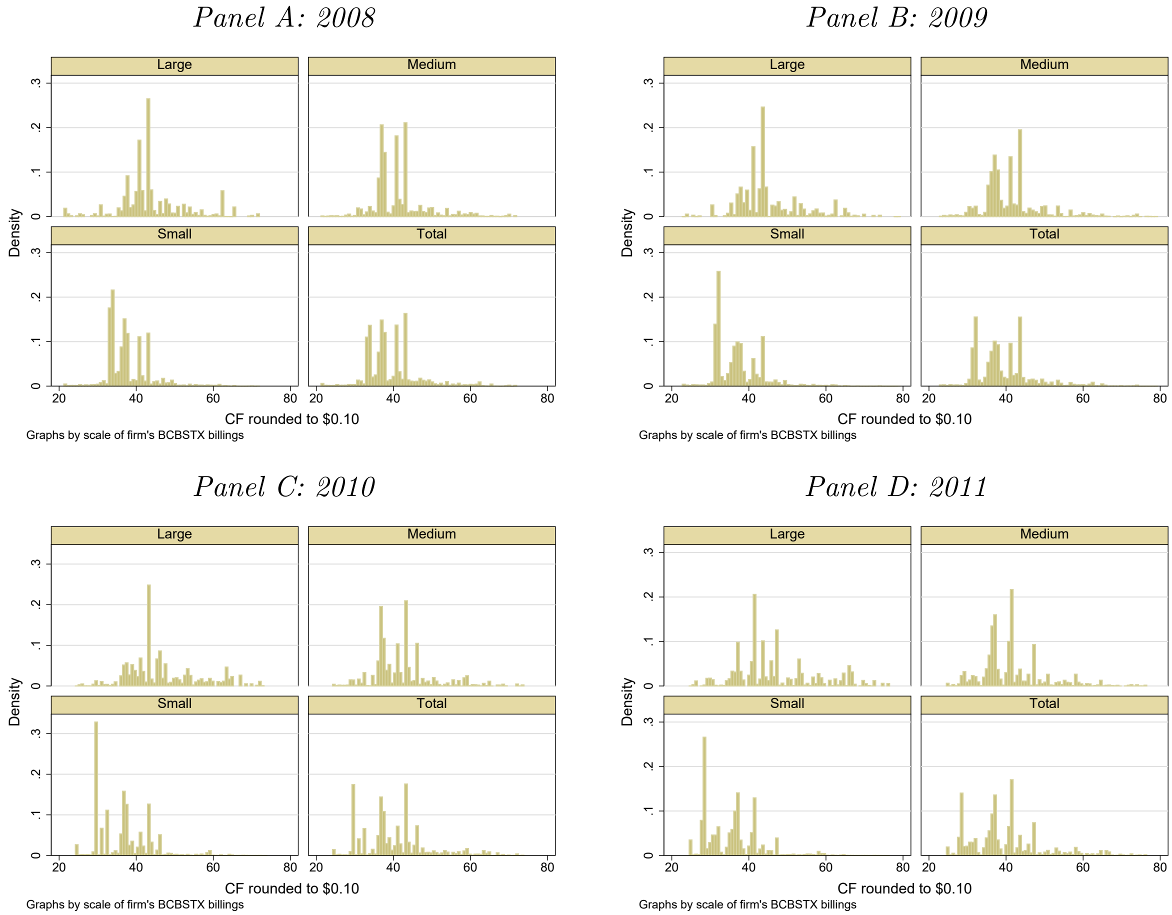
Figure C.1: Examples of Updates to Individual Services



Note: the figure presents monthly averages of BCBS's log payment for the service named in each panel's title. All data are from calendar year 2010. BCBS implemented its update from the 2009 to 2010 relative value scales on July 1, 2010, as indicated by the vertical dashed line each panel.

[140]Both Medicare and private sector payment policies generate nonlinear payments in certain circumstances when multiple instances of the same service are provided per claim.

Figure C.2: Distribution of ICFs by Firm Size



Note: the figure reports the distributions of common Implied Conversion Factors that we compute in each year. We require that common ICFs account for 10 percent of a group's claims, when rounded to the nearest $0.02. Each year's distributions are split according to the sizes of the physician groups, measured as the dollar value of the group's BCBS billings.

Table C.1: Data Cleaning

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Year:	2008		2009		2010		2011	
Measure:	Claims	Spending	Claims	Spending	Claims	Spending	Claims	Spending
(A) Initial dataset	45.5m	$3.49b	54.7m	$4.09b	57.6m	$4.29b	61.7m	$4.64b
(B) Basic cleaning	90.0%	80.2%	90.7%	80.8%	90.0%	80.0%	90.3%	80.4%
(C) In-network outpatient	74.0%	59.6%	75.5%	61.1%	76.5%	61.5%	77.3%	62.3%
(D) Exclude rare codes	73.9%	59.3%	75.3%	60.8%	76.5%	61.3%	77.3%	62.1%
(E) Medicare code merge	41.3%	47.3%	40.3%	47.1%	41.7%	47.8%	41.3%	47.8%

Note: this table quantifies the data lost at each step of our data cleaning and merge process. We show calculations for each of the four years of BCBS claims data. For each year, row (A) shows the raw number of claims (odd-numbered columns) and money spent (even-numbered columns) in that year's claims data. All subsequent rows show the share of claims on row (A) that remain after each set of cleaning steps. Row (B) shows the share of data remaining when we keep only claim lines for which the payment is non-missing, the service quantity is 1, and the observation is an "original" claim line rather than an adjustment to a past payment. These basic cleaning steps eliminate about ten percent of claims and twenty percent of spending. Row (C) further restricts our sample to the universe we consider, namely outpatient in-network claims. This eliminates approximately 15 percent more claims, and twenty percent more spending per year. Row (D) drops those relatively rare service codes for which we have fewer than 10 observations prior to the RVU updates in a given year; this has minimal effect on the sample sizes. Finally, row (E) drops claims that don't merge with Medicare's RBRVS codes. This loses 12–15 percent of observations per year. Source: Authors' calculations using claims data from BCBS.

Table C.2: Alternative Measures of Pricing According to Common Implicit Conversion Factors

Panel A: 2008				
Benchmarking Measure:	Services	Dollars	Services Q1	Dollars Q1
Rounding for ICFs:				
\$0.02	67%	60%	68%	62%
\$0.10	73%	66%	74%	67%
\$0.20	77%	71%	78%	72%
Panel B: 2009				
Benchmarking Measure:	Services	Dollars	Services Q1	Dollars Q1
Rounding for ICFs:				
\$0.02	67%	60%	68%	62%
\$0.10	73%	66%	74%	67%
\$0.20	77%	70%	78%	71%
Panel C: 2010				
Benchmarking Measure:	Services	Dollars	Services Q1	Dollars Q1
Rounding for ICFs:				
\$0.02	87%	83%	88%	84%
\$0.10	89%	86%	89%	86%
\$0.20	89%	87%	90%	87%
Panel D: 2011				
Benchmarking Measure:	Services	Dollars	Services Q1	Dollars Q1
Rounding for ICFs:				
\$0.02	86%	81%	86%	82%
\$0.10	87%	85%	88%	85%
\$0.20	88%	85%	88%	85%

Note: each cell shows the share of services for which payments are associated with a common Implied Conversion Factor (cICF), as defined in the main text. The different cells within a panel show this statistic according to slightly different measures and using different rounding thresholds to define cICFs. The column labeled "Rounding" indicates the rounding applied to each estimated ICF. We then declare an ICF to be "common" for the payments to a physician group if it accounts for at least 5 percent of the group's services in a given year. The first column shows the share of services priced using cICFs, just as in Table 4.2. The column labeled "Dollars" shows a dollar-weighted measure. The dollar-weighted estimates are lower than the service-weighted measure because lower-value services are more likely to be priced using common ICFs. The remaining columns report equivalent measures for which the claims data are restricted to the first quarter of a given year. Source: Authors' calculations using claims data from BCBS.

Table C.3: Firm Size and Implied Conversion Factors

	(1)	(2)	(3)	(4)	(5)
	Panel A: Heterogeneity in Medicare Benchmarking				
Dependent variable:	Share of claims linked to Medicare				
Firm Size (Log Spending)	-0.042**			-0.043**	-0.017**
	(0.005)			(0.005)	(0.005)
Firm Market Share		-0.149**		0.110**	0.041
		(0.022)		(0.031)	(0.032)
Market Concentration			-0.139**	-0.191**	-0.078*
			(0.024)	(0.035)	(0.039)
	Panel B: Firm Size and Implied Conversion Factors				
Dependent variable:	Log implied conversion factor (ICF)				
Firm Size (Log Spending)	0.058**			0.058**	0.040**
	(0.004)			(0.005)	(0.006)
Firm Market Share		0.241**		-0.158**	-0.092**
		(0.015)		(0.037)	(0.029)
Market Concentration			0.238**	0.318**	0.159**
			(0.020)	(0.036)	(0.028)
	Panel C: Firm Size and Deviations from ICFs				
Dependent variable:	Log deviation from ICF				
Firm Size (Log Spending)	0.047**			0.045**	0.030**
	(0.008)			(0.008)	(0.007)
Firm Market Share		0.183**		0.021	0.080^{+}
		(0.038)		(0.054)	(0.047)
Market Concentration			0.151**	0.068	-0.048
			(0.042)	(0.061)	(0.049)
N	20,736,449	20,736,449	20,736,449	20,736,449	20,736,449
No. of Clusters	23,098	23,098	23,098	23,098	23,098
Code Effects	No	No	No	No	Yes
HSA Fixed Effects	No	No	No	No	Yes

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. This table shows estimates of relationship between features of physicians' contracts and measures of firm size and/or market concentration. The construction of all variables is discussed in the main text. Source: Authors' calculations using claims data from BCBS.

Table C.4: Medicare Benchmarking by Betos Category

	(1)	(2)	(3)	(4)
Dependent variable:	Payments with Common Conversion Factors			
	Spending Share		Service Share	
	Panel A: 2008 (N=516,189)			
Imaging	-0.172**	-0.287**	-0.132*	-0.256**
	(0.046)	(0.048)	(0.053)	(0.057)
Procedures	-0.169**	-0.282**	-0.150*	-0.271**
	(0.039)	(0.045)	(0.059)	(0.069)
Tests	-0.188**	-0.275**	-0.111*	-0.206**
	(0.041)	(0.044)	(0.054)	(0.059)
Constant	0.730**	0.405**	0.750**	0.398**
	(0.028)	(0.042)	(0.037)	(0.055)
	Panel B: 2009 (N=593,779)			
Imaging	-0.158**	-0.270**	-0.105*	-0.228**
	(0.040)	(0.042)	(0.047)	(0.050)
Procedures	-0.159**	-0.287**	-0.144**	-0.282**
	(0.033)	(0.040)	(0.050)	(0.059)
Tests	-0.174**	-0.261**	-0.098+	-0.193**
	(0.036)	(0.042)	(0.050)	(0.060)
Constant	0.712**	0.397**	0.736**	0.393**
	(0.027)	(0.036)	(0.033)	(0.048)
	Panel C: 2011 (N=651,901)			
Imaging	-0.391**	-0.441**	-0.273**	-0.336**
	(0.058)	(0.053)	(0.031)	(0.026)
Procedures	-0.318**	-0.371**	-0.351**	-0.417**
	(0.026)	(0.025)	(0.053)	(0.051)
Tests	-0.384**	-0.420**	-0.258**	-0.303**
	(0.040)	(0.034)	(0.048)	(0.045)
Constant	0.895**	0.808**	0.921**	0.814**
	(0.012)	(0.013)	(0.019)	(0.018)
Omitted Category	Evaluation & Management			
Additional Controls	Group Size	None	Group Size	None

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. This table shows estimates of the η_b coefficients in equation (4.12), namely the relationship between Betos category and the Medicare-linked share of services (columns 1 and 2) or spending (columns 3 and 4) at the group-service code level. Medicare links are measured using the common Implied Conversion Factors (cICFs) defined in section 4.4.1. Columns 1 and 3 show estimates after controlling for vigintile of group size, as measured with BCBS spending, and columns 2 and 4 show estimates without group size controls. Standard errors are two-way clustered (Cameron et al., 2011) by Betos category and physician group. Sources: Authors' calculations using claims data from BCBS.

C.3 Estimation in Changes and Threats to Identification

This appendix justifies our measure of Medicare benchmarking based on estimation in simple differences, in section C.3.1. Appendix C.3.2 then discusses potential bias from active renegotiations of physician-insurer contracts contemporaneously with the implementation of Medicare RVU updates. Finally, Appendix C.3.3 computes the bias that would result in such a case.

C.3.1 Estimation in Changes

We simplify our main estimating equations to two time periods in order to see the Medicare-private price relationships as transparently as possible. This approach will also clearly highlight the assumptions necessary for our estimate of $\hat{\beta}$ to equal the true Medicare-linked share σ. Averaging equation (4.10) within each time period, and then taking the difference across the two, yields:

$$\Delta\overline{\ln(P_{g,j})} = \alpha + \beta\Delta\ln(RVU_j) + (1-\sigma)\overline{\varepsilon_{g,j}}. \tag{C.7}$$

In the context of price changes for one service, this equation shows how we can directly interpret the evidence from Figure 4.2C. This graph showed BCBS average log payments for a standard office visit increasing by 70 percent of the Medicare log RVU change. Hence the implied estimate of σ, in the absence of contemporaneous active negotiations, is also 70 percent.

C.3.2 Threats to Identification From Active Renegotiations

This interpretation is threatened by the possibility of actively negotiated changes in $\ln(\theta_g)$ and $\ln(\rho_{g,j,p})$, which would show up in the error term. If they also covary with the updates to Medicare's relative values, then our estimate of $\hat{\beta}$ would be biased relative to the true parameter σ. (We compute the bias in Appendix C.3.2 below.) This might arise endogenously because changes in Medicare's relative values could alter groups' bargaining positions, and perhaps do so differentially across services. We quantify the potential influence of these changes on our estimates of Medicare's benchmarking in two ways.

First, note that when we estimate β on the full sample of physician groups, it could be biased away from σ by active renegotiations of both $\ln(\rho_{g,j,t})$ and $\ln(\theta_{g,t})$. If we estimate β on the data for a single firm, however, $\Delta\ln(\theta_g)$ is a constant. In the levels specification of equation (4.10), we can similarly account for changes in each group's average log payment by allowing for a full set of group-by-period effects. If estimates of β change little as a result of adding firm-by-period effects to such a specification, we can rule out the possibility that changes in the overall level of each firm's payments are biasing our attempt to recover σ.

Second, the channel through which active renegotiations might bias our attempt to recover σ involves changes in bargaining power *induced* by the RVU changes.[141] The threat to our estimation takes the following form: BCBS may pursue renegotiations with firms whose average Medicare payment has fallen, with these negotiations resulting in declines in their payments. Similarly, physician groups whose average Medicare payment has increased may pursue renegotiations with BCBS, with these negotiations resulting in increases in their payments. This pattern would imply a positive bias to our estimates of σ. To investigate the potential relevance of this source of bias, we first construct the average change in the RVUs for the specific services provided by each firm. This allows us to gauge the extent to which each firm is affected. We then investigate whether we obtain larger estimates $\hat{\beta}$ on a sample of firms that were significantly affected compared with firms that experienced little change in their average RVUs.

C.3.3 Deriving the Bias in our Medicare Link Estimate

The biased coefficient $\hat{\beta}$ we would estimate from equation (C.7) in the presence of simultaneous updates to non-benchmarked prices or group-specific markups is:

$$\begin{aligned}
\hat{\beta} &= \frac{Cov[\Delta\overline{\ln(P_{g,j})}, \Delta\ln(RVU_j)]}{Var[\Delta\ln(RVU_j)]} \\
&= \frac{Cov[\sigma\Delta\overline{\ln(\phi_g)} + \sigma\Delta\ln(RVU_j) + (1-\sigma)\Delta\overline{\ln(\rho_{g,j})} + \Delta\epsilon_{g,j,p}, \Delta\ln(RVU_j)]}{Var[\Delta\ln(RVU_j)]} \\
&= \sigma\frac{Cov[\Delta\ln(RVU_j), \Delta\ln(RVU_j)]}{Var[\Delta\ln(RVU_j)]} + \sigma\frac{Cov[\Delta\overline{\ln(\phi_g)}, \Delta\ln(RVU_j)]}{Var[\Delta\ln(RVU_j)]} \\
&+ (1-\sigma)\frac{Cov[\Delta\overline{\ln(\rho_{g,j})}]}{Var[\Delta\ln(RVU_j)]} + \frac{Cov[\Delta\epsilon_{g,j,p}, \Delta\ln(RVU_j)]}{Var[\Delta\ln(RVU_j)]} \\
&= \sigma + \sigma\frac{Cov[\Delta\overline{\ln(\phi_g)}, \Delta\ln(RVU_j)]}{Var[\Delta\ln(RVU_j)]} + (1-\sigma)\frac{Cov[\Delta\overline{\ln(\rho_{g,j})}, \Delta\ln(RVU_j)]}{Var[\Delta\ln(RVU_j)]},
\end{aligned} \tag{C.8}$$

[141] Actively negotiated payment changes that are driven by the RVU updates themselves may plausibly covary with these changes. There is no *a priori* reason to suspect that changes renegotiated for other reasons would covary with the RVU updates and bias our estimates.

where the third equality follows from the properties of covariances and the fourth from the fact that $\frac{Cov[\Delta \ln(RVU_{j,t}), \Delta \ln(RVU_j)]}{Var[\Delta \ln(RVU_j)]} = 1$ and $\frac{Cov[\Delta \epsilon_{g,j,t}, \Delta \ln(RVU_j^M)]}{Var[\Delta \ln(RVU_j)]} = 0$.

One separate source of bias in the estimate of $\hat{\beta}$ could arise if the linked share σ varies across firms and services. This would imply additional terms in equation (C.8) describing our regression estimates, involving covariances between the RVU updates used for identification and the service-by-group linked shares $\sigma_{j,g}$. Recovering σ also requires us to assume that these covariance terms are 0, which will be true if updates to Medicare's rates are uncorrelated with the $\sigma_{j,g}$. In section 4.1, we will allow for heterogeneity across various dimensions in the linked shares.

When thinking about our estimates and any potential bias, it is essential to remember that our estimates are based only on those services with RVU updates. The service code fixed effects ensure that codes without updates don't influence $\hat{\beta}$.

C.3.4 Checks for the Relevance of Active Contract Renegotiation

The estimates presented in Figure 4.2 and Table 4.4 may differ from the true Medicare benchmarking parameter σ if changes in other terms of providers' contracts covary with the changes in RVUs. Indeed, payment changes that significantly alter physician groups' average Medicare payment can move private payments in subsequent years, due in part to the resulting changes to their bargaining positions (Clemens and Gottlieb, 2017). We thus draw on institutional detail and theoretically motivated specification checks to explore how much our estimates might deviate from the true share of payments benchmarked to Medicare's relative values.

The most relevant institutional detail is the relatively short time horizon of our event studies. Dunn and Shapiro (2015) report that physician contracts tend to remain in force for around 3 years. Within each of our single-year event studies, we thus anticipate that roughly one-third of the groups in our sample engage in active contract re-negotiations, which could affect our estimates. Unlike the payment changes analyzed by Clemens and Gottlieb (2017), which significantly shifted certain specialties' average Medicare payments, those we consider here are relatively diffused across specialties, so unlikely to affect groups' overall outside options.

Nevertheless, we investigate the potential relevance of active contract renegotiation with two analyses. First, we consider the potential effect of scheduled RVU changes on a firm's bargaining position. We construct a variable that, for each firm, reports the average change in RVUs for the services it provides. Firms experiencing a negative average change have seen their bargaining positions deterioriate. Firms experiencing an average RVU increase have seen their bargaining positions improve. Using the average RVU change to which each firm was exposed, we construct an indicator for groups whose bargaining positions were significantly affected.

Second, we investigate the potential relevance of changes in groups' average log reimbursement by adding full sets of group-by-period fixed effects to our specification. For this regression, we restrict our sample to the 100 largest firms in each year, primarily for computational ease. Note, however, that large firms are precisely those for which we would expect active renegotiations to be most frequent.

Table C.7 presents these results. Column 1 reports our baseline specification, unchanged from Table 4.4. Column 2 allows our coefficient of interest to vary with an indicator for whether a firm's average Medicare reimbursement rate was significantly affected by a year's RVU updates. The point estimate on this interaction varies across years, but is negative in each case. This is the opposite of what we would expect if significant RVU updates were driving active contract renegotiations. Column 3 limits the baseline specification to the services provided by the 100 largest physician groups. A comparison of column 3 with column 1 reveals that, on average across the years we analyze, the largest firms have contracts that are less linked to Medicare than are contracts in the full sample, a result that we explore further in section 4.6. Most relevant for our current purposes, however, column 4 reveals that adding group-by-period effects to the previous specification has essentially no impact on our coefficient of interest. These results provide evidence against the concern that that active contract renegotiations confound the relationship between BCBS's and Medicare's payments over the intervals we analyze. Thus they bolster the case for interpreting our estimates of $\hat{\beta}$ as unbiased estimates of the fraction of services tied directly to Medicare.

Table C.5: Other Years' Estimates of Benchmarking Using RVU Changes

	(1)	(2)	(3)	(4)
Dependent variable:	*Log private reimbursement rate*			
	Panel A: All Services: 2008 RVU Updates			
Log RVU Change × Post	0.602**	0.597**	0.539**	0.602**
	(0.061)	(0.061)	(0.060)	(0.061)
N	19,552,096	19,552,096	19,552,096	19,552,096
No. of Clusters	3,505	3,505	3,505	3,505
	Panel B: All Services: 2009 RVU Updates			
Log RVU Change × Post	0.778**	0.778**	0.792**	0.778**
	(0.081)	(0.078)	(0.070)	(0.081)
N	21,941,227	21,941,227	21,941,227	21,941,227
No. of Clusters	3,807	3,807	3,807	3,807
	Panel C: All Services: 2011 RVU Updates			
Log RVU Change × Post	0.704**	0.689**	0.679**	0.704**
	(0.046)	(0.052)	(0.048)	(0.046)
N	25,404,007	25,404,007	25,404,007	25,404,007
No. of Clusters	4,091	4,091	4,091	4,091
Group-by-Code Effects	Yes	No	Yes	Yes
Code Effects	No	Yes	No	No
Cubic Time x RVU Change	No	No	Yes	No
Cubic Time x Post	No	No	No	Yes
Weighting	Service	Service	Service	Service

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. The table shows the results of OLS specifications of the forms described in section 4.4.2. Each column in each panel reports an estimate of $\hat{\beta}$ from equation (4.10). Panel A shows estimates using RBRVS updates and BCBS claims data for 2008, Panel B for 2009, and Panel C for 2011. Observations are at the claim-line level and are equally weighted. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Additional features of each specification are described within the table. The construction of all variables is further described in the main text. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

Table C.6: Dollar-Weighted Estimates of Benchmarking Using RVU Changes

	(1)	(2)	(3)	(4)
Dependent variable:	*Log private reimbursement rate*			
	Panel A: All Services: 2008 RVU Updates			
Log RVU Change × Post	0.421**	0.413**	0.359**	0.420**
	(0.075)	(0.075)	(0.071)	(0.075)
N	19,552,096	19,552,096	19,552,096	19,552,096
No. of Clusters	3,505	3,505	3,505	3,505
	Panel B: All Services: 2009 RVU Updates			
Log RVU Change × Post	0.618**	0.627**	0.669**	0.618**
	(0.046)	(0.045)	(0.052)	(0.046)
N	21,941,227	21,941,227	21,941,227	21,941,227
No. of Clusters	3,807	3,807	3,807	3,807
	Panel C: All Services: 2011 RVU Updates			
Log RVU Change × Post	0.749**	0.739**	0.738**	0.749**
	(0.044)	(0.043)	(0.047)	(0.044)
N	25,404,007	25,404,007	25,404,007	25,404,007
No. of Clusters	4,091	4,091	4,091	4,091
Group-by-Code Effects	Yes	No	Yes	Yes
Code Effects	No	Yes	No	No
Cubic Time × RVU Change	No	No	Yes	No
Cubic Time × Post	No	No	No	Yes
Weighting	Dollars	Dollars	Dollars	Dollars

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. The table shows the results of OLS specifications of the forms described in section 4.4.2. Each column in each panel reports an estimate of $\hat{\beta}$ from equation (4.10). Panel A shows estimates using RBRVS updates and BCBS claims data for 2008, Panel B for 2009, and Panel C for 2011. Observations are at the claim-line level and are weighted according to each service's average payment during the baseline period. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Additional features of each specification are described within the table. The construction of all variables is further described in the main text. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

Table C.7: Checks for the Relevance of Active Contract Negotiations

	(1)	(2)	(3)	(4)
Dependent variable:	*Log private reimbursement rate*			
	Panel A: All Services: 2009 RVU Updates			
Log RVU Change × Post	0.778**	0.847**	0.696**	0.666**
	(0.081)	(0.085)	(0.093)	(0.081)
Log RVU Change × Post × Update Impact		-0.077		
		(0.114)		
N	21,941,227	21,941,227	4,097,283	4,097,283
No. of Clusters	3,807	3,807	3,496	3,496
	Panel B: All Services: 2010 RVU Updates			
Log RVU Change × Post	0.750**	0.992**	0.740**	0.747**
	(0.038)	(0.076)	(0.048)	(0.052)
Log RVU Change × Post × Update Impact		-0.393**		
		(0.099)		
N	23,933,577	23,933,577	4,708,213	4,708,213
No. of Clusters	3,681	3,681	3,450	3,450
	Panel C: All Services: 2011 RVU Updates			
Log RVU Change × Post	0.704**	0.804**	0.544**	0.523**
	(0.046)	(0.084)	(0.051)	(0.067)
Log RVU Change × Post × Update Impact		-0.162		
		(0.106)		
N	25,404,007	25,404,007	5,069,260	5,069,260
No. of Clusters	4,091	4,091	3,825	3,825
Group × Post-Update Effects	No	No	No	Yes
Sample	Full	Full	Largest Firms	Largest Firms

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. The table shows the results of OLS specifications of the forms described in section 4.4.2. Column 1 replicates the baseline specification from column 1 of Table 4.4. Column 2 augments the baseline specification with interaction terms allowing the effect of RVU updates to vary with the extent of the average impact of each year's RVU updates on a physician group's average Medicare reimbursement rate. In columns 3 and 4 the sample is restricted to each year's 100 largest physician groups, as sorted by total bills submitted. The specification in column 3 is the baseline specification, while the specification in column 4 includes a full set of post-by-group interactions. Panel A shows estimates using RBRVS updates and BCBS claims data for 2009, Panel B for 2010, and Panel C for 2011. Observations are at the claim-line level and are equally weighted. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Additional features of each specification are described within the table. The construction of all variables is further described in the main text. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

Table C.8: Public-Private Payment Links Across Service Categories

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Dependent variable:	*Log private reimbursement rate*						
	Panel A: 2008 RVU Updates by Betos Categories						
	Evaluation	Imaging	Procedures	Tests	Imaging Sub-Categories:		
					Global	Technical	Professional
Log RVU Change	0.541***	0.644***	0.495***	0.786***	0.665***	0.494***	0.945***
× Post-Update	(0.115)	(0.092)	(0.116)	(0.055)	(0.103)	(0.112)	(0.228)
N	9,851,995	3,221,634	3,851,609	1,292,912	1,688,102	192,569	1,340,963
No. of Clusters	207	1,069	1,817	385	400	235	434
	Panel B: 2009 RVU Updates by Betos Categories						
	Evaluation	Imaging	Procedures	Tests	Imaging Sub-Categories:		
					Global	Technical	Professional
Log RVU Change	0.857**	0.775**	0.399**	0.933**	0.702**	0.769**	0.680**
× Post-Update	(0.209)	(0.066)	(0.064)	(0.052)	(0.072)	(0.068)	(0.184)
N	11,498,770	3,524,642	3,861,539	1,449,803	1,769,522	222,026	1,533,094
No. of Clusters	219	1,133	2,036	388	422	262	449
	Panel C: 2011 RVU Updates by Betos Categories						
	Evaluation	Imaging	Procedures	Tests	Imaging Sub-Categories:		
					Global	Technical	Professional
Log RVU Change	0.794**	0.616**	0.900**	0.439*	0.816**	0.692**	0.709**
× Post-Update	(0.065)	(0.100)	(0.075)	(0.221)	(0.048)	(0.067)	(0.058)
N	13,116,657	3,696,733	5,233,336	1,659,485	1,929,095	193,577	1,574,061
No. of Clusters	238	1,143	2,246	436	424	264	455

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. The table shows the results of OLS specifications of the forms described in section 4.4.2. The cells in each panel report estimates of $\hat{\beta}$ from equation (4.10), with samples selected to contain the HCPCS codes falling into individual broad service categories. The name of the relevant service category accompanies each point estimate. Panel A shows estimates using RBRVS updates and BCBS claims data for 2008, Panel B for 2009, and Panel C for 2011. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Additional features of each specification are described within the table. The construction of all variables is further described in the main text. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

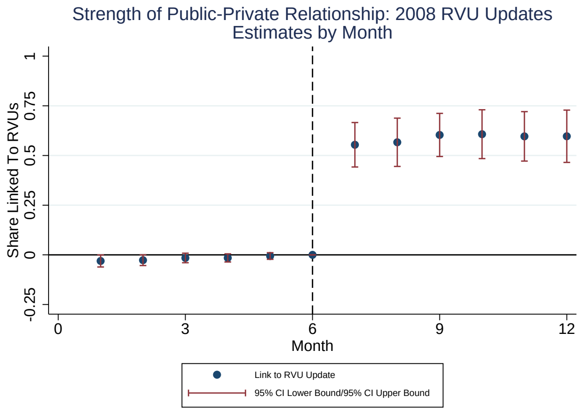
Table C.9: Medicare Benchmarking by Firm Size

	(1)	(2)	(3)	(4)
Dependent variable:	*Log private reimbursement rate*			
Panel A: 2008 RVU Updates (N = 19,552,096)				
Log RVU Change	0.602**	0.560**	0.421**	0.418**
× Post-Update	(0.061)	(0.074)	(0.075)	(0.089)
Log RVU Change		0.130*		-0.059
× Post-Update × Midsize		(0.065)		(0.072)
Log RVU Change		-0.000		0.064
× Post-Update × Large		(0.101)		(0.085)
Panel B: 2009 RVU Updates (N = 21,941,227)				
Log RVU Change	0.778**	0.755**	0.618**	0.756**
× Post-Update	(0.081)	(0.090)	(0.046)	(0.070)
Log RVU Change		0.078		-0.110
× Post-Update × Midsize		(0.059)		(0.071)
Log RVU Change		-0.035		-0.271*
× Post-Update × Large		(0.094)		(0.109)
Panel C: 2011 RVU Updates (N = 25,404,007)				
Log RVU Change	0.704**	0.812**	0.749**	0.774**
× Post-Update	(0.046)	(0.063)	(0.044)	(0.052)
Log RVU Change		-0.140+		-0.036
× Post-Update × Midsize		(0.075)		(0.100)
Log RVU Change		-0.183*		-0.023
× Post-Update × Large		(0.075)		(0.116)
Firm Size × Post-Update Controls	No	Yes	No	Yes
Weighting	Services	Services	Dollars	Dollars

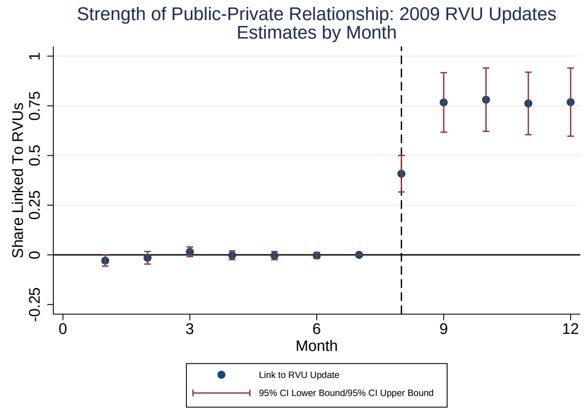
Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. Columns 1 and 3 report the baseline estimates from Table 4.4 Panels A and B respectively. In columns 2 and 4 we augment these specifications to include interactions between firm size indicators variables and both the "Post" indicator and the interaction between the "Log RVU Change" and "Post" indicator. The omitted category is small firms, defined as those with less than $200,000 in billings. Mid-sized firms are those with billings between $200,000 and $1 million, and large firms are those with billings exceeding $1 million. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

Figure C.3: Strength of Public Private Payment Relationships

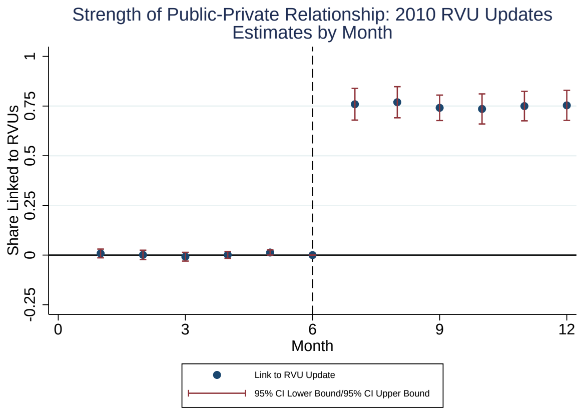
Panel A



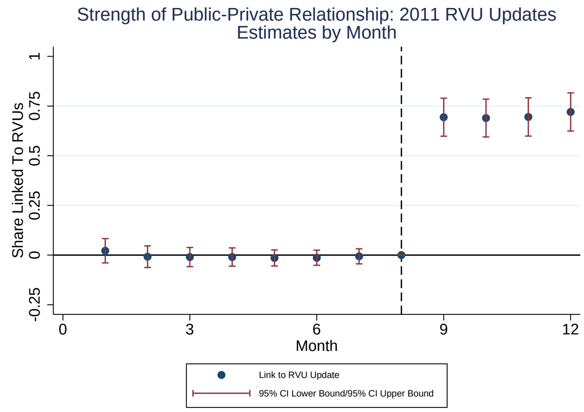
Panel B



Panel C

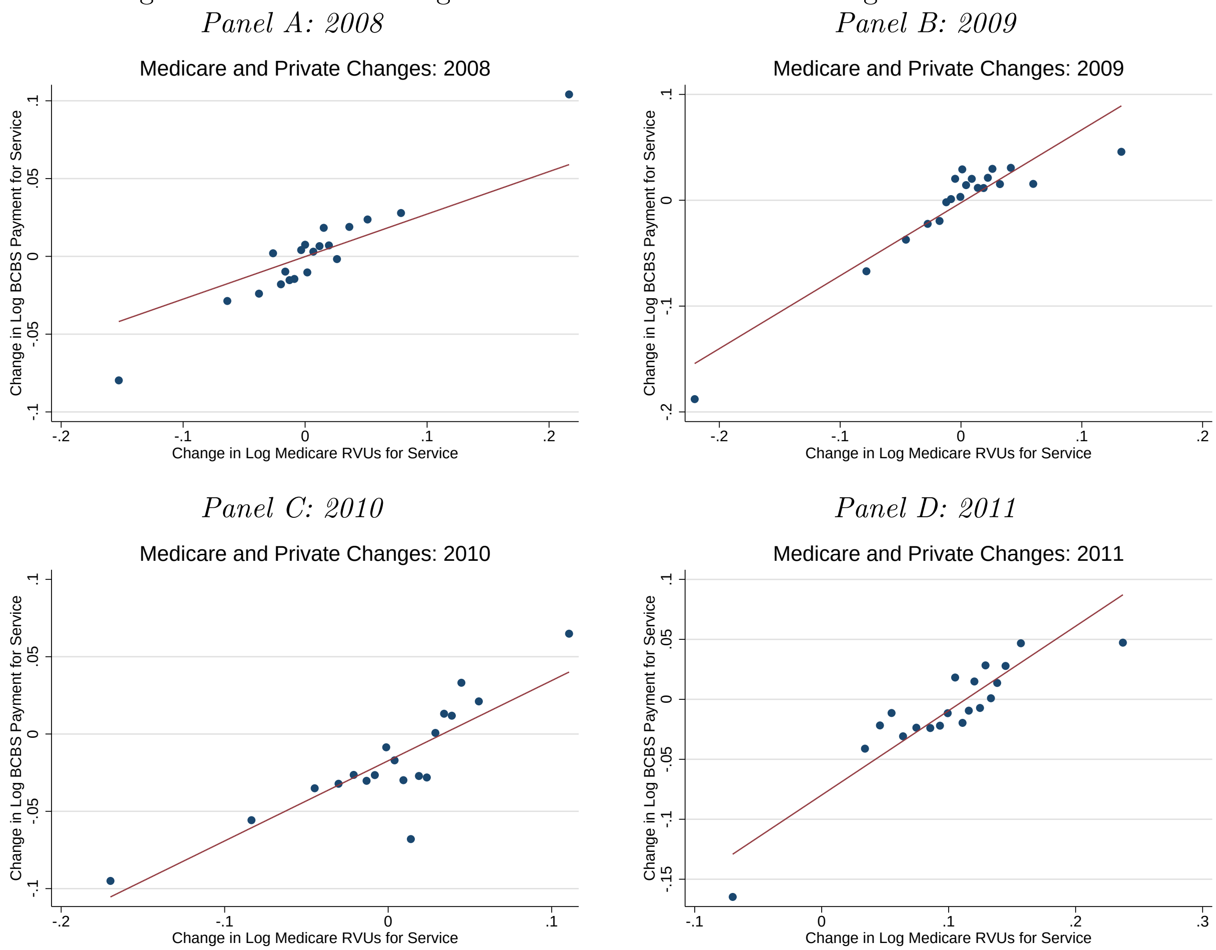


Panel D



Note: the figure reports estimates of the β_p from estimates of equation (4.11). The vertical dashed line in each panel corresponds with the month during each year in which BCBS implemented its update from the prior year's relative value scale. These updates occurred on July 1, 2008, August 15, 2009, July 1, 2010, and September 1, 2011.

Figure C.4: Benchmarking Estimates Based on Price Changes Across Services



Note: the figure reports the relationships described by equation (C.7) for RVU updates in each year, and estimates of that equation. Each panel shows a separate year's estimates, measured as log differences between the period before BCBS implemented the Medicare RVU updates and the period after this update. The years are split at July 1, 2008, August 15, 2009; July 1, 2010; and September 1, 2011. The regressions are run at the underlying service level, but observations are grouped into twenty bins for each year, based on vigintiles of the Medicare log RVU change.

C.4 Extensions

C.4.1 External Validity: Data from Colorado

To gauge the external validity of our BCBS findings, we also use data from the state of Colorado's All-Payer Claims Database (APCD). The APCD is administered by the Center for Improving Value in Health Care (CIVHC), from whom we obtain the data. These data are structured similarly to the BCBS data, but include many insurers. Our analysis focuses on the payments of one Colorado insurer for which we were able to obtain the necessary information on fee schedule updates to implement our primary estimation framework.

In this exercise, we conduct a similar analysis to our baseline changes estimation using the claims data from this Colorado insurer. Just as in Table 4.4 in the main text, we use institutional detail on the timing with which the insurer adopted updates to Medicare's payments. This insurer adopts Medicare's payment updates less frequently than BCBS of Texas. Over the period for which we have the Colorado claims data, this insurer shifted from the 2008 version of Medicare's relative rate structure to the 2010 version on July 1, 2011.

The results are shown in Figure C.6 and Table C.10. The table reports our estimates of equation (4.11) using these changes. The estimates imply that roughly 40 percent of its payments are linked to Medicare's relative rate structure. This insurer thus appears to deviate from Medicare payments to a greater degree than does BCBS of Texas.

Beyond the insurers we analyze directly, anecdotal and documentary evidence suggest that our results apply more broadly. Provider newsletters and industry magazines describe price-setting that frequently uses Medicare as a benchmark, but sometimes deviates. Among many examples, *Managed Care* magazine writes that insurers' "talks with doctors on fee-for-service rates often begin with Medicare's rates" (Carroll, 2007). Clemens and Gottlieb (2017) provide other related examples.

C.4.2 Supply Responses

To determine whether pricing at this granular level has impacts on real resource use, we estimate how relative price changes across services affect physicians' supply of care. We use the same Medicare price changes as in our main analysis, which means that our estimates have three economically salient features. First, they involve short-run responses within a calendar year. Second, they involve responses to changes in the profitability of some services relative to others rather than to across-the-board changes in reimbursement rates.[142] Finally, they involve private payment changes that result from contractual links to changes in Medicare's relative rates.[143]

To measure supply responses, we estimate an analogue of the changes regression shown in Panel C of Figure 4.2 in which the dependent variable is now the change in log quantity of care. We again split the year into two time periods: before and after BCBS implemented the year's Medicare updates. The change in the log number of instances that a given physician group provided a particular service across these two time periods is our dependent variable.

Figure C.7 shows the results of this estimation for each year, along with a binned scatterplot of the underlying data. Note that this is a reduced-form estimate; it relates the Medicare price change to the supply responses for privately insured patients.

In order to estimate the BCBS own-price supply elasticities, we next develop an IV framework. We use the same reimbursement changes that follow from Medicare's RVU updates in the following two-stage least squares setup:

$$\Delta\overline{\ln(P_{g,j})} = \alpha + \beta\Delta\ln(RVU_j) + \varepsilon_{g,j} \tag{C.9}$$

$$\Delta\ln(Q_{g,j}) = \gamma + \delta\Delta\widehat{\overline{\ln(P_{g,j})}} + \epsilon_{g,j}. \tag{C.10}$$

The first stage, equation (C.9), is taken from equation (C.7) in the text. This estimates the share of private prices that respond to the Medicare RVU updates. This generates a predicted price change, which we use in the second stage equation (C.10).

The coefficient δ that we estimate in equation (C.10) is close to providing an estimate of the physicians' supply elasticity for BCBS patients, in response to BCBS prices. It is somewhat confounded, however, by the fact that the BCBS prices are changing at the same time as the prices of physicians' outside option—treating Medicare patients.[144] This would tend to bias the estimates down relative to a pure own-price supply estimate.

Table C.11 shows the results. The IV estimates scale up the reduced form estimates substantially, and range from 0.15 to 0.66. The median estimate of 0.37 occurs in 2011. For comparison, the conceptually most similar estimates in the literature are those of Brekke et al. (2015). Brekke et al. (2015) estimate physicians' supply responses to a reimbursement change for one particular service, which is also the type of price change we consider here. These are different types of elasticities than those of Clemens and Gottlieb (2014), who consider market-wide changes, or the relative price changes of Gruber et al. (1999) and Jacobson et al. (2010).

These positive supply elasticities imply that the pricing decisions we examine have meaningful implications for how physicians provide treatment. If Medicare sets prices inefficiently, then copying Medicare's relative prices leads to inefficient care. When insurers deviate from Medicare rates, these positive supply responses suggest that physicians respond to payment adjustments as the insurers presumably intend.

[142] Our focus here on relative supply responses across services makes this analysis somewhat comparable to Gruber et al. (1999) or, more recently, Brekke et al. (2015).

[143] This final characteristic makes it natural to think about a clear causal chain in our setting where prices influence the subsequent supply of care. When Medicare reimbursement changes lead to a renegotiation, as might happen over a longer time horizon, then we would have to consider the price renegotiations and supply decisions as jointly determined.

[144] Clemens and Gottlieb (2013, Appendix B) model these forces.

C.4.3 Out-of-Network Payments

Our analysis thus far only includes in-network payments—those made to physician groups that have agreed with BCBS on mutually acceptable payment rates. We next show analogous results for out-of-network payments, which occur when providers have not reached any such agreement. When a BCBS-insured patient sees an out-of-network provider, the ultimate payment reflects a complex interaction of the provider's charge, after-the-fact negotiations (Mahoney, 2015, as in), and the insurance plan's coverage. So out-of-network payments are less likely to depend on a convenient benchmark such as the Medicare fee schedule. This analysis allows us to determine whether the benchmarking that we document reflects active decisions as opposed to a purely mechanical force.

Table C.12 replicates Table 4.4 in the main text, but for out-of-network payments. Table C.13 is a dollar-weighted version of the same regressions. In both cases, we obtain small and precisely estimated coefficients. This means that out-of-network payments—which don't represent the outcome of the *ex ante* negotiations we described in section 4.2—are not priced in the same way.

Table C.14 complicates the analysis somewhat. It reveals that around half of out-of-network services appear to be priced according to cICFs. This share is much larger than the results from Tables C.12 and C.13 would suggest, though still far below the in-network results from Table 4.2 in the main text. The difference with the in-network results is especially pronounced in 2010 and 2011, and when using a more stringent cICF threshold (20 percent). In these cases, only 30 percent of out-of-network prices appear to be benchmarked to Medicare, compared with 70 percent of in-network payments. Nevertheless, the ambiguity over the correct definition again demonstrates the advantage of the update-based benchmarking measure in Tables C.12 and C.13.

In short, we find much weaker—if any—Medicare benchmarking in out-of-network payments. The difference between these results and our in-network estimates suggests that the in-network prices reflect active efforts to negotiate around a simplified payment schedule.

Table C.10: Estimating Medicare Benchmarking Using RVU Changes: Colorado

	(1)	(2)	(3)	(4)
Dep. variable:	Log private reimbursement rate			
Log RVU change × Post	0.368**	0.354**	0.455**	0.408**
	(0.145)	(0.146)	(0.098)	(0.102)
Provider-Code FE		Yes		Yes
Insurance Plan Controls			Yes	Yes
N	509,929	509,929	508,038	508,038
No. Clusters	1,471	1,471	1,471	1,471

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. The table shows the results of OLS specifications of the forms described in section 4.4.2. Each column reports an estimate of $\hat{\beta}$ from equation (4.10). Observations are at the claim-line level and are equally weighted. Data are from 2011. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Additional features of each specification are described within the table. The construction of all variables is further described in the main text. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from the Colorado APCD.

Table C.11: Supply Elasticity Estimates

	(1)	(2)	(3)	(4)
Year:	2008	2009	2010	2011
Dependent variable:	Change in log service quantity			
	Panel A: Reduced Form			
Log RVU change for service	0.027	0.095*	0.339***	0.252***
	(0.047)	(0.047)	(0.050)	(0.038)
	Panel B: IV Estimates			
Log BCBS payment change for service	0.052	0.152*	0.658***	0.365***
	(0.090)	(0.076)	(0.102)	(0.055)
N	63,526	71,354	81,294	89,936
First Stage F-Statistic	358.9	776.2	483.9	1843.0

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. Panel A estimates reduced form relationships analogous to the one shown in Panel D of Figure 4.2 in the main text. Panel B shows the second stage estimates from the IV framework in equation (C.10). The robust first-stage F-statistics all easily satisfy the weak instruments test of Olea and Pflueger (2013). Source: Authors' calculations using claims data from BCBS.

Table C.12: Estimating Medicare Benchmarking for Out-of-Network Payments Using RVU Changes

	(1)	(2)	(3)	(4)
Dependent variable:	*Log private reimbursement rate*			
	Panel A: All Services: 2009 RVU Updates			
Log RVU Change × Post	0.018	0.007	0.084*	0.018
	(0.043)	(0.047)	(0.037)	(0.043)
N	2,585,681	2,585,681	2,585,681	2,585,681
No. of Clusters	2,456	2,456	2,456	2,456
	Panel B: All Services: 2010 RVU Updates			
Log RVU Change × Post	0.302**	0.351**	0.170**	0.302**
	(0.073)	(0.074)	(0.044)	(0.073)
N	2,386,575	2,386,575	2,386,575	2,386,575
No. of Clusters	2,051	2,051	2,051	2,051
	Panel C: All Services: 2011 RVU Updates			
Log RVU Change × Post	0.106*	0.094+	0.047	0.105*
	(0.047)	(0.054)	(0.037)	(0.047)
N	2,626,264	2,626,264	2,626,264	2,626,264
No. of Clusters	2,473	2,473	2,473	2,473
Group-by-Code Effects	Yes	No	Yes	Yes
Code Effects	No	Yes	No	No
Cubic Time × RVU Change	No	No	Yes	No
Cubic Time × Post	No	No	No	Yes
Weighting	Service	Service	Service	Service

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. The table shows the results of OLS specifications of the forms described in section 4.4.2, except using data from out-of-network payments. Each column in each panel reports an estimate of $\hat{\beta}$ from equation (4.10). Panel A shows estimates using RBRVS updates and BCBS claims data for 2009, Panel B for 2010, and Panel C for 2011. Observations are at the claim-line level and are equally weighted. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Additional features of each specification are described within the table. The construction of all variables is further described in the main text. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

Table C.13: Dollar-Weighted Estimates of Medicare Benchmarking for Out-of-Network Payments Using RVU Changes

	(1)	(2)	(3)	(4)
Dependent variable:	*Log private reimbursement rate*			
	Panel A: All Services: 2009 RVU Updates			
Log RVU Change × Post	0.036	0.004	-0.043	0.036
	(0.048)	(0.055)	(0.079)	(0.048)
N	2,585,681	2,585,681	2,585,681	2,585,681
No. of Clusters	2,456	2,456	2,456	2,456
	Panel B: All Services: 2010 RVU Updates			
Log RVU Change × Post	0.244**	0.315**	0.203*	0.242**
	(0.063)	(0.066)	(0.082)	(0.063)
N	2,386,575	2,386,575	2,386,575	2,386,575
No. of Clusters	2,051	2,051	2,051	2,051
	Panel C: All Services: 2011 RVU Updates			
Log RVU Change × Post	-0.016	-0.045	0.053	-0.016
	(0.068)	(0.075)	(0.066)	(0.067)
N	2,626,264	2,626,264	2,626,264	2,626,264
No. of Clusters	2,473	2,473	2,473	2,473
Group-by-Code Effects	Yes	No	Yes	Yes
Code Effects	No	Yes	No	No
Cubic Time × RVU Change	No	No	Yes	No
Cubic Time × Post	No	No	No	Yes
Weighting	Dollars	Dollars	Dollars	Dollars

Note: **, *, and + indicate statistical significance at the 0.01, 0.05, and 0.10 levels respectively. The table shows the results of OLS specifications of the forms described in section 4.4.2, except using data from out-of-network payments. Each column in each panel reports an estimate of $\hat{\beta}$ from equation (4.10). Panel A shows estimates using RBRVS updates and BCBS claims data for 2009, Panel B for 2010, and Panel C for 2011. Observations are at the claim-line level and are equally weighted. Standard errors are calculated allowing for arbitrary correlation among the errors associated with each HCPCS service code (including modifiers for the professional and technical components of diagnostic imaging services). Additional features of each specification are described within the table. The construction of all variables is further described in the main text. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from BCBS.

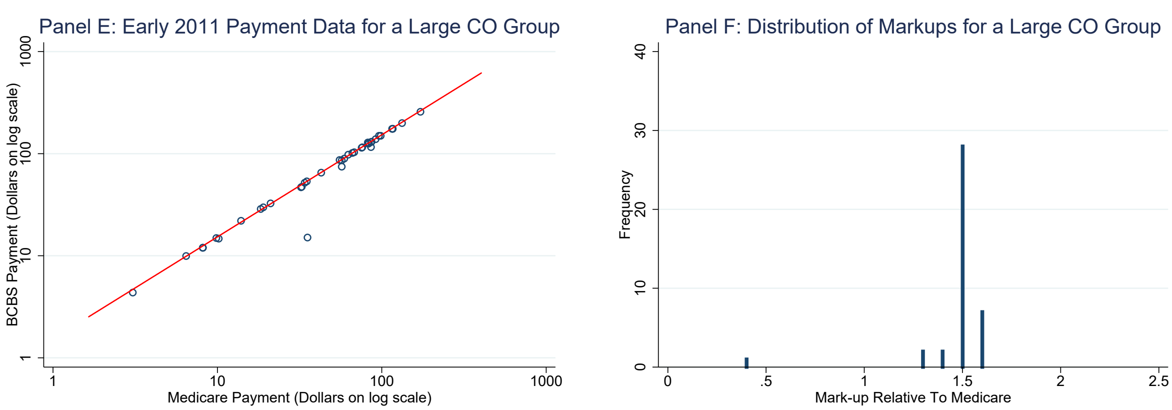
Table C.14: Out-of-Network Services Priced According to Common Implied Conversion Factors

	Frequency Threshold:		
	5%	*10%*	*20%*
Panel A: 2009			
Rounding for ICFs:			
$0.02	54%	42%	26%
$0.10	60%	46%	30%
$0.20	64%	52%	35%
Panel B: 2010			
Rounding for ICFs:			
$0.02	57%	45%	32%
$0.10	61%	48%	34%
$0.20	65%	52%	37%
Panel C: 2011			
Rounding for ICFs:			
$0.02	57%	43%	29%
$0.10	61%	47%	32%
$0.20	66%	51%	35%

Note: each cell shows the share of out-of-network services for which payments are associated with a common Implied Conversion Factor (cICF), as defined in the main text. The cells within each panel show how this share varies as we apply different thresholds for the frequency required to quality as a cICF. The column labeled "Rounding" indicates the rounding applied to each estimated ICF. An ICF is defined as "common" for the payments to a physician group if it accounts for at least the fraction of services associated with the specified Frequency Threshold. Source: Authors' calculations using claims data from BCBS.

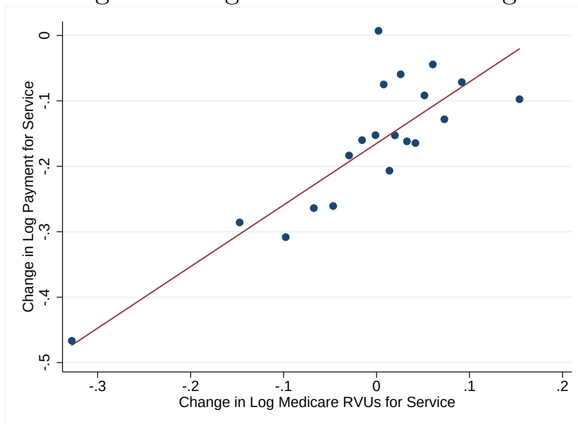
Figure C.5: Raw Payments For Illustrative Physician Group: Colorado

Panel A *Panel B*



Note: this figure shows a payment scatterplot and distribution of markups analogous to those in Figure 4.1, but for a physician group in Colorado. In Panel A, each observation is a unique reimbursement paid for a particular service to the group. The line has a slope of 1 (in logs) and represent the group's most common Implied Conversion Factor. Panel B plots the distribution of markups relative to the Medicare rates for all payments the group received. It shows a clear spike at the value that we identify as a common Implied Conversion Factors in Panel A. To comply with confidentiality rules, we omit from these graphs a small share of each group's claims. The share of claims whose observations are suppressed is 17.17%. Sources: Authors' calculations using claims data from CO APCD.

Figure C.6: Validating Bunching-Based Benchmarking Measure: Colorado

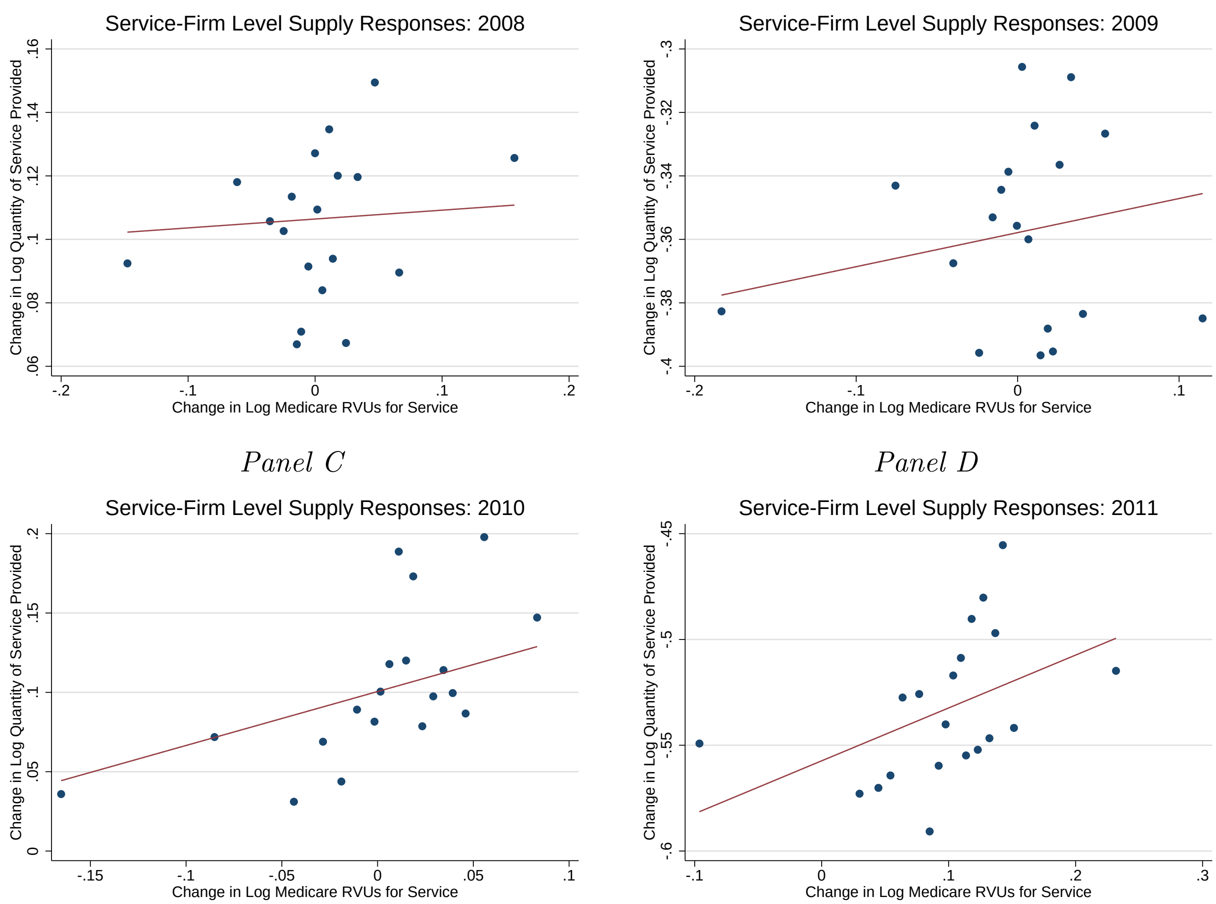


Note: this graph is analogous to Panel C of Figure 4.2, but using 2011 data from one insurer in the Colorado APCD, with the sample split at July 1, 2011. Private price changes are computed as the difference between service-level average payments after and before July 1, 2010. The regression is estimated at the underlying service-code level. Sources: Authors' calculations using updates to Medicare's RBRVS as reported in the Federal Register and claims data from Colorado APCD.

Figure C.7: Short-Run Supply Responses to Medicare Price Changes

Panel A

Panel B



Note: the figure reports estimates of physicians' supply responses to Medicare price changes that BCBS implemented in a given year. Quantities, the dependent variable, are computed at the service-by-firm level. Each panel shows a separate year's estimates, measured as log differences between the period before BCBS implemented the Medicare RVU updates and the period after this update. The years are split at July 1, 2008, August 15, 2009; July 1, 2010; and September 1, 2011. The estimates have very different intercepts across the three panels because of the differences in the share of the year's data that are included in the periods before *vs.* after each year's update.