Essays on Gender and Children

Daniel Nicholas Drew Brown

Lincoln College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2017

Essays on Gender and Children

Daniel Nicholas Drew Brown

Lincoln College
University of Oxford

Michaelmas 2017

Abstract

This thesis investigates the causes of childhood outcomes. In Chapter 1, written with Elisabetta De Cao, we show that unemployment causes child neglect in the United States, during the period from 2004 to 2012. We identify this effect using a Bartik instrument, which we create by combining national-level unemployment rates across industries with differences in the initial industrial structure across counties. We find evidence that parents in the United States lack the social and private safety nets needed to provide for their children's basic needs in the event of a job loss.

In Chapter 2, written with Sonia Bhalotra, we find that fathers work less after having a first born son rather than a daughter in the United Kingdom, during the period from 2009 to 2016. The effect occurs on both the extensive and intensive margins, and in the first year of the child's life. Our identification strategy exploits the randomness of the gender of the first child. We find evidence that degree educated fathers and their partners both take time off work to increase the total parental time invested in sons, because boys' development is less advanced than girls' development at a young age. We find evidence that non-degree educated fathers take time off work to reduce the burden of childcare on their partners, which helps their partners to return to work sooner after childbirth.

Finally, in Chapter 3, I find that female leaders elected to the State Legislative Assembly improve child learning in their districts in rural India, during the period from 2008 to 2014. The effect occurs for both reading and mathematics, and the gender gap in mathematics narrows. To identify this effect, I instrument for the fraction of seats won by a woman in a district with the fraction of seats won by a woman in a close election against a man. The most plausible explanation for the effects is that female leaders try to improve educational outcomes around the time of an election, in the hope of re-election.

Supervisors

This thesis was supervised by James Fenske (University of Warwick) and Paul Collier (University of Oxford).

Total Words

This thesis contains approximately 84,000 words.

Statement of Authorship

Chapter 1 is co-authored with Elisabetta De Cao (Centre for Health Service Economics and Organisation, University of Oxford). I conducted the empirical analysis, including gaining access to and setting up the NCANDS dataset. I wrote the paper. I had the initial idea for the paper and took the lead on the conceptual content. I contacted the Child Protective Services in each state to understand the process of child maltreatment reporting. Elisabetta conducted the literature review, which forms part of Section 1. She created the three maps, and contributed to the conceptual content of the paper.

Chapter 2 is co-authored with Sonia Bhalotra (Department of Economics, University of Essex). I conducted the empirical analysis and wrote the paper. Sonia and I both contributed to the conceptual content of the paper. Sonia had the initial idea for the paper and took a supervisory role on the project.

Chapter 3 is wholly my own work.

Acknowledgement of Funding

I am very grateful to the Economic and Social Research Council, the Economics Department at the University of Oxford, and Lincoln College for their generous financial support.

Acknowledgements

I would like to start by thanking my co-authors, Sonia Bhalotra and Elisabetta De Cao, for their work and guidance on Chapters 1 and 2. It was a genuine pleasure working with you both. I am grateful to my supervisors James Fenske and Paul Collier for their advice over the past four years. I would also like to thank Abhijeet Singh for his help at a difficult time.

A few of my friends in the Economics Department have provided me with invaluable technical and moral support. I might finally stop pestering my desk-mate Viviana Perego, your advice has been incredibly useful. Thanks also to Rachel Cassidy and Jonathan Lain.

I owe a long-overdue thank you to my first Economics teacher, Alan Fearnley. His passion for the subject, unshackled by the school curriculum, was the main inspiration behind my choice to do Economics at university. He commuted the length of London every day, and by delaying retirement for a year gave my classmates and I the best possible chance with our university applications.

Graduate study is famously arduous, but it's considerably easier when your life outside of work is as enjoyable as mine has been over the past few years. For that reason, I would like to say thank you to the friends I have made during my time in Oxford. Listing everyone's name could end up being a fourth chapter in itself, so I'll keep it short. Firstly, thanks to a golden generation in the Lincoln MCR. Secondly, thanks to all my teammates at OUAFC. We shared once-in-a-lifetime experiences, from Beijing to California, and it was an absolute privilege both on and off the football pitch.

I have been surrounded by selfless people throughout my life, none more so than my family. To my Mum, Penny, my Dad, Patrick, and my Sister, Chloe Brown; and to my Grandma, Pamela, and my Grandad, Clifford Morris, thank you. The family you get is a lottery, and I won. Your sacrifices are the main reason behind any successes I will ever have.

Finally, thank you to my partner and best friend, Claire Simon. Writing this thesis has come with more than its fair share of setbacks, and I truly wouldn't have managed without you.

Contents

Introduction

This thesis is a collection of three essays that investigate the causes of childhood outcomes. I look at the role of direct parental investments and the influence of the wider political context in which a child grows up. In Chapters 1 and 2, I consider the interaction between the labour market experiences of parents, parental investments in their children, and child outcomes, in two developed countries. In Chapters 2 and 3, I consider how gender relates to child development. By contrast to Chapters 1 and 2, Chapter 3 focuses on a developing country.

A growing literature, spanning several academic disciplines, demonstrates that early childhood cognitive and non-cognitive skills matter for outcomes in adulthood (for a review, see Heckman & Mosso (2014)). Mild shocks to parental investments in pre-school years, which affect those childhood skills, can in turn have large effects on adult educational attainment, employment, earnings and health (e.g., Heckman et al. (2013); Almond et al. (2017)). In addition, child abuse and neglect can have harmful long-term consequences, for example leading to mental health problems, substance abuse and even a higher probability of engaging in crime (e.g., Dube et al. (2003); Springer et al. (2007); Currie & Tekin (2012)). Disadvantages in early life can be difficult to reverse. Interventions that target adolescents are generally ineffective, at least compared to those aimed at young children (Heckman & Mosso, 2014). Together, this body of evidence shows just how vital it is to create an environment conducive to development in a child's formative early years.

In Chapter 1, which is co-authored with Elisabetta De Cao, we ask: what is the impact of unemployment on child maltreatment in the United States? Child maltreatment includes neglect, physical, sexual and emotional abuse and therefore captures some of the worst child outcomes. There currently exists little causal evidence about the effect of economic conditions on child maltreatment, which we address in this chapter.

We use a new county-level dataset containing one observation for every reported incident of child maltreatment made to the Child Protective Services, covering the whole of the United States for the period from 2004 to 2012. In the baseline regressions, we identify a causal effect of unemployment on child maltreatment using a Bartik instrument. We create this instrument as the weighted average of the national-level unemployment rates across each of twenty industries, where the weights are the county-level fractions of the employed working-age population in each industry the year before the start of the sample period. We check the robustness of the results to four alternative choices of instrument, three of which are variations on the baseline Bartik instrument, and one of which is based on changes in the real exchange rate.

We find that a one percentage point increase in the unemployment rate causes a 20 percent increase in neglect. In a county with the median prevalence of neglect, which has 536 cases per year, this is equivalent to an additional 110 cases per year. The effect is driven by the sharp increase in the unemployment rate following the onset of the recession, and is largest for children aged 0-4 years old. According to the aforementioned literature on child development, this could have damaging long-term consequences for that generation in adulthood. We show that the results capture an effect of unemployment on the actual incidence of neglect and not reporting behaviour. We do so in three ways, including by demonstrating that there is no effect of unemployment on unsubstantiated reports.

We hypothesize that one reason unemployment causes neglect is that parents in the United States lack safety nets. We consider both private safety nets and those provided by the state. We find that the effect of unemployment is significantly smaller in states that introduce longer extensions to unemployment benefits through the Extended Benefits and Emergency Unemployment Compensation programs. We also find that the effect is significantly greater in counties where a larger fraction of children start the sample period without health insurance.

In addition, we show that unemployment causes repeated neglect, which we might expect because the unemployment shock following the onset of the recession persisted, meaning that income is persistently below the level required to meet children's basic needs. Further, we show that unemployment causes a decrease in real expenditure on basic goods such as food and beverages. A deprivation of these goods could contribute to the increase in neglect.

In Chapter 2, which is co-authored with Sonia Bhalotra, we ask: how does the gender of the first child affect parental labour supply in the UK? Therefore, unlike Chapters 1 and 3, the main dependent variables in Chapter 2 are not child outcomes. Nonetheless, this chapter does have implications for child outcomes because parental labour supply directly affects the time that parents can invest in their children, which in turn affects child development.

We use the six waves of the UK Household Longitudinal Study, a panel dataset which covers the period from 2009 to 2016. We run fixed effects regressions, identifying the gender of first child effect by exploiting changes in behaviour over time within individuals before and after they have their first child. Our identification strategy relies on the assumption that the gender of the first child is random, conditional on the decision to have a child.

We find that fathers work significantly less after having a first born son rather than a daughter. Fathers are 4.5 percentage points less likely to be employed and work 2.2 fewer hours per week on the intensive margin, on average. The effects on both the extensive and intensive margins occur in the first year of the child's life. The effects are approximately one quarter the size of the effect of having a child of either gender for mothers.

The partners of degree educated fathers also take more time off work after having a first born son rather than a daughter. This behaviour is consistent with the hypothesis that parents take time off work to increase the total parental time invested in sons, because boys' cognitive and non-cognitive development is less advanced than that of girls at a young age. We refer to this as the child production function mechanism. We additionally find that these fathers spend their additional time at home engaged with the child, and that parental time engaged with a child is positively correlated with development.

By contrast, the partners of non-degree educated fathers work more after having a first born son rather than a daughter. We hypothesize that these fathers take time off work to reduce the burden of childcare on their partners, which helps their partners to return to work sooner after childbirth. We refer to this as the 'new man' mechanism. Consistent with this, we find that within the group of non-degree educated fathers, it is those with the most progressive attitudes about women's role in the labour market who drive the effect.

We aggregate the effects across the fathers and their partners, and find that a first born son of a degree educated father could receive up to an additional 702 hours of parental time before his fourth birthday, on average, compared to a first born daughter. This difference could help to reduce the sizeable gender gap in early child cognitive and socio-emotional development that we document in both the UKHLS and the Millennium Cohort Study, which is in the region of one fifth to one third of a standard deviation depending on the measure.

Finally, in Chapter 3, I ask: what effect does a female Member of the Legislative Assembly have on child learning in her district, in rural India? Whilst the first two chapters look at the effect of direct parental investments on child development, this chapter looks at the effect of the wider political context.

I use a dataset which contains basic literacy and numeracy test scores for more than two and a half million school-aged children, known as the Annual Status of Education Reports. This dataset covers the period from 2008 to 2014, and can be matched at the district level to electoral data. To identify a causal effect of female leaders, I instrument for the fraction of seats in a district won by a woman with the fraction of seats won by a woman in a close election against a man.

I find that an additional woman elected to the State Legislative Assembly leads to an increase in the probability that a child can do subtraction by 4 percentage points and can read at Grade 1 level by 3.4 percentage points, in a district with the median number of constituency seats. In addition, there is a decrease in the gender gap in mathematics equivalent in size to over half its mean. These effects occur for primary school children, who are aged 6-14 years old, and not secondary. The children studied in this chapter are therefore older than those for whom we find effects in Chapters 1 and 2. Nonetheless, the levels of child development that I consider here are levels that children are expected to achieve at a young age (for example, children are expected to read at Grade 1 level by about age 6).

The effect occurs in the female leader's next election year, which is consistent with an explanation that political leaders increase effort around the time of an election, and for female leaders this effort is particularly focused on education. The effects die out again the following year, and so we should be cautious about interpreting this as evidence of a deep or persistent improvement in child learning. It is perhaps more plausible that female leaders find a way to motivate children and schools to exert short-term effort during an election year. The results must also be taken cautiously because placebo tests reveal that women won in close elections against men in districts where test scores were already higher the year before the women were elected.

Chapter 1

The Impact of Unemployment on Child Maltreatment in the United States

with Elisabetta De Cao (University of Oxford)

Abstract

In this paper, we show that unemployment increases child neglect in the United States during the period from 2004 to 2012. A one percentage point increase in the unemployment rate leads to a 20 percent increase in neglect. We identify this effect by instrumenting for the county-level unemployment rate with a Bartik instrument, which we create as the weighted average of the national-level unemployment rates across each of twenty industries, where the weights are the county-level fraction of the employed working-age population in each industry at the start of the sample period. An important mechanism behind this effect is that parents lack social and private safety nets. The effect on neglect is smaller in states that introduce longer extensions to unemployment benefits, and is greater in counties where an initially larger fraction of children are not covered by health insurance. We find no evidence that the effect is driven by alcohol consumption or divorce.

JEL: I10, J13, K42

Keywords: child maltreatment, neglect, unemployment rate, recession, safety nets.

We thank Sonia Bhalotra, Steve Bond, Marianne Bruins, Paul Collier, James Fenske, Mike Keane and Simon Quinn for helpful comments. We thank participants in the workshop on the Economics of Domestic Violence at the University of Leicester, the European Association of Labour Economists Conference 2017, the 1st IZA Workshop on Gender and Family Economics, the Socioeconomic Differences and Health Later in Life Workshop organised by the University of Groningen, and the Dondena Seminars at Bocconi University for useful discussions. We thank Michael Dineen for his help with accessing the NCANDS data. We thank Fatih Karahan for sharing data on unemployment benefit durations with us. Some data used within this analysis were derived from National Data Archive on Child Abuse and Neglect (NDACAN) restricted data. These data were accessible through contractual arrangements with NDACAN, and are solely available through the Cornell Restricted Access Data Center at CRADC@Cornell.edu. The collector of the original data, the funder, NDACAN, Cornell University, and the agents or employees of these institutions bear no responsibility for the analyses or interpretations presented here. The information and opinions expressed reflect solely the opinions of the authors. All errors are our own.

1.1 Introduction

Child maltreatment is the abuse or neglect of children under 18 years old.[1] In the United States, child maltreatment is a major problem. The U.S. Department of Health and Human Services estimates that there were approximately 683,000 victims in 2015 alone, equivalent to a rate of 9.2 victims per 1,000 children (US Department of Health and Human Services, 2015). Child maltreatment can have adverse effects on child development. It can interrupt the development of the brain and other vital organs, and worsen cognitive function including attention, abstract reasoning, impulse control and long-term memory (e.g., Hagele (2005); Know Violence in Childhood (2017)). In addition, maltreated children can experience insufficient weight gain (Block & Krebs, 2005), and develop abnormal base cortisol levels and cortisol reactivity (Tarullo & Gunnar, 2006). Further, child maltreatment can have long-term consequences on the victims, including mental health problems, substance abuse, lower levels of education, productivity or earnings, and a higher probability of engaging in crime (e.g., Dube et al., 2003; Springer et al., 2007; Currie & Widom, 2010; Currie & Tekin, 2012).

The correlation between economic conditions and child maltreatment has been studied in both the economics and sociology literatures (Elder Jr, 1974; Becker, 1993). Low-income parents are more likely to use physical forms of discipline to alter their children's behaviour because they cannot provide other types of incentives (Weinberg, 2001). At the same time they may lack the resources to provide for their children's basic needs. In addition, the Great Recession was associated with more hospital-diagnosed instances of child head trauma (Huang et al., 2011), high frequency spanking (Brooks-Gunn et al., 2013), and harsh parenting (Lee et al., 2013).

Causal empirical evidence is scarce (Berger & Waldfogel, 2011).[2] Early economic stud-

[1]Child maltreatment is a narrower concept than 'adverse childhood experiences', which also includes, for example, parental separation or divorce, household mental illness or having an incarcerated household member. See https://www.samhsa.gov/capt/practicing-effective-prevention/prevention-behavioral-health/adverse-childhood-experiences for more information. The exact definitions of specific types of child abuse and neglect are presented in Section 1.2.

[2]Berger & Waldfogel (2011) provides an extensive literature review of studies focusing on the economic

ies found only a weak relationship between economic conditions and child maltreatment (Paxson & Waldfogel, 2002; Bitler & Zavodny, 2004; Seiglie, 2004). More recently, Lindo et al. (2013) use administrative data from California and find that male layoffs increase child maltreatment whilst the opposite is true for female layoffs. Stephens-Davidowitz (2013) considers reported cases of maltreatment, rates of child mortality and Google searches for child maltreatment, and finds that the Great Recession led to a decrease in reporting of child maltreatment, but an increase in actual incidences. Schneider et al. (2017) use data from the Fragile Families and Child Wellbeing Study and find that the Great Recession led to a higher risk of child abuse, but a lower risk of child neglect. Nonetheless, these papers assume that area-level unobservable characteristics related to child maltreatment are not correlated with the economic variable considered.

In this paper, we ask whether unemployment causes child maltreatment. We make three main contributions. First, we use a new dataset containing one observation for every reported incident of child abuse and neglect made to the Child Protective Services for nearly every state in the U.S. during the period from 2004 to 2012. Our dataset comes from the National Child Abuse and Neglect Data System (NCANDS), produced by the National Data Archive on Child Abuse and Neglect (NDACAN). A handful of previous papers have used an earlier and more limited version of this dataset, with information available only at the state level and for the mid 1990s (Paxson & Waldfogel, 1999, 2002, 2003). To the best of our knowledge, we are the first authors to use a restricted-access version of this dataset with county-level identifiers. This dataset covers the whole of the U.S., which allows us to improve upon recent articles that consider only one state (Lindo et al., 2013; Frioux et al., 2014; Raissian, 2015).

Second, we identify the causal impact of unemployment on child maltreatment using an instrumental variables approach. We instrument for the county-level unemployment

determinants and consequences of child abuse and neglect. In addition, three studies present suggestive but not conclusive evidence of a causal link between low income and child maltreatment. They are based in the US and focus on a cash assistance program to families at risk of having their child placed in foster care in Illinois (Shook & Testa, 1997), on a welfare reform program in Delaware (Fein & Lee, 2003), and on a randomised child support and welfare reform experiment in Wisconsin (Cancian et al., 2013).

rate using a predicted county-level unemployment rate, which we create as the weighted average of the national-level unemployment rates across each of twenty industries, where the weights are the county-level fraction of the employed working-age population in each industry the year before the start of the sample period, in 2003. This type of instrument has been widely used in the labour economics literature, and is commonly referred to as the Bartik instrument (Bartik, 1991; Blanchard & Katz, 1992). Our approach improves on previous work (Paxson & Waldfogel, 1999, 2002; Bitler & Zavodny, 2004; Seiglie, 2004; Stephens-Davidowitz, 2013; Schneider et al., 2017) since it allows us to isolate exogenous shocks to local labour demand, based on how locally important industries are performing at the national level. The Bartik instrument has also been used in a related literature that investigates the effect of economic conditions on intimate partner violence (e.g., Aizer, 2010; Anderberg et al., 2016).

Third, we investigate the lack of safety nets as a potential mechanism for our results. We ask whether unemployment benefits, a social safety net, can mitigate the adverse effects of poor economic conditions on child maltreatment. We also explore the role of two private safety nets. We look at the effect of health insurance and having two employed parents living in the household. We ask whether unemployment causes a decrease in expenditure on basic goods, and whether unemployment causes repeated cases of child maltreatment.

We find that a one percentage point increase in the unemployment rate leads to a 10 percent increase in overall abuse. We look at the effect of unemployment on different types of maltreatment, and find that the effect on overall abuse is driven by an increase in neglect. A one percentage point increase in the unemployment rate leads to a 20 percent increase in neglect. In the county with the median prevalence of neglect, which has 536 cases per year, this is equivalent to an additional 110 cases per year. We show that the results are robust and capture an effect on the actual incidence of neglect and not reporting behaviour, for example by demonstrating that there is no effect of unemployment on unsubstantiated reports.

We find evidence that both social and private safety nets mitigate the effect of unemployment on neglect. The effect is smaller in states that introduce longer extensions to the duration of unemployment benefits. A one percentage point increase in the unemployment rate at the 25th percentile of the 2008-12 distribution of the duration of benefits (55 weeks) leads to a 21 percent increase in neglect, whilst at the 75th percentile (87 weeks) it leads to only a 14 percent increase. In addition, the effect is greater in counties in which an initially larger fraction of children are not covered by health insurance. For a county with the median poverty rate, a one percentage point increase in the unemployment rate at the 25th percentile of the distribution of the fraction of uninsured children (0.081) leads to a 10 percent increase in neglect, whilst at the 75th percentile (0.156) it leads to a 14 percent increase. We also find that unemployment causes a decrease in real expenditure on basic goods, as we might expect in the absence of safety nets. Finally, we find that unemployment causes repeated neglect. If unemployment persists then income may stay persistently below the level required to meet children's basic needs, in the absence of safety nets. We rule out two alternative explanations, namely that alcohol consumption or divorce drive the effect.

The paper is structured as follows. In Section 1.2, we provide definitions of each type of child abuse and neglect, describe the system of the Child Protective Services and outline the process of child maltreatment reporting in the United States. We present our empirical strategy and describe the NCANDS dataset in Section 1.3. In Section 1.4, we outline the main results. In Section 1.5, we explore mechanisms for those results. We present robustness checks in Section 1.6, and conclude in Section 1.7.

1.2 Context: The Child Protective Services and the Process of Child Maltreatment Reporting in the United States

At the Federal level, child maltreatment is defined by the Child Abuse Prevention and Treatment Act (CAPTA) as: 'Any recent act or failure to act on the part of a parent or caregiver, which results in death, serious physical or emotional harm, sexual abuse or exploitation, or an act or failure to act which presents an imminent risk of serious harm' (Child Welfare Information Gateway, 2014). There exist some differences in the way that specific types of child abuse or neglect are defined across states. Neglect is generally defined as the failure of a parent or other caregiver to provide the necessary food, clothing, shelter, medical care or supervision to the point that the child's health, safety and well-being are threatened with harm. Physical abuse is generally defined as any non-accidental physical injury to the child. The definition of sexual abuse generally includes the encouragement or coercion of a child to engage in any sexually explicit conduct or simulation of such conduct for the production of child pornography, as well as rape, molestation, incest, or prostitution. Emotional abuse is generally defined as injury to the psychological capacity or emotional stability of the child, as evidenced by an observable or substantial change in behaviour, emotional response or cognition (Child Welfare Information Gateway, 2014).

All fifty states and the District of Columbia have a Child Protective Services (CPS) agency, which is responsible for investigating reports of child abuse and neglect.[3] The process of child maltreatment reporting varies by state, but typically works as follows.[4]

[3]The CPS falls under different departments in different states, for example the Department of Health and Welfare in Idaho or the Office of Children and Family Services in New York.

[4]We have contacted every Child Protective Services agency in the United States by phone and email to understand the process of child maltreatment reporting. The information provided in Section 1.2 is based on those phone calls and emails. A useful review of procedures is available for New York state, at the website of the Office of Children and Family Services: `http://ocfs.ny.gov/ohrd/ccg/`, which is similar to procedures in many other states. Recommended procedures for CPS caseworkers can be found

All but ten states have a centralised statewide hotline that reporters can call if they suspect child abuse or neglect.[5] Individuals in some professions, such as teachers and doctors, are mandated to report any suspicion of child maltreatment, but reports can come from any member of the public, for example neighbours, family or friends.[6] Trained specialists at either the state or county hotline receive the call, obtain as much information about the case as possible from the reporter, and make a judgement about whether the case warrants an investigation in accordance with state law. This often requires that the specialists call other agencies, such as law enforcement and schools, to gather additional information.

After the initial phone call, the case is allocated to the CPS office in the county in which the child resides. A CPS caseworker makes initial face to face contact with the family, before undertaking an investigation.[7] During the investigation, the caseworker may talk to the child, to the child's family, as well as professionals who are involved in the child's life. The caseworker will decide whether there is sufficient evidence that child abuse or neglect has taken place. In the event that the report is substantiated, a range of actions can be taken. In extreme cases, the child can be removed from his or her family home for protection. More often, the caseworker will recommend a plan to the family involving, for example, cognitive-behavioural therapy, school-based training, or counselling and other supportive services (Children's Bureau, 2003). The CPS cannot directly prosecute the parents, but they can recommend cases to law enforcement agencies and the courts.

In 2012, the CPS agencies received approximately 3.4 million reports of child abuse or neglect involving approximately 6.3 million children. Of these, 62.0% were investigated,

at Children's Bureau (2003).

[5]Alabama, California, Hawaii, Maryland, Minnesota, North Carolina, North Dakota, South Carolina, Wisconsin and Wyoming do not have a statewide hotline, and so reporters must call the specific county office in the county in which the child resides. In states with a statewide hotline, the initial call is often processed at the state level, but it can be directed to county hotlines.

[6]In some states, such as New Hampshire, everyone is a mandated reporter by law.

[7]Initial face to face contact is made within one or few days of the call depending on the degree of emergency of the report, after which an investigation is undertaken over a longer period, which can be a few months.

leading to a national rate of investigated reports of 28.3 per 1,000 children (US Department of Health and Human Services, 2012). Professionals made 58.7% of reports, with 16.7% made by legal and law enforcement personnel, 16.6% by education personnel and 11.1% by social services personnel (US Department of Health and Human Services, 2012). Of the child-reports that were investigated, 19% of cases were found to be substantiated.[8]

1.3 Identification Strategy and Data

1.3.1 Identification Strategy: The Bartik Instrument

We wish to understand the effect of unemployment on the incidence of child abuse and neglect. Unobservable worker characteristics within a county might be correlated with both the unemployment rate and the incidence of child abuse or neglect in that county. To deal with this concern, we use an instrumental variables approach. We instrument for the county-level unemployment rate using a predicted county-level unemployment rate, which combines national-level unemployment rates across industries with differences in the initial industrial structure across counties. This instrument isolates a measure of local labour demand that is unrelated to local labour supply. It therefore allows us to separate demand-driven unemployment rate shocks from supply-driven shocks that could be correlated with unobservables that are also related to child maltreatment. This approach was introduced by Bartik (1991). It has been used many times in the labour economics literature (e.g., Blanchard & Katz, 1992; Luttmer, 2005; Wozniak, 2010), and has been used recently in papers on violence against women (e.g., Aizer, 2010; Anderberg et al., 2016).

Our instrument is a weighted average of the national-level unemployment rates across each of twenty industries,[9] where the weights are the fraction of the employed working-age population in each industry the year before the start of the sample period, in 2003, in

[8]These figures apply to the Federal Fiscal Year 2012.

[9]For the industry classification, we use the North American Industry Classification System (NAICS), see www.census.gov/eos/www/naics.

the county. National-level unemployment rates are plausibly exogenous to county-level worker characteristics in any individual county, since counties are small in size relative to the whole of the United States (the U.S. consists of 3,143 counties). The initial industrial structure in a county is likely correlated with its workers' characteristics, which is a threat to the validity of the instrument. However, the initial industrial structure is by definition time invariant at the county level, and so we can deal with this threat by controlling for county-level fixed effects, which we do in all regressions. We estimate the following, where equation (1.2) is the first stage of the IV procedure:

$$Y_{cst} = \beta Unemp_{cst} + X'_{cst}\phi + \psi_{st} + \eta_c + \epsilon_{cst} \tag{1.1}$$

$$Unemp_{cst} = \delta(\Sigma_j w_{csj} N_{tj}) + X'_{cst}\theta + \pi_{st} + \tau_c + \upsilon_{cst} \tag{1.2}$$

Here, Y_{cst} is the natural logarithm of the number of abuses per year in county c, in state s, in year t, for the type of maltreatment of interest.[10] We count only allegations that were found to be substantiated.[11] $Unemp_{cst}$ is the unemployment rate in county c, in state s, in year t. The Bartik instrument comprises of the weights, w_{csj}, which are the fractions of employed working-age individuals in each industry, j, the year before the start of the sample period in county c, in state s; and the national-level unemployment rate, N_{tj} in each industry, j, in each time period, t.[12] We control for county fixed effects through η_c, to ensure the validity of our instrument as discussed above.

We take the natural logarithm of the number of abuses because its distribution (even when expressed as a rate per child) has a long right-hand tail, and we do not want the estimates to be dominated by the effects in counties with the most extreme levels

[10] Some counties have zero abuses for particular types of maltreatment in a given year. This varies by the type of maltreatment, but, for example, 10% of county-years have no incidents of physical abuse and 8% of county-years have no incidents of neglect. To ensure that no county-year has a zero, we add 0.001 abuses to every county-year before taking the natural logarithm. We check the robustness of the results to adding 0.01 and 0.0001 abuses in columns (5)-(8) of Table A.8.

[11] In the robustness checks in Section 1.6.1, we also consider unsubstantiated cases.

[12] The industry assigned to an unemployed person is the industry of the last job that person held.

of maltreatment. We control for the natural logarithm of the child population in X'_{cst}, rather than consider the abuse rate per child as the dependent variable, because doing so is a more general approach which does not impose that the coefficient on the natural logarithm of the child population is equal to one (see also Aizer (2010)).

The coefficient of interest is β. This coefficient tells us the percentage change in the number of abuses in a county as a result of an increase in the unemployment rate by one percentage point.

A further identification concern comes from the measurement of the left hand side variables. The measurement of abuse and neglect differs across states and may change over time within states. If the unobservable reasons for these differences are correlated with unemployment, our estimate of β will be biased. Differences in measurement arise from several sources. First, the definitions of child abuse and neglect vary across states and may vary over time within states. Whilst the Child Abuse Prevention and Treatment Act (CAPTA) provides federal definitions, state definitions can differ (Child Welfare Information Gateway, 2014). For example, Washington state does not recognise emotional abuse.[13] Second, some states include specific exceptions in their definitions of child abuse and neglect. For example, in thirty-one states and D.C., an exception is made for parents who choose not to seek medical care for their children due to religious beliefs. Third, states differ in who is mandated to report child abuse. Fourth, states have different systems to determine whether a referral should be classified as substantiated. Since these differences only occur at the state-year level, we can deal with this concern by controlling for state-year fixed effects through ψ_{st}.

In X'_{cst}, we also control for the fractions of the population that are Black, Hispanic, and Other Race (which includes individuals who are Asian, Alaskan Native or American Indian, Native Hawaiian or Other Pacific Islander and Two or More Races). We weight observations by the child population in the county-year, because the child population

[13] There are several other examples, which are summarised in Child Welfare Information Gateway (2014). For example, seven states explicitly include human trafficking in their definition of child sexual abuse. Twenty-five states and D.C. include a failure to educate a child as required by law in the definition of neglect.

varies considerably across counties and we wish to estimate an effect that is representative for children across the U.S.[14]

We cluster standard errors at the state level. Our main concern here is that the CPS is organised at the state level, and so the number of substantiated incidents may be correlated across counties within the same state. This is true even though we control for state-year fixed effects. For example, if CPS workers are re-allocated within the state from one county office to another, changes in the quality of investigation (and so perhaps the number of incidents found to be substantiated) may be correlated across those two counties. We later demonstrate the robustness of results to clustering at the county level.

1.3.2 Data

1.3.2.1 Outcome Variables: Child Abuse and Neglect

We use a dataset which contains every reported incident of child abuse and neglect made to the Child Protective Services in nearly every state in the U.S. for the years 2004-12. To our knowledge, we are the first authors to use a restricted-access version of this dataset with information on the county of the report. This dataset comes from the National Child Abuse and Neglect Data System (NCANDS), produced by the National Data Archive on Child Abuse and Neglect (NDACAN). We focus on reports of neglect, physical, sexual and emotional abuse. For each child maltreatment report, we observe the gender, age and ethnic group of the perpetrator and victim, the report date, the type of maltreatment alleged, the county of the report and the outcome of the investigation.

For each county and year, we create a count of the total number of incidents of each type of maltreatment.[15] If multiple children are maltreated within a single report, we

[14] For example, the year before the start of the sample period in 2003, 1% of counties had 222 or fewer children. By contrast, 25% of counties had a child population of greater than or equal to 15,297, whilst the largest county, Los Angeles, had 2,678,788 children. In Table A.10, we check that the results are robust to dropping counties with a population of more than one million children in 2003, given that those counties receive a large weight in the regressions.

[15] Emotional abuse is not recorded in Washington state in any year, or in: D.C. in 2004, Idaho in 2006, Indiana in 2004-7, Rhode Island 2007, or Vermont in 2012.

count an incident for each of the children. We count only incidents where that specific type of maltreatment is found to be substantiated by the Child Protective Services.[16] We take the natural logarithm of those counts, after first adding 0.001 to all county-years, since some county-years have zero reports of some types of maltreatment. We later demonstrate the robustness of the results to alternative methods of dealing with zeros. In each year, a small number of states do not submit information to NCANDS, which is a voluntary reporting system. The median number of states reporting in each year is 49 (including D.C.), and the lowest is 45 in 2004. Our analysis focuses on a final sample of 2,803 counties from forty-six states.[17]

1.3.2.2 Unemployment Rate

We focus on the annual unemployment rate at the county level, using data from the Local Area Unemployment Statistics (LAUS) produced by the Bureau of Labour Statistics (BLS). The BLS calculates unemployment rates using information collected in the Current Population Survey, Current Employment Statistics survey, and state Unemployment Insurance systems.

1.3.2.3 Bartik Instrument

To calculate the weights for the instrument, we use county-level data from the Quarterly Census of Employment and Wages (QCEW) in 2003. We measure the national-level unemployment rates by industry using the Current Population Survey. Both datasets are

[16] Specifically, we count incidents that are coded as 'substantiated', 'indicated or reason to suspect', and 'alternative response disposition - victim' for the given type of maltreatment, since states differ slightly in the way that they classify the outcome of investigations.

[17] We exclude Alaska, South Dakota, Illinois, North Dakota and Oregon from the analysis. The Child Protective Services in Alaska is organised by boroughs, which have different boundaries to the FIPS counties used in all federal reporting systems including NCANDS. The county of report that Alaska submits to NCANDS is created after the fact, based on computer codes that have changed over time and has a tenuous link to the borough boundaries. Only 25 out of 66 counties in South Dakota are included in the NCANDS data, whilst in Illinois the fraction of counties reporting decreases from all to less than one third between 2010 and 2011. Finally, North Dakota and Oregon do not report any information before 2009, and so are not observed before the start of the recession, the time at which the main unemployment shock occurs in our sample period.

produced by the BLS. We consider a classification of twenty industries provided by the North American Industry Classification System (NAICS).

1.3.2.4 Control Variables

We measure the child population using the Population and Housing Unit Estimates (PHUE), produced by the Census Bureau. The Census Bureau uses data on births, deaths and migration to update decennial census data to produce these estimates. We define a child as any individual between the ages of 0 and 17. We also use the PHUE to measure the fraction of the population that are of each ethnic group (Black, Hispanic and Other Race).

1.3.2.5 Summary Statistics

In Table 1.1, we present means and standard deviations for the main variables we use in the regression analysis. Neglect is considerably more common than the other three types of maltreatment, with the mean number of incidents of neglect more than four times greater than the mean number of incidents of physical abuse, the next most common type of maltreatment. The variance in the number of incidents of all types of maltreatment per year is large, but particularly so for emotional abuse.

The main results are driven by the sharp increase in the unemployment rate between 2007 and 2009, following the onset of the recession. We show this increase in Figure 1.1, in which we plot the weighted average unemployment rate across counties over the sample period. We weight county-years by the child population so that the unemployment rate plotted is representative of where children reside in the U.S. The median county in our sample experienced an increase in the unemployment rate of 4.5 percentage points between 2007 and 2009.

This unemployment shock disproportionately affected the long-term unemployed. In Figure 1.2, we plot the fraction of all unemployed individuals in each of four categories by the duration of unemployment. The long-term unemployed, defined as individuals who have been unemployed for twenty-seven weeks or more,[18] comprised more than forty percent of total unemployment from 2010 onwards, having comprised less than one fifth of total unemployment in 2008.

We show the main sources of variation underlying the regression results in Figures 1.3 and 1.4. In Figure 1.3, we plot the change in the unemployment rate between 2004-6 and 2010-12 after controlling for state-year fixed effects. This map demonstrates that the size of the unemployment shock varies considerably across counties within states. Similarly, in Figure 1.4, we plot the change in the overall abuse rate per 100,000 children between 2004-6 and 2010-12 after controlling for state-year fixed effects. This map highlights the heterogeneity in the change in the overall abuse rate across counties within states, which we seek to explain in the regression analysis.

Finally, in Figures A.1 to A.3, we present some descriptive statistics by the demographic characteristics of the perpetrator and victim. These patterns are interesting as it is rare for datasets on domestic violence to contain detailed demographic information about the victim and perpetrator. Figure A.1 shows the abuse rates by the victim's age. The rate of neglect amongst children aged 0-4 is nearly twice the rate amongst children aged 5-17. The rate of sexual abuse amongst children aged 5-17 is more than three times greater than the rate amongst 0-4 year olds. Figure A.2 shows the overall abuse rate by the perpetrator and victim's gender. Women are relatively more likely to abuse or neglect their children. However, women usually spend more time with children, both within two-parent households and because the incidence of single motherhood is greater than single fatherhood. Whilst these figures are not restricted to incidents perpetrated by parents, for 86.8% of substantiated incidents, at least one of the perpetrators is a parent.[19] Figure A.3 shows the overall abuse rate trends for the least poor and poorest 10% of counties,

[18] This is the BLS definition of long-term unemployment. For more information, see: https://www.bls.gov/bls/cps_fact_sheets/ltu_mock.htm.

[19] To define a parent, we include the following categories from the NCANDS dataset: parent, relative foster parent, non-relative foster parent, legal guardian, foster parent. We exclude: other relative, group home or residential facility staff, child daycare provider, unmarried partner of parent, other professionals, friends or neighbours, and other. The relationship of the perpetrator to the child is recorded in most, but not all, state-years.

categorising counties by their overall poverty rate. We caution that differences in abuse rates between the poorest and least poor counties may be partly driven by differences in definitions of abuse at the state level, by changes in definitions of abuse over time, or by differences in which counties are included in the calculation of the weighted mean in each year.[20] However, with those cautions in mind, this Figure demonstrates that overall abuse is more common in the poorest 10% of counties than the least poor 10% of counties.

1.4 Results

1.4.1 Unemployment Causes Child Neglect

In Table 1.2, we present the main results. In columns (1) and (2), we estimate the effect of unemployment on overall abuse using OLS and IV respectively. The measure of overall abuse combines any incident of neglect, physical, sexual or emotional abuse. We find that the coefficients on the unemployment rate are positive, and for the IV regression the effect is statistically significant at the 5% level. A one percentage point increase in the unemployment rate leads to a 10 percent increase in overall abuse.

The effect on overall abuse is driven by an effect of unemployment on neglect. In columns (3) to (10), we separate out the four different types of maltreatment. The IV results in column (6) demonstrate that a one percentage point increase in the unemployment rate leads to a 20 percent increase in neglect. This effect is statistically significant at the 1% level. In the county with the median prevalence of neglect, which has 536 cases per year, a one percentage point increase in the unemployment rate leads to an increase in neglect by 110 cases per year.[21]

Our results, which are for the U.S. as a whole, contrast with the estimates presented in Lindo et al. (2013) for California. In that single state, they do not find a statistically

[20] Not all counties are included in the calculation in every year because not all states submit information to NCANDS in every year.

[21] The weighted median county-year in the final sample (weighting by the child population) has 536 cases of neglect. Using the estimates in column (6) of Table 1.2, a 20.44 percent increase in neglect translates into an increase by $(0.2044 \times 536 = 110)$ cases.

significant effect of the predicted employment rate on neglect.

We find no statistically significant effect on sexual or emotional abuse in either the OLS or IV regressions, and the point estimates are small. For physical abuse, we do find a small statistically significant effect in the OLS, which is no longer significant in the IV regression. Though the point estimate is larger in the IV regression, it is still less than one third the size of the effect on neglect. We therefore focus on understanding the effect on neglect in the rest of the paper.

The Bartik instrument is highly relevant. We present the first stage results in Table A.1. The Kleibergen-Paap F-statistic on the first stage is 34.9. The coefficient on the Bartik instrument suggests that a one percentage point increase in the predicted unemployment rate leads to a 0.63 percentage point increase in the actual unemployment rate.

For both overall abuse and neglect, the two main IV results, the OLS estimates are smaller and not statistically significant. There are two explanations for these differences. The first is that the OLS estimates suffer from attenuation bias due to classical errors-in-variables measurement error in the county-level unemployment rate. The county-level unemployment rate from the LAUS program is partially model-based. The LAUS program uses county-level data on the number of unemployment insurance claimants. However, new entrants and re-entrants to the labour force who are unemployed cannot claim unemployment insurance, because their current period of unemployment is not preceded by a period of employment. Therefore, the BLS estimates the number of unemployed new entrants and re-entrants, by combining state-level data on labour force entry with county-level demographic information, and this introduces measurement error.[22]

The second explanation is that the OLS estimates suffer from omitted variables bias. One possible omitted variable is the county-level government budget. Following a negative shock to the county government budget,[23] public sector workers may be made

[22]See http://www.bls.gov/lau/laumthd.htm for details.

[23]A shock to the county government budget may arise from a change in its ability to raise local revenues, or simply from the mismanagement of revenues from any source (including state and federal transfers). Local governments rely heavily on revenues raised themselves. For example, in 2012, 63% of

unemployed. In addition, funding for the CPS may decrease.[24] A decrease in funding for the CPS leads to a cut in CPS staff (Eckholm, 2009), which can lead to a decrease in substantiated reports of abuse and neglect for two reasons. First, county hotlines become understaffed, and so reporters who are left on hold for long periods of time may hang up the call and fail to make the report (Stephens-Davidowitz, 2013). Second, the remaining CPS caseworkers are given higher caseloads, which can lead to shortcuts in investigations or in extreme cases reports may not be investigated at all (Buckley & Secret, 2011; Preston, 2013; Santos, 2013). In this case, the county government budget will be negatively correlated with the unemployment rate and positively correlated with maltreatment, which creates a downward bias in the OLS estimates. Shocks to a county government budget will not affect national-level unemployment rates and so the Bartik instrument removes this source of omitted variables bias. To the best of our knowledge, there does not exist a county-level dataset of local government budgets covering our sample period, and so this variable is unobservable. This may be because the structure of county and sub-county governments is complicated and varies by state.[25]

Given the difference between the OLS and IV results, we later demonstrate the robustness of the IV results to four alternative choices of instrument, in Section 1.6.2. One instrument is based on the change in the real exchange rate, and the other three are variations on the baseline Bartik instrument. Further details are given in that Section. Whilst this does not allow us to distinguish between these two explanations for the difference, it does ensure that the results are not a quirk of our preferred choice of instrument.

local government revenue came from its own sources (this figure comes from the Tax Policy Center, at: http://www.taxpolicycenter.org/statistics/local-general-revenue).

[24]Resources for the investigation of child maltreatment can vary widely across counties within states. This can be seen, for example, in Pennsylvania, where the Department of Human Services records a review of expenditure per case at the county level in Pennsylvania Department of Human Services (2016).

[25]County governments exist in every state other than Connecticut, Rhode Island and D.C. There exist four levels of sub-county government: municipal and township governments (general purpose), and school and special districts (special purpose). Which levels of government exist, and which levels have responsibility for the many functions of government varies considerably across states and can even vary across regions within states (U.S. Census Bureau, 2013).

1.4.2 Heterogeneity

In this subsection, we ask whether the effect of unemployment differs by the victim's age and the perpetrator's gender. In Table A.2, we see that the point estimate for the effect on neglect of young children (0-4 years) is more than double that for older children (5-17 years). A one percentage point increase in the unemployment rate leads to a 25 percent increase in neglect of children aged 0-4 years old. In Table A.3, we see that the point estimate for the effect of unemployment on female perpetrated neglect is more than double the size of that for male perpetrated neglect. A one percentage point increase in the unemployment rate increases female perpetrated neglect by 28 percent.

1.5 Mechanisms

1.5.1 Unemployment Causes Neglect Because Parents Lack Safety Nets

We hypothesise that one important mechanism behind the effect of unemployment on neglect is that parents lack safety nets. We define a safety net as any source of cash or in-kind benefit that helps to prevent an individual from falling into poverty in response to a transitory negative income shock, such as a job loss. In the absence of safety nets, unemployment will force parents to decrease expenditure, perhaps to a level that they can no longer provide for their children's basic needs. According to this mechanism, the effect of unemployment on neglect is pecuniary. Parents are financially less able to provide for their children after being made unemployed.

We test this hypothesis in several ways. In Sections 1.5.1.1 and 1.5.1.2, we ask whether the effect of unemployment on neglect is greater where safety nets are lacking. We distinguish between social safety nets, which are safety nets provided by the state, in Section 1.5.1.1, and private safety nets, in Section 1.5.1.2. In Section 1.5.1.3, we directly ask whether unemployment causes a decrease in expenditure on basic goods, which could in turn lead to child neglect. Finally, in Section 1.5.1.4, we ask whether unemployment leads to repeated neglect.

1.5.1.1 Social Safety Nets: Unemployment Insurance

One of the most important social safety nets is Unemployment Insurance (UI), which provides cash benefits to involuntarily unemployed individuals. During the sample period, unemployed individuals who had exhausted benefits from the regular UI program could claim additional benefits through two programs.

The first is the Extended Benefits (EB) program. The EB program allowed individuals to claim an additional 13 to 20 weeks of unemployment benefits in states with high unemployment rates. This program existed before the recession. It was initially partly funded by the state and federal governments, and so many states chose to opt out of the program at the time of onset of the recession. However, the program became entirely funded by the federal government under the American Recovery and Reinvestment Act in February 2009. After that date, many states joined and chose low state-level unemployment rate 'triggers', above which unemployed individuals become eligible for additional benefits (Hagedorn et al., 2013).

The second is the Emergency Unemployment Compensation (EUC) program. The EUC program was introduced in June 2008 in response to the recession, and ended in January 2014. It was entirely federally funded, and again allowed unemployed individuals to claim additional weeks of unemployment benefits if their state-level unemployment rate hit certain thresholds. The program had four tiers. All states were eligible for Tier 1, which provided an extra fourteen weeks of benefits. To qualify for Tiers 2, 3 and 4, a state's seasonally adjusted total unemployment rate had to reach 6%, 7% and 9% respectively.[26] Under EUC, unemployed individuals could potentially receive up to fifty-three additional weeks of unemployment benefits (Hagedorn et al., 2013).

We have a dataset on the duration of unemployment benefits, which comes from

[26] See `https://workforcesecurity.doleta.gov/unemploy/pdf/euc08.pdf`.

state-specific trigger reports provided by the Department of Labor,[27] and was compiled in Hagedorn et al. (2013).[28] The dataset is provided at the quarterly level, and so we take the mean across quarters to make the dataset annual. In Figure 1.5, we plot the distribution of the duration of unemployment benefits across the forty-five states that we include in the regression analysis.[29] From 2004-7, almost every state provided twenty-six weeks of unemployment benefits. In 2008, almost every state provided between thirty-three and thirty-five weeks. After 2008, there was a large increase in the variation in the duration of benefits across states, and it is this variation that drives the results in Table 1.3. The combination of the EB and EUC programs meant that unemployment benefits were extended from the standard twenty-six weeks to as many as ninety-nine weeks in some states.

We can now see that the EB and EUC programs directly targeted the long-term unemployed, since the long-term unemployed are by definition the group who have exceeded the standard twenty-six weeks of unemployment benefits. As we saw in Figure 1.2, long-term unemployment increased sharply during the sample period, and so the EB and EUC programs extended a core social safety net to some of a rapidly growing group of previously uninsured individuals.[30]

We ask whether unemployment has a smaller effect on neglect in the states that introduce longer extensions to the duration of benefits. To do so, we interact the unemployment rate with the duration of unemployment benefits, and to obtain a second instrument, we interact the Bartik instrument with the duration of unemployment benefits. We already control for the level of the duration of unemployment benefits through the state-year fixed effects.

The coefficient on $Benefit\ \ Duration \times Unemployment\ \ Rate$ in column (3) of Table

[27] See `http://ows.doleta.gov/unemploy/trigger/` and `http://ows.doleta.gov/unemploy/euctrigger/`.

[28] We are grateful to Fatih Karahan at the Federal Reserve Bank of New York for providing us with this dataset.

[29] Data on unemployment benefits is not available for Hawaii.

[30] We cannot directly look at the effect of long-term unemployment on neglect, as no measure of the unemployment rate by duration exists at the county level.

1.3 is negative and statistically significant at the 10% level, which demonstrates that extending the duration of benefits is indeed associated with a smaller effect of unemployment on neglect. The size of the effect is large. A one percentage point increase in the unemployment rate at the 25th percentile of the 2008-12 distribution of the duration of benefits (55 weeks) leads to a 21 percent increase in neglect, whilst at the 75th percentile (87 weeks) it leads to only a 14 percent increase. In the county with the median prevalence of neglect, this difference equates to 37 fewer cases of neglect per year in response to a one percentage point increase in the unemployment rate.

At the maximum duration of benefits of ninety-nine weeks, a one percentage point increase in the unemployment rate still leads to an 11 percent increase in neglect. Part of this residual effect may exist because two groups of parents remain ineligible for unemployment insurance even when benefits are extended to the ninety-nine week maximum. The first is the group of parents who have been unemployed for more than ninety-nine weeks. The second is the group of parents who earned less than a state-determined threshold level of earnings in their previous period of employment (Isaacs & Healy, 2014).

In columns (1) and (2), we present the first stage results from this regression. The negative coefficient on the interaction instrument in the first stage for the unemployment rate suggests that states that intervened with longer benefit extensions had counties whose unemployment rates were less badly affected by national shocks. This is intuitive if the states who use policy interventions more pro-actively have been better able to insulate local economic conditions from national shocks.

Contiguous Border Counties Methodology

Unemployment benefits were not extended randomly, and so it is possible that the coefficient on $Benefit\ \ Duration \times Unemployment\ \ Rate$ in column (3) of Table 1.3 does not capture a causal policy effect. Unobservable worker characteristics at the county level may be correlated with the state-level unemployment rate, and so whether triggers for benefit extensions were reached, as well as child neglect. To deal with this, we adopt a contiguous border counties methodology following Dube et al. (2010).[31]

We restrict the sample to pairs of contiguous counties that straddle a state border. We include only county-pair-years where both counties are observed in the NCANDS dataset. We have 1,090 county-pairs, involving 987 different counties, for whom all the variables for the regression analysis are measured.[32] We highlight those 987 counties in the map in Figure A.4. We run the baseline IV regressions on this sample of county-pairs, but now control for county-pair-year fixed effects. We then identify effects by comparing the variation in neglect over time within counties between contiguous counties that straddle a state border. If unobservable worker characteristics are similar across contiguous counties, this methodology allows us to overcome the identification concern expressed above. We therefore estimate the following, where equations (1.4) and (1.5) are the first stages of the IV procedure:

$$Y_{cspt} = \beta_1 Unemp_{cspt} + \beta_2 (Ben_{st} \times Unemp_{cspt}) + X'_{cst}\phi + \psi_{st} + \eta_c + \gamma_{pt} + \epsilon_{cspt} \quad (1.3)$$

$$Unemp_{cspt} = \delta_1(\Sigma_j w_{csj} N_{tj}) + \delta_2(Ben_{st} \times (\Sigma_j w_{csj} N_{tj})) + X'_{cst}\theta + \pi_{st} + \tau_c + \omega_{pt} + \upsilon_{cspt} \quad (1.4)$$

$$(Ben_{st} \times Unemp_{cspt}) = \mu_1(\Sigma_j w_{csj} N_{tj}) + \mu_2(Ben_{st} \times (\Sigma_j w_{csj} N_{tj})) + X'_{cst}\vartheta + \kappa_{st} + \nu_c + \iota_{pt} + \varepsilon_{cspt} \quad (1.5)$$

Here, p denotes the county-pair. The key difference in equation (1.3) from equation (1.1) is the addition of the county-pair-year fixed effects, γ_{pt}. The unit of observation is now the county-pair-year.[33] Each county therefore enters the regression as many times as

[31]This approach has been widely used in the minimum wage literature (for a review, see Allegretto et al. (2017)).

[32]In the United States as a whole, there are 1,308 contiguous county-pairs that straddle a state border.

[33]Y and $Unemp$ vary at the county-year level, not the county-pair-year level, but we add the subscript p to each to highlight that the unit of observation is now the county-pair-year. We must organise the data at the county-pair-year level rather than the county-year level for the following reason. There are more county-pair-year fixed effects than we have degrees of freedom. If we included a dummy variable for every county-pair-year we would be unable to estimate all these coefficients. However, if we organise the data at the county-pair-year level, there is only a single non-zero county-pair-year dummy variable for each observation. This property allows us to demean all the observations within the county-pair-year, treating the county-pair-year dummies as nuisance parameters rather than explicitly including them in the regression. We can then estimate the model using these county-pair-year-demeaned variables (this is the approach taken in Dube et al. (2010)). To estimate equation (1.3), we first demean all the observations

it has border pairs, in a given year. The modal number of contiguous border pairs a county has is two, and the maximum is nine. A county's inclusion in multiple county-pairs will induce a correlation in the error term between those county-pairs. To deal with this, we follow Dube et al. (2010) and use two-way clustering of the standard errors.[34] We cluster at the state level as before, and at the level of the border segment. A border segment is a group of county-pairs that run along the same boundary between two states.[35]

In column (4) of Table 1.3, we first check whether the main effect of unemployment on neglect still exists in the sub-sample of contiguous county-pairs, which it does. In column (5), we check whether the policy heterogeneity still exists in this sub-sample. The point estimate on the interaction term in column (5) is very similar in size to the estimate for the main sample in column (3), but the effect is no longer statistically significant. This may be due to low statistical power since the border counties constitute a little over one third of the number of counties in the main sample. Bearing that in mind, we present the results from the regression which uses the contiguous border counties methodology, in column (6). We find that the absolute size of the coefficient on $Benefit\ \ Duration \times Unemployment\ \ Rate$ increases. The point estimate is nearly fifty percent larger than the estimate in the main sample in column (3). This suggests that the estimates in column (3) may underestimate the true mitigating effect of extending the duration of benefits. However, the effect is not statistically significant, and so the results are not conclusive. This is perhaps unsurprising given that the policy heterogeneity was not significant in column (5), and may be explained by low statistical power.

We look at two other social safety nets for which we have measures of state-level policy differences. These are the Supplemental Nutrition Assistance Program (SNAP) and the Earned Income Tax Credit (EITC).[36] We take the same approach as for UI, but find that

within counties (to implicitly control for county fixed effects), and then demean the resulting variables within county-pair-years.

[34]In addition, we run regressions in which we adjust the weights to mechanically ensure that the weight received by each county-year does not depend on the number of contiguous border pairs it has. To do this, we multiply the original weights by the inverse of the number of contiguous border pairs a county has in each year. The results are very similar and are available upon request.

[35]We have 45 states and 94 border segments in the regressions in Table 1.3.

[36]UI, SNAP, EITC and the Temporary Assistance for Needy Families (TANF) programs constitute

the first stages are not identified. See Table A.4 and Section 1.B.1 in the Appendix for further details. Whilst this is unfortunate, the literature suggests that SNAP and EITC are less counter-cyclically responsive than UI (Bitler & Hoynes, 2016), which is the part of the social safety net that we have assessed.

1.5.1.2 Private Safety Nets

Parents can rely on private safety nets as well as those provided by the state. We observe two county-level measures of private safety net coverage at the start of the sample period. In Section 1.5.1.2.1, we ask whether the effect of unemployment on neglect is greater in counties where a larger fraction of children are not covered by health insurance. Since some children are eligible for free or low-cost health insurance through Medicaid and the Children's Health Insurance Program (CHIP), this is a mix between a private and a social safety net. In Section 1.5.1.2.2, we ask whether the effect of unemployment on neglect is smaller in counties where a larger fraction of children have two employed parents living in the household with them. Private safety nets can take several other forms, for example parents may receive income transfers from family and friends, or dissave. However, we do not observe county-level measures of these other private safety nets.

1.5.1.2.1 Health Insurance

In the absence of health insurance, an unemployed parent may be unable to cover the financial costs of adequate medical care for their children. An existing literature demonstrates that health insurance coverage is associated with greater access to and utilisation of health care services for children in the United States, including well-child preventative care, ambulatory care, and dental care (e.g., Selden & Hudson (2006)). In addition,

the core cash or near-cash elements of the social safety net for the non-elderly (Bitler & Hoynes, 2016). We do not have a measure of state-level policy differences for TANF. States have considerable flexibility in how TANF funds are spent. They can use TANF funds for a range of programs other than direct cash transfers, for example family planning and marriage counselling (Bruins, 2016). Since so many programs exist under the umbrella of TANF, it is difficult to measure state-level policy differences in one or even few dimensions. Measures of total caseloads or total expenditure are likely to capture differences in demand for the program by state, more than they capture differences in the generosity of the program.

the uninsured are less likely to have a regular source of care, which can in turn lead to later diagnoses of new conditions and decrease preventative care (Starfield & Shi, 2004; Leininger, 2009). The uninsured can receive emergency medical care regardless of their ability to pay, due to the 1986 Emergency Medical Treatment and Labor Act.[37] Nonetheless, parents with uninsured children who have received emergency care still face large medical bills that may in turn affect their ability to provide for the children's other basic needs. In addition, doctors may refuse to see the uninsured for non-emergency medical needs (Feldman, 2010).

Our measure of neglect includes medical neglect. We have a county-level dataset of the fraction of children who are not covered by health insurance in 2000, which comes from the Small Area Health Insurance Estimates (SAHIE). We interact the fraction of children initially not covered by health insurance, *Fraction Uninsured Children*, with the unemployment rate, as well as with the Bartik instrument to create a second instrument. We already control for the level of *Fraction Uninsured Children* through the inclusion of the county fixed effects.

We are concerned that the existence of private safety nets could just proxy for the level of poverty. We wish to understand the mitigating effect of private safety nets conditional on the interaction between poverty and unemployment. This matters because the effect of unemployment on neglect is not concentrated in the poorest households. To see this, we use a variable in the NCANDS dataset that indicates whether the house in which the child resides is substandard, overcrowded, unsafe or otherwise inadequate, including homelessness. We separately count the number of cases of neglect for children living in inadequate and adequate housing, and find that the effect of unemployment on neglect occurs for children living in adequate housing in Table A.5. The lack of effect for children in the poorest households might be explained if their parents were not affected by the unemployment shock because they were already unemployed. For further details on this analysis, see Section 1.B.2 in the Appendix.

[37] See https://www.cms.gov/Regulations-and-Guidance/Legislation/EMTALA/.

In column (1) of Table A.6, we then demonstrate that the poverty rate, which we measure using the Small Area Income and Poverty Estimates (SAIPE), is positively correlated with the fraction of children who are not covered by health insurance. This tells us that a failure to control for the interaction between the poverty rate and the unemployment rate will create a downward bias in the estimate of the coefficient on $Fraction\ Uninsured\ Children \times Unemployment\ Rate$. We therefore control for the interaction between the poverty rate, also measured in 2000,[38] and the unemployment rate, and we create a third instrument as the interaction between the poverty rate and the Bartik instrument. We present results both including and not including this control in Table 1.4.

We find that unemployment has a greater effect on neglect in counties where a larger fraction of children were initially not covered by health insurance, in column (2) of Table 1.4. This effect is significant at the 5% level. For a county with the median poverty rate, a one percentage point increase in the unemployment rate at the 25th percentile of the distribution of the fraction of uninsured children (0.081) leads to a 10 percent increase in neglect, whilst at the 75th percentile (0.156) it leads to a 14 percent increase in neglect. Controlling for the interaction between the poverty rate and the unemployment rate removes a downward bias in the estimate of the coefficient on $Fraction\ Uninsured\ Children \times Unemployment\ Rate$ as predicted, as can be seen by comparing the results in column (2) to those in column (1).

1.5.1.2.2 Having Two Employed Parents

The labour supply of other family members can act as a private safety net for an unemployed person (Cullen & Gruber, 2000). If one parent becomes unemployed in a household without a second employed parent, it is more likely that expenditure will decrease below

[38] We were concerned that the estimates of health insurance coverage and the poverty rate may be related and not truly capture different concepts. However, the models used for the SAIPE and SAHIE are very different. Further, the model used to estimate the poverty rate does not include any measure of health insurance coverage. Details on the methodology can be found at: https://www.census.gov/did/www/saipe/methods/statecounty/2000county.html and https://www.census.gov/did/www/sahie/methods/2000/model.html.

the level required to provide for the children's basic needs. We create a variable that measures the fraction of children in the county with two employed parents living in the household, using data from the Census in 2000. We use the same identification strategy as described in Section 1.5.1.2.1. We interact *Fraction Children Two Employed Parents* with the unemployment rate, and control for the interaction between the poverty rate and the unemployment rate. We create two additional instruments by interacting each of *Fraction Children Two Employed Parents* and the poverty rate with the Bartik instrument.

Whilst the coefficient on the interaction term is negative, it is not statistically significant, as demonstrated in column (4) of Table 1.4. Controlling for the interaction between the poverty rate and the unemployment rate removes an upward bias in the estimate of the coefficient on *Fraction Children Two Employed Parents* $\times$ *Unemployment Rate*, as can be seen by comparing the results in column (4) to those in column (3). That upward bias is to be expected, since the poverty rate is negatively correlated with the fraction of children with two employed parents, as demonstrated in column (2) of Table A.6.

Before concluding that this second private safety net has no impact, we explore the heterogeneity in the effect of unemployment on neglect by the ethnic group of the victim. Across the United States, Black and Hispanic children are significantly less likely to live in a household with two employed parents than White children, as demonstrated in column (1) of Table A.7. Further, Black children are less likely to live with two employed parents than Hispanic children, as demonstrated in column (2) of Table A.7. We separately count incidents of neglect for White, Black and Hispanic children as the dependent variables for the regressions in columns (5) to (7) of Table 1.4. Both the unemployment rate and the Bartik instrument are not ethnic-group specific. We control for the child population of the relevant ethnic group. We weight observations by the child population of the relevant ethnic group so that the estimates are representative of the effect where the children of that ethnic group reside in the U.S.

We find that unemployment has a statistically significant effect on the neglect of Black children in column (6), at the 5% significance level. The effects on White and Hispanic children are positive but not statistically significant. A one percentage point increase in the unemployment rate leads to a 31 percent increase in the neglect of Black children. A one standard deviation increase in the within-county variation in the unemployment rate leads to an increase in the neglect of Black children equivalent to 0.28 standard deviations of its within-county variation (the equivalent effect sizes for Hispanic and White children are 0.19 and 0.08 standard deviations respectively).[39]

In summary, whilst the interaction between the unemployment rate and *Fraction Children Two Employed Parents* is not significant in column (4), we do find evidence that the effect of unemployment is largest for the group of children who are least likely to have two employed parents living in the household with them. However, there may be other explanations for this heterogeneity by ethnic group that are unrelated to the employment status of the parents.

1.5.1.3 Unemployment Reduces Expenditure on Basic Goods

In the absence of safety nets, unemployment will force parents to decrease expenditure. If they decrease expenditure on basic goods, such as food or clothing, this may in turn lead to child neglect. We have a state-level dataset from the Bureau of Economic Analysis (BEA) which measures expenditure per capita on two groups of basic goods that could be related to neglect. The first is expenditure per capita on food and beverages purchased for off-premises consumption, and the second is expenditure per capita on clothing and footwear. We convert both variables to be in real terms.[40] We ask whether unemployment causes a decrease in real expenditure per capita on basic goods, using the following IV regression:

[39] To calculate these standardised effect sizes, we use the within-county variation in the variables. Since we control for county fixed effects, this is the source of identifying variation in the regressions. We take this approach throughout the paper whenever we present standardised effect sizes.

[40] To do so, we use the All Items All Urban Consumers CPI, produced by the BLS.

$$Expenditure_{st} = \beta Unemp_{st} + X'_{st}\rho + \lambda_s + \zeta_t + \epsilon_{st} \tag{1.6}$$

$$Unemp_{st} = \delta(\Sigma_j w_{sj} N_{tj}) + X'_{st}\varrho + \sigma_s + \varphi_t + \upsilon_{st} \tag{1.7}$$

We run separate regressions for expenditure on food and beverages, and clothing and footwear. This IV approach follows the baseline regressions as closely as possible, except that the unit of observation is the state-year. The dependent variable $Expenditure_{st}$ is the natural logarithm of real expenditure per capita on the goods of interest. We create a state-level Bartik instrument, $\Sigma_j w_{sj} N_{tj}$, which is identical to the original Bartik instrument except that the weights, w_{sj}, are at the state level. The validity of the Bartik instrument in state-level regressions relies on the assumption that no individual state dominates national-level production in an industry, since otherwise state-level unobservable worker characteristics may be related to the national-level unemployment rate in that industry. We control for the fractions of the population that are Black, Hispanic, and Other Race in X'_{st}, we control for state and year fixed effects, we weight observations by the child population in the state-year, and we cluster standard errors at the state level. We restrict the sample to the four hundred state-years that are included in the baseline regressions, but the results are very similar if we include all state-years.

We present the results in Table 1.5. We find that unemployment is associated with a decrease in real expenditure per capita on both food and beverages, and clothing and footwear, in the OLS regressions in columns (1) and (3). These associations are significant at the 1% level. In column (2), we find that unemployment causes a decrease in real expenditure per capita on food and beverages, using the IV approach. This effect is also significant at the 1% level. A one percentage point increase in the unemployment rate causes a 1.1 percent decrease in real expenditure on food and beverages per capita. A one standard deviation increase in the unemployment rate leads to a 0.82 standard deviation decrease in real expenditure on food and beverages per capita, using the within-state variation. The effect on clothing and footwear is not statistically significant in the IV regression in column (4), and the point estimate is small.

We therefore conclude that unemployment leads to a decrease in real expenditure on basic goods, which could lead to child neglect. This decrease in expenditure could come from a decrease in the quantity of basic goods purchased, or from the purchase of cheaper goods in the same quantity.

1.5.1.4 Unemployment Causes Repeated Child Neglect

In the absence of safety nets, if unemployment persists, then income may stay persistently below the level required to meet children's basic needs. In that case, unemployment may cause repeated neglect, which we test in Table 1.6. Unemployment indeed persisted during the sample period after the initial shock, as demonstrated in Figure 1.1. Between 2009 and 2012, the unemployment rate fell just 1.06 percentage points.

Thirty-eight states record a unique child ID that takes the same value across the panel.[41] In these states, we can track each child over time and so understand whether each substantiated report is the first time the child has been the victim of neglect, or whether they have been a victim before. We rerun the baseline regressions but using four new dependent variables. We separately count first time, second time, third time and fourth or more time cases of neglect.[42] We present the results in columns (2) to (5) of Table 1.6, after first demonstrating that the main effect still exists in this sub-sample of

[41] Alabama, Minnesota and South Carolina do not record the child ID consistently over time. Arkansas, Indiana, New Mexico, Texas and Wisconsin do not provide mapping files for the Child ID variable and so we cannot verify whether they record the child ID consistently over time. To be cautious we drop these five states too. We cannot track children over time if they move across states, because the child ID variable is only unique within states.

[42] Our ability to observe whether or not a child has previously been the victim of neglect is limited by the start date of the sample period. For example, if a child is neglected for the second time in 2004, having been neglected in 2003 before the start of the sample period, we will count this second report as a first time incident. The more recent the year, the more years of preceding data we have, and so the less likely it is that we miss a previous case of neglect for a given child due to the truncation of the data. For that reason, it is comforting that the distribution of first to fourth or more time cases is quite similar throughout the panel. For example, in the full sample period, 78.2% of cases of neglect are first time, 15.0% second time, 4.3% third time and 2.6% fourth or more times. In 2012, 73.7% are first time, 16.8% second time, 5.6% third time and 4.0% fourth or more times. Nonetheless, we should treat the results in this subsection with a little caution.

states in column (1).

Unemployment has a statistically significant effect on first time, second time and third time cases of neglect at least at the 5% level. The point estimates of the effect of unemployment on second time and third time cases of neglect are, respectively, more than double and more than three times the size of the effect on first time cases. The largest effects of unemployment are on repeat cases of child neglect.

We have interpreted the results in columns (3) and (4) of Table 1.6 as demonstrating that the unemployment shock following the onset of the recession causes repeated child neglect. An alternative explanation is that this unemployment shock only causes one-time child neglect, but that children who were neglected before the recession are neglected again. However, for cases of neglect from 2008 onwards, the median lag to the previous case of neglect is just 411 days. This relatively short lag between reports suggests that, indeed, children have often been the victim of multiple cases of neglect after the unemployment shock.

1.5.2 Alternative Explanations

1.5.2.1 Substance Abuse

Individuals may increase alcohol consumption or drug use to cope with stress after being made unemployed (e.g., Boardman et al. (2001); Eliason & Storrie (2009)). Substance abuse can in turn increase child maltreatment. Previous evidence has demonstrated that substance abuse can lead to violent behaviour within the household (e.g., Lee Luca et al. (2015)). Markowitz & Grossman (1998) find that alcohol regulations reduce violence against children. Further, alcohol and drug use can drain resources that could otherwise be used to pay for the children's basic needs, which can lead to neglect. Alternatively, this effect on neglect may be non-pecuniary. Alcohol or drug use might directly incapacitate the parent, leading to neglect because they are unable to adequately supervise their children.[43]

To explore this mechanism, we use data on the estimated county-level prevalence of heavy and binge drinking over the period from 2004 to 2012.[44] These measures are estimated in a paper by Dwyer-Lindgren et al. (2015), using the Behavioural Risk Factor Surveillance System (BRFSS) dataset. The BRFSS is a health-related telephone survey of U.S. households, which samples approximately 400,000 adults each year and is the largest continuously conducted health survey system in the world.[45]

In Table 1.7, we test whether unemployment increases overall abuse or neglect by increasing alcohol consumption. In columns (1) to (6), we ask whether unemployment increases heavy or binge drinking. We look at the prevalence rate of heavy and binge drinking at the county level, overall and for men and women separately. We use the same Bartik IV strategy as in the baseline regressions. We weight the observations by the child population in the county-year as in the baseline regressions because we want to estimate an effect of unemployment on alcohol consumption that is representative of the counties where children reside in the U.S.

We in fact find that unemployment causes a decrease in female heavy and binge drinking. These effects are significant at the 5% and 1% levels respectively. A one standard deviation increase in the unemployment rate leads to a 0.39 standard deviation decrease in the prevalence of female heavy drinking and a 0.59 standard deviation decrease in the prevalence of female binge drinking, using the within-county variation. This result is in line with previous studies where economic downturns have been found to be positively correlated with health, as individuals can no longer afford costly unhealthy behaviours (Ruhm, 2000, 2003, 2005; Ruhm & Black, 2002). The coefficients on the unemployment

[43] Whilst the main explanation for the effect of unemployment on neglect in this paper is pecuniary, it is possible that unemployment has non-pecuniary effects that lead to neglect. Substance abuse is one possibility, and mental health is another. Unemployment can lead to anxiety or depression (Tefft, 2011), which can also impair a parent's ability to provide basic care or supervision to their children. Unfortunately, we do not have the data to explore mental health as a possible mechanism.

[44] Heavy drinking is classified as more than one drink per day for women, or more than two drinks per day for men. Binge drinking is classified as having more than four drinks for women or five drinks for men on a single day at least once in the previous thirty days.

[45] For details see: http://www.cdc.gov/brfss/.

rate for overall heavy and binge drinking are also negative, though not statistically significant. When we control for the prevalence of overall binge and heavy drinking in the baseline regressions in columns (7) and (8), there is virtually no change in the size or statistical significance of the effect of unemployment on overall abuse or neglect.

We can therefore rule out that unemployment causes an increase in overall abuse or neglect through an increase in alcohol consumption. Combined with the finding that the point estimate for the effect of unemployment on neglect is smaller for White children than Black or Hispanic children in Section 1.5.1.2.2, this also suggests that the results in this paper are not related to the rising substance abuse and morbidity of middle-aged White Americans documented in Case & Deaton (2015).

1.5.2.2 Divorce

Unemployment can lead to divorce (e.g., Charles & Stephens (2004); Doiron & Mendolia (2012); Eliason (2012)). Divorce can in turn increase child maltreatment, for at least three reasons (Lindo et al., 2013). Divorce might increase the time that children spend with unrelated adults (the new partners of their parents), who may be particularly prone to abusive behaviour (Sedlak et al., 2010). The children of divorced parents may grow up in single parent households, which may have fewer resources to provide for children's basic needs, leading to an increase in neglect. Finally, divorce may lead to stress or substance abuse, which may lead to abuse or neglect.

In Table 1.8, we test whether divorce can explain the effect of unemployment on overall abuse or neglect. We use the American Community Survey, available for the period from 2005 to 2012, to measure the prevalence of divorce at a county level. As in the regressions for alcohol consumption, we weight the observations by the child population in the county-year.

In column (1), we find that there is a large, positive and statistically significant effect of unemployment on divorce. A one percentage point increase in the unemployment rate increases the divorce rate by 0.6 percentage points. A one standard deviation increase in the unemployment rate leads to a 0.99 standard deviation increase in the rate of divorce, using the within-county variation. We allow for an effect at a one year lag in column (2). Unemployment might trigger divorce with a lag because divorce procedures can be time consuming. As we might expect, the effect is greater at a one year lag than the contemporaneous effect.

In columns (3) and (4), we control for the divorce rate in the baseline regression. If divorce is driving the effect of unemployment on overall abuse or neglect, we would expect the coefficient of interest on unemployment to fall to zero and become statistically insignificant, whilst the coefficient on divorce would be positive and significant. However, controlling for divorce has virtually no effect on the size or statistical significance of the effect of unemployment on overall abuse or neglect. Further, the coefficients on the divorce rate are not statistically significant and are even negative for both types of maltreatment. This suggests that although unemployment does cause an increase in divorce, the effect of unemployment on overall abuse and neglect is not driven by divorce.

1.6 Robustness

1.6.1 An Effect on Actual Incidence or Reporting?

The main concern with the results is whether they capture an effect of unemployment on the reporting of maltreatment, rather than the actual incidence. We test this in three ways in Table 1.9. First, an unemployment shock may result in a reallocation of labour to high-reporting sectors, such as schools, social services, health care, police, clergy and childcare, which could increase reporting rates for a given level of actual maltreatment. To deal with this, we follow Lindo et al. (2013). In columns (1) and (2), we control for the fraction of the working-age population employed in these six high-reporting sectors. We find that the coefficient on the unemployment rate barely changes in either the regression for overall abuse or neglect, and the fraction employed in high-reporting sectors variables are statistically insignificant in all but one case.

Second, unemployed individuals spend more time at home and may therefore have more opportunity to observe the maltreatment of a neighbour, friend or extended family member's children, for a given level of actual maltreatment. The NCANDS dataset contains a variable that records the identity of the reporter for each report. We therefore create a new dependent variable which does not count incidents where the reporter is a friend, neighbour or other relative.[46] We only count reports made by groups of reporters who are not affected by this concern. These are: professionals (social services, medical, mental health, legal, law enforcement, criminal justice and education personnel, child day care providers and substitute care providers), parents and the alleged victim or perpetrator. By definition, professionals only interact with the maltreated children through their job, and parents are sufficiently close to their own children that they have an opportunity to observe maltreatment whether they are employed or unemployed. In columns (3) and (4), we rerun the baseline regressions using this dependent variable. We restrict the sample to state-years where the reporter's identity is non-missing for more than 80% of substantiated reports. Again, we find that the coefficient on the unemployment rate barely changes in size or statistical significance in either regression.

Third, we have so far only considered substantiated reports, which comprise 19 percent of all reports. If unemployment affects reporting, then we should estimate a very similar effect on unsubstantiated reports. If unemployment affects the actual incidence of maltreatment, then we should only see an effect on substantiated reports. In columns (5) and (6), we run the baseline regressions but now count only unsubstantiated reports in the dependent variable. There is no significant effect of unemployment on unsubstantiated reports of either overall abuse or neglect, and the point estimates are approximately one third the size of the estimates for substantiated reports. In summary, all three tests demonstrate that unemployment affects the actual incidence of overall abuse and neglect and not reporting behaviour.

[46]These groups of reporters account for 10.7% of all reports with a non-missing reporter in the state-years included in these regressions. We also do not count reports made by 'other' or 'anonymous reporter', as we cannot know whether they are also affected by this concern.

1.6.2 Alternative Choices of the Instrument

The OLS estimates in the main results in Table 1.2 are smaller than the IV estimates and not statistically significant for both overall abuse and neglect. Whilst there are two very plausible explanations for these differences, as described in Section 1.4.1, we still wish to check that the IV estimates are robust to alternative choices of the instrument. We consider four alternative instruments in Table 1.10.

In columns (1) and (2), we instrument using the change in the real exchange rate (RER) from the previous year, calculated as $(RER_t - RER_{t-1})/RER_{t-1}$, multiplied by the county-level fraction of the employed working-age population in manufacturing at the start of the sample period, in 2003. An appreciation of the real exchange rate harms the competitiveness of U.S. manufacturers who compete with foreign producers, thereby increasing unemployment. This approach follows Lin (2008). To measure the real exchange rate, we use the broad real trade-weighted US dollar index, produced by the Federal Reserve Bank of St. Louis.

In columns (3) to (8), we use three alternative Bartik instruments. For the first two, we use alternative proxies of the economic performance of each industry at the national level. We replace the national-level unemployment rate by industry from the original Bartik instrument with the national-level real value added by industry (measured in billions of 2009 dollars) in columns (3) and (4), and with the national-level total employed by industry in columns (5) and (6). Both measures are produced by the BEA.[47] Finally, in columns (7) and (8), we replace the weights in the original Bartik instrument with the fraction of the employed working-age population in each industry in the county in 1995.

The effect of unemployment on neglect is robust across these alternative choices of the instrument. In every case, the effect is still statistically significant at least at the 10% level, and the point estimate ranges from 0.15 to 0.24. The instruments are highly relevant,

[47]We use the total employed rather than the employment rate by industry, because the CPS (used to create the unemployment rate in the original Bartik instrument) does not ask about the industry of individuals who are out of the labour force. For more details see the Data Appendix.

with Kleibergen-Paap F-statistics ranging from 12.9 to 41.1, and the coefficients on the instrument in the first stage always have the expected sign (positive for the regressions in columns (1), (2), (7) and (8), and negative in columns (3) to (6)). The effect on overall abuse is always positive, but is smaller in size and not always statistically significant.

1.6.3 Other Robustness Tests

1.6.3.1 Alternative Choices of the Dependent Variable

We check the robustness of the results to five alternative choices for the dependent variable in Table A.8. In columns (1) and (2), we use the natural logarithm of the rate of maltreatment per 100,000 children, and accordingly no longer control for the natural logarithm of the child population on the right hand side.[48] This is a restricted version of the baseline regression in which the coefficient on the natural logarithm of the child population equals one. In columns (3) to (8), we use three alternative methods of dealing with zeros. In columns (3) and (4), we take the inverse hyperbolic sine transformation of the rate of maltreatment per 100,000 children, following Woolley (2011). In columns (5) and (6), we add 0.01 to the number of incidents before taking the natural logarithm, and in columns (7) and (8), we add 0.0001. Finally, in columns (9) and (10), we count the number of children who are victims of maltreatment in each county-year, rather than the number of incidents. In those regressions, we must restrict the sample to the states that report a unique child ID.

The results are again robust. The effect of unemployment on neglect is statistically significant at the 1% level in every case, and the size of the point estimate is very similar to the main results. The effect on overall abuse is statistically significant at least at the 10% level in all but one case (in column (9), the p-value is 0.113), and again the point estimates are similar in size to the main results.

[48] To do this, we add 0.001 to the number of incidents before calculating the rate.

1.6.3.2 Other Changes to the Specification

We make several other changes to the specification, and present the results in Table A.9. In columns (1) and (2), we control for linear county trends. In columns (3) and (4), we drop the state-year fixed effects. In columns (5) and (6), we run reduced form OLS regressions with the Bartik instrument as the right hand side variable. In columns (7) and (8), we weight observations using the child population in the county at the start of the sample period, in 2003. In columns (9) and (10), we weight observations by the total population of all ages in the county-year. Finally, in columns (11) and (12), we cluster standard errors at the county level. The results are robust. The effect on neglect is statistically significant at least at the 5% level in every case, whilst the effect on overall abuse is significant at least at the 10% level in all but two cases.

1.6.3.3 Changes to the Sample

As a final set of robustness tests, we check that the results are not driven by outlier counties with very small or large populations of children. In columns (1) and (2) of Table A.10, we drop the counties with a child population of more than one million at the start of the sample period, in 2003.[49] In columns (3) and (4), we drop the smallest 10 percent of counties in terms of their child population at the start of the sample period, in 2003.[50] In both cases, the results for overall abuse and neglect are robust, both in terms of the size and statistical significance of the effect.

1.7 Conclusion

Child maltreatment is a severe public health problem with long-term consequences for the victims. Despite this, little causal empirical evidence on its determinants currently exists.

[49] There are two counties with a child population of more than one million in 2003. These are: Los Angeles, California (2,678,788 children), and Harris, Texas (1,043,580 children).

[50] We drop the counties with a population of less than 1,231 children in 2003 (the 10th percentile among all counties in the U.S. in that year).

In this paper, we address that gap by estimating the causal effect of unemployment on child maltreatment. We use a new dataset containing every reported incident of child abuse and neglect made to the Child Protective Services for nearly every state in the U.S. from 2004 to 2012. To identify the effect of unemployment, we use a Bartik instrument. We instrument for the county-level unemployment rate with the weighted average of the national-level unemployment rates across each of twenty industries, where the weights are the county-level fraction of the employed working-age population in each industry the year before the start of the sample period, in 2003. We find that a one percentage point increase in the unemployment rate leads to a 20 percent increase in neglect. In the county with the median prevalence of neglect, which has 536 cases per year, this is equivalent to an additional 110 cases per year. We show that our results are robust and capture an effect on the actual incidence of neglect and not reporting behaviour.

One important mechanism behind the effect of unemployment on neglect is that parents lack safety nets. We find evidence that both social and private safety nets mitigate the effect of unemployment on neglect. The effect of unemployment is smaller in states that introduce longer extensions to the duration of unemployment benefits. The effect is greater in counties where an initially larger fraction of children are not covered by health insurance. We also find evidence that unemployment causes a decrease in real expenditure on basic goods and leads to repeated neglect, as we might expect in the absence of safety nets.

These results suggest that a potentially important benefit of the social safety net has been ignored in policy evaluations in the academic literature. The existing literature weighs up a trade-off between the successful poverty-reducing impact of the social safety net (e.g., Ben-Shalom et al. (2011); Bitler et al. (2017)), and the possibility that safety nets have harmful effects on incentives, for example creating a disincentive to engage in job search (e.g., Card & Levine (2000); Barro (2010); Rothstein (2011)). In terms of the former, our paper quantifies an impact of the social safety net that has not previously been considered. Unemployment benefits can prevent children from being neglected when a parent is made unemployed, which may in turn have positive long-term consequences for those children.

These results also suggest that we need to understand whether parents face barriers to creating private safety nets. Private safety nets can prevent children from being neglected when a parent is made unemployed, and yet many households lack even very low-levels of private safety nets. For example, in 2017, thirty-seven percent of U.S. households are liquid asset poor, meaning that they do not have enough in liquid assets to replace income at the poverty level for three months in the event of a job loss. Even if they sold non-liquid assets such as a home or a car, more than one quarter of U.S. households could not replace this level of income (Wiedrich et al., 2017).

1.8 Tables and Figures

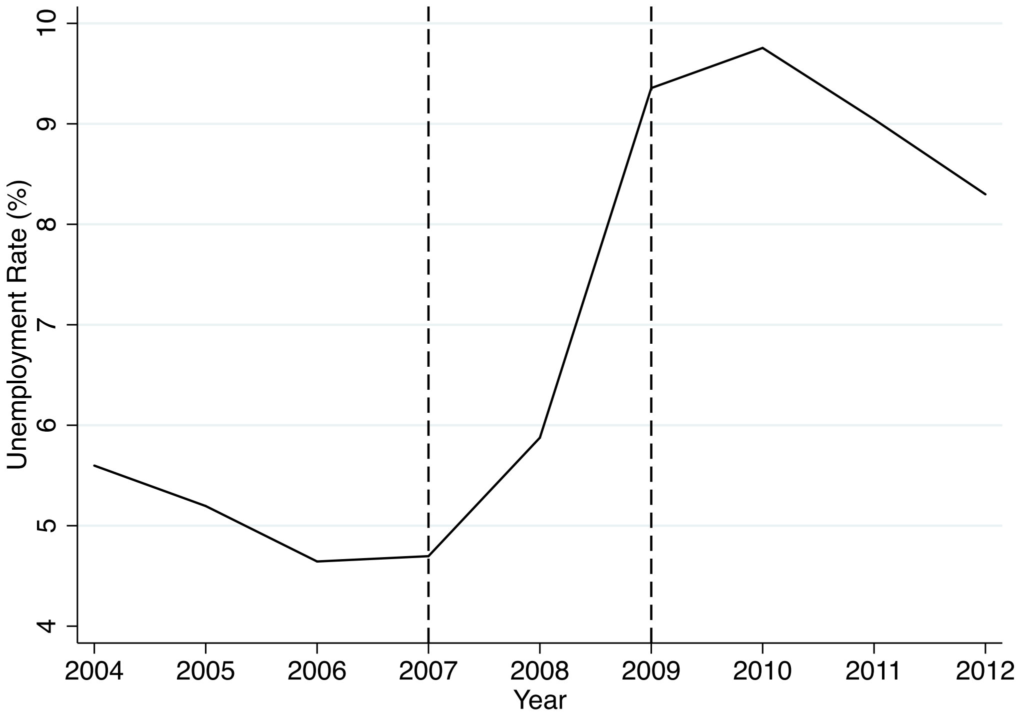


Figure 1.1: *Notes.* **Trend in the Unemployment Rate**. In this Figure, we present the trend in the weighted average unemployment rate across the 2,803 counties in our final sample. We weight observations by the child population, so that the unemployment rate is representative of where children reside in the U.S. Unemployment rates jumped during the period from 2007 to 2009, with the onset of the recession.

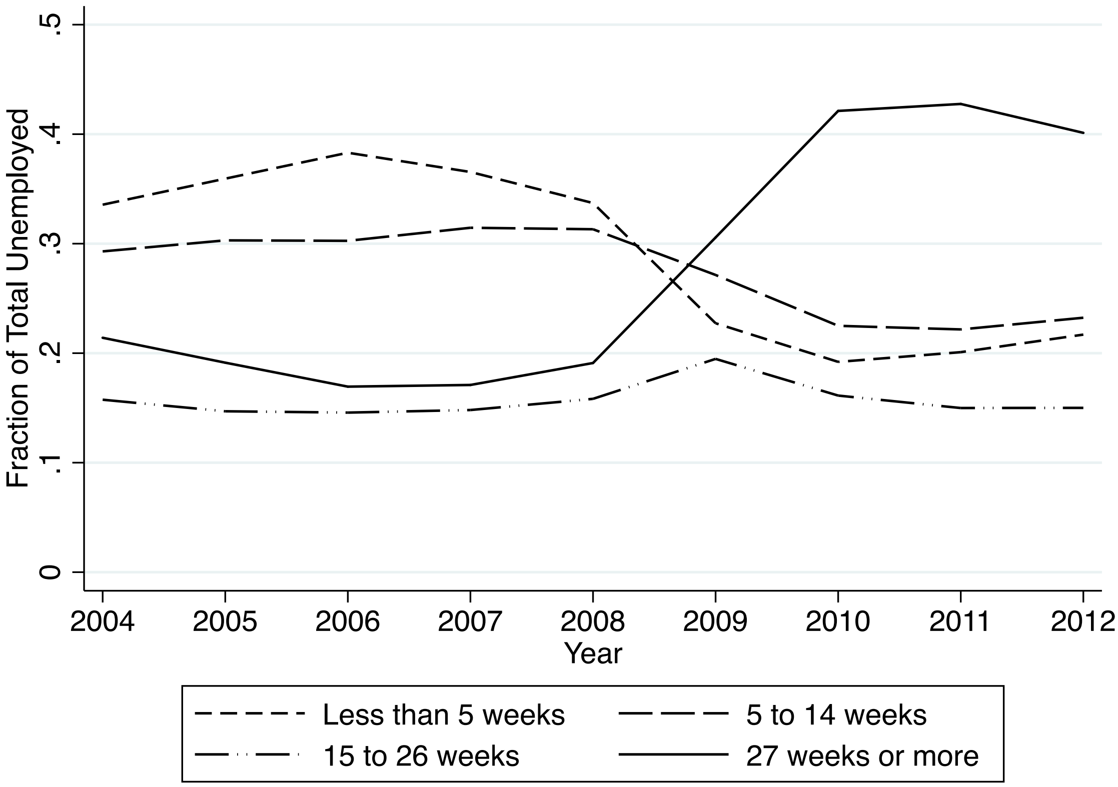


Figure 1.2: *Notes.* **The Increase in Long-Term Unemployment as a Proportion of Total Unemployment**. In this Figure, we plot the fraction of the total unemployed in each of four duration categories: less than five weeks, five to fourteen weeks, fifteen to twenty-six weeks, and twenty-seven or more weeks. Long-term unemployment is defined as twenty-seven or more weeks. We take a weighted average of the fraction in each category across state-years, where the weights are the number of children in the state-year. We include only the 400 state-years in our final sample.

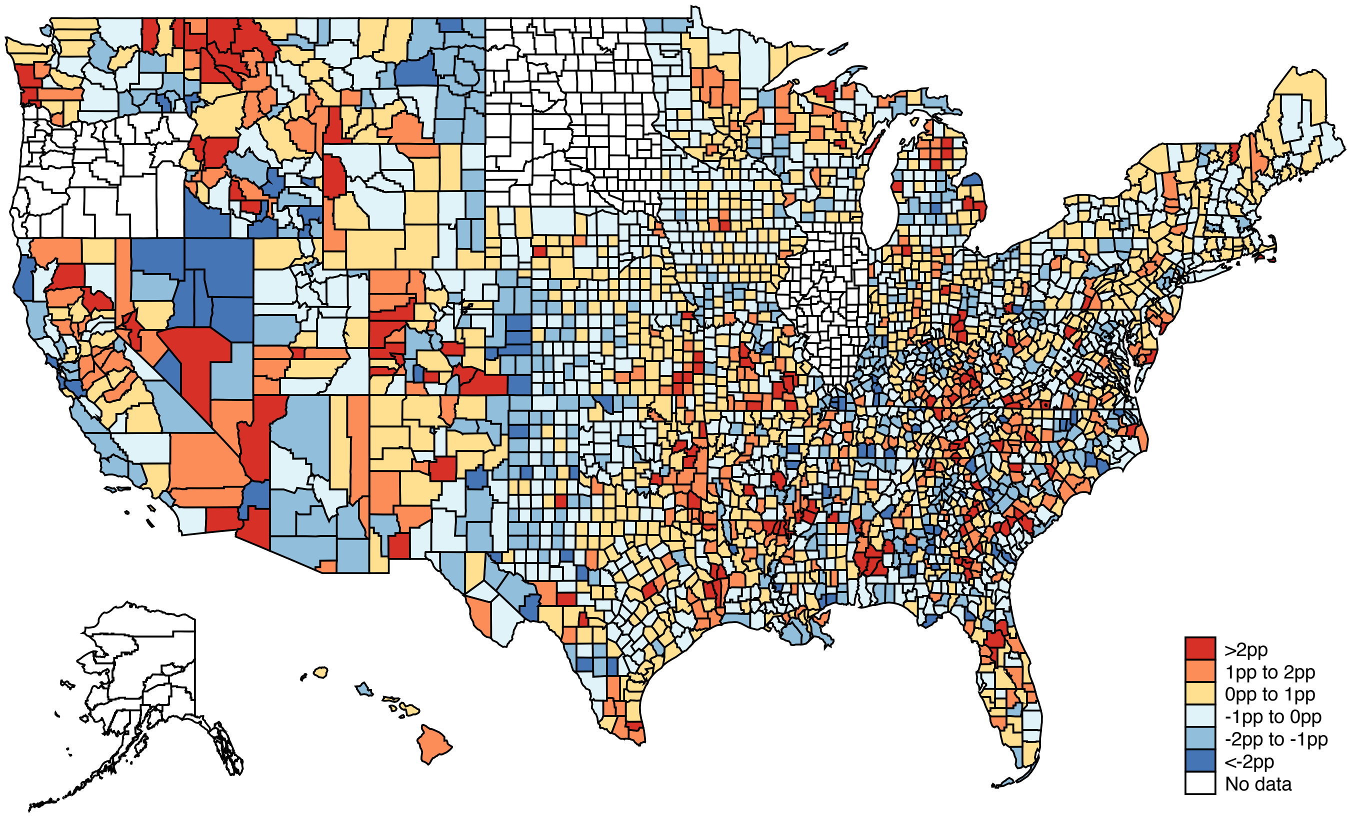


Figure 1.3: *Notes.* **Change in the Unemployment Rate Before and After the Recession**. This map presents the change in the unemployment rate before and after the recession. We regress the unemployment rate on state-year fixed effects. We take the residuals from this regression, and calculate the average for the pre-recession period 2004-2006 and for the post-recession period 2010-2012. We then subtract the pre-recession average from the post-recession average. Alaska and Hawaii are in the bottom left of the figure.

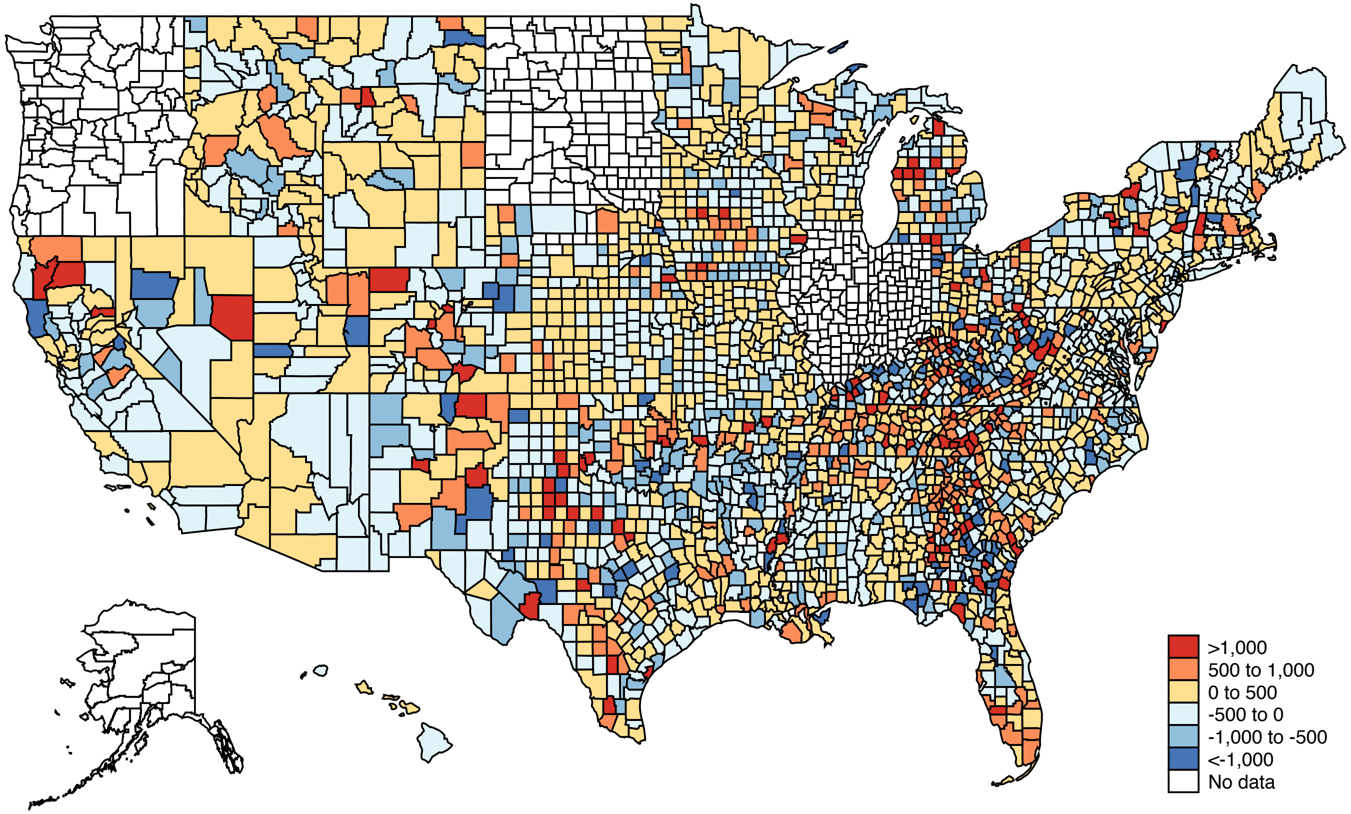


Figure 1.4: *Notes.* **Change in the Overall Abuse Rate Before and After the Recession**. This map presents the change in the overall abuse rate before and after the recession. We regress the overall abuse rate per 100,000 children on state-year fixed effects. We take the residuals from this regression, and calculate the average for the pre-recession period 2004-2006 and for the post-recession period 2010-2012. We then subtract the pre-recession average from the post-recession average. Alaska and Hawaii are in the bottom left of the figure. Washington does not record emotional abuse in any year, and Indiana does not record emotional abuse for the years 2004-7. Therefore, we cannot calculate the change in the overall abuse rate for either state and so they are missing in this map.

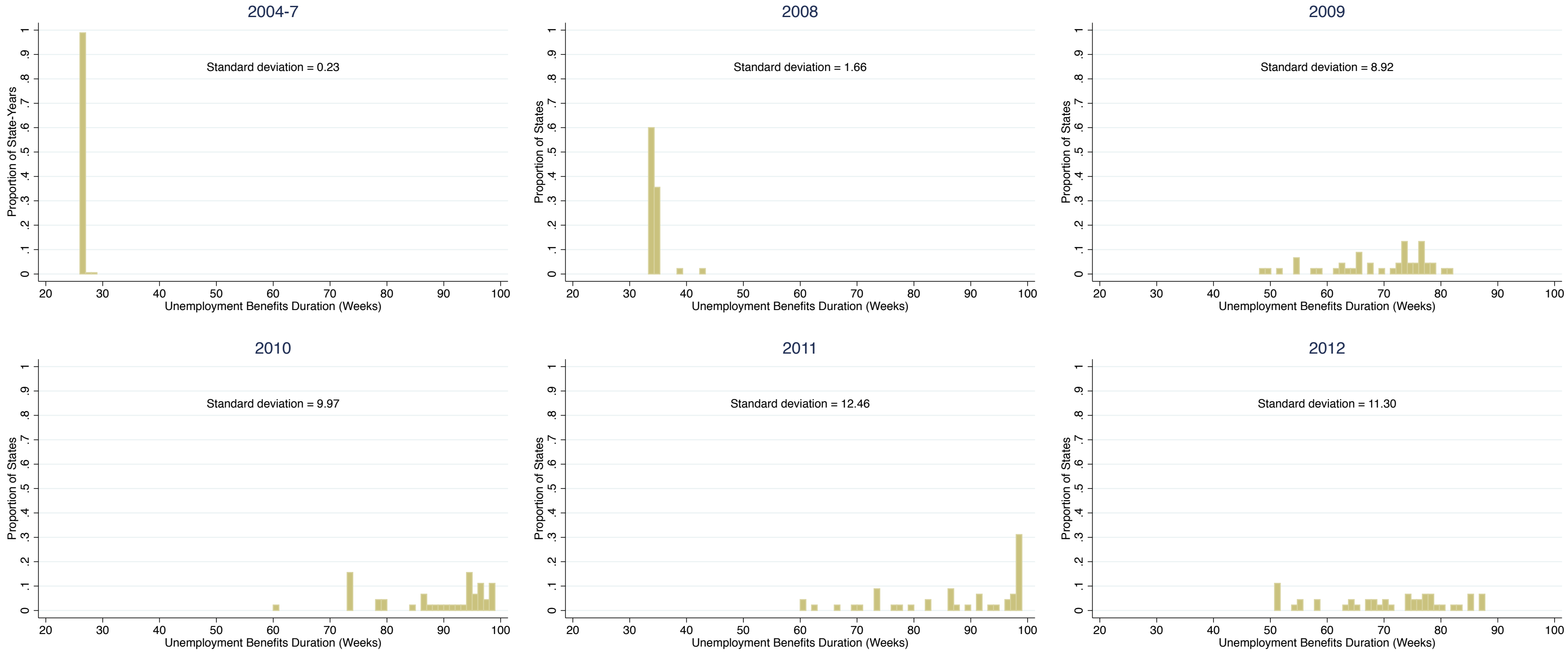


Figure 1.5: *Notes.* **The Duration of Unemployment Benefits Over Time.** In this Figure, we plot the duration of unemployment benefits over time. We choose bins with a width of one week (our annual dataset is an average across quarterly data and so we have non-integer values). We plot the proportion of states whose duration of unemployment benefits falls into each bin. The top left panel presents the distribution before the onset of the recession, over the period 2004-7. Nearly every state had 26 weeks of unemployment benefits. The remaining five panels show how that distribution changed in each year from 2008 onwards. States increased the duration of unemployment benefits, and no state provided fewer than 33 weeks over the period 2008-12. In each year, we include all 45 states that are included in the regressions which investigate the effect of unemployment benefits, in Table 1.3.

Table 1.1: Summary Statistics

	Unweighted Mean	Weighted Mean
Number of Incidents of Physical Abuse	42.22	407.05
	(135.39)	(704.90)
Number of Incidents of Sexual Abuse	22.60	204.92
	(67.80)	(394.51)
Number of Incidents of Emotional Abuse	19.68	391.86
	(176.13)	(1499.67)
Number of Incidents of Neglect	182.51	1806.71
	(616.47)	(3171.64)
Unemployment Rate (%)	6.85	6.91
	(2.99)	(2.83)
Fraction Black (%)	9.47	12.28
	(14.75)	(12.79)
Fraction Hispanic (%)	8.42	16.95
	(13.59)	(17.61)
Fraction Other Race (%)	3.63	6.99
	(5.90)	(7.38)
Child Population	24,386	
	(79,772)	
Observations	24,181	24,181
Counties	2,803	2,803
States	46	46

Notes. In this Table, we present summary statistics for the full sample of 24,181 county-years included in the baseline regressions. We present means, with standard deviations in parentheses. The statistics in column (1) are unweighted, whilst those in column (2) are weighted by the child population in the county-year.

Table 1.2: Main Results

	Overall		Physical		Neglect		Sexual		Emotional	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV	OLS	IV
Unemployment Rate	0.013	**0.096****	**0.026****	0.060	0.0020	**0.20*****	0.010	0.0037	-0.045	0.032
	(0.010)	**(0.049)**	**(0.0097)**	(0.060)	(0.012)	**(0.068)**	(0.014)	(0.062)	(0.044)	(0.24)
State-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,468	23,468	24,181	24,181	24,181	24,181	24,181	24,181	23,468	23,468
Counties	2,771	2,771	2,803	2,803	2,803	2,803	2,803	2,803	2,771	2,771
States	45	45	46	46	46	46	46	46	45	45
Mean of outcome	3.77	3.77	1.65	1.65	3.05	3.05	0.94	0.94	-2.19	-2.19
Mean of Unemployment Rate	6.86	6.86	6.85	6.85	6.85	6.85	6.85	6.85	6.86	6.86
Kleibergen-Paap F-stat		34.5		34.9		34.9		34.9		34.5

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. This Table contains the results of the OLS and IV regressions which look at the effect of unemployment on the incidence of overall, physical, sexual, emotional abuse and neglect. In each case, the dependent variable is the natural logarithm of the number of incidents of that abuse type per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of abuse across states. We include state-year fixed effects to control for any changes in the definitions of abuse over time within states. Finally, we control for the natural logarithm of the child population and the fraction of the overall population that is Black, Hispanic or Other Race (which includes individuals who are Asian, Alaskan Native, American Indian, Native Hawaiian or Other Pacific Islander and Two or More Races). We weight observations by the child population in the county-year. Emotional abuse is not observed in some state-years, which is why there are fewer observations in the regressions in which overall abuse and emotional abuse are the dependent variables.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3: Unemployment Benefits Mitigate the Effect of Unemployment on Neglect

	Main Sample			Contiguous Border Pairs Sample		
	First Stage		Second Stage	Second Stage		
	Unemployment Rate	Interaction	Neglect			
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
Bartik IV	1.35***	153.2***				
	(0.26)	(37.1)				
Benefit Duration x Bartik IV	-0.0073***	-1.18***				
	(0.0023)	(0.40)				
Unemployment Rate			**0.32*** **	**0.18* **	**0.26* **	0.18
			(0.12)	**(0.10)**	**(0.14)**	(0.18)
Benefit Duration x Unemployment Rate			**-0.0021* **		**-0.0025**	**-0.0030**
			(0.0012)		**(0.0016)**	**(0.0025)**
State-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
County-Pair-Year Fixed Effects	No	No	No	No	No	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,145	24,145	24,145	17,915	17,915	17,915
Counties	2,799	2,799	2,799	987	987	987
States	45	45	45	45	45	45
Mean of outcome			3.05	3.06	3.06	3.06
Mean of Unemployment Rate			6.85	6.80	6.80	6.80
Kleibergen-Paap F-stat				11.4		
SW F-stat for Unemployment Rate			32.9		47.5	50.0
SW F-stat for Unemployment Rate x Benefit Duration			6.43		9.76	14.6

Notes. Standard errors (in parentheses) are clustered at the state level in columns (1)-(3). In columns (4)-(6), we use two-way clustering of the standard errors at the level of the state and border segment. This Table contains the results of IV regressions in which we ask whether the effect of unemployment on neglect is mitigated by unemployment benefits. In columns (1)-(3), we present the results from a regression in which we interact the unemployment rate with the duration of unemployment benefits that an individual can claim in the state-year. In columns (1) and (2), we present results from the first stages, and in column (3), the results from the second stage. In columns (4) and (5), we restrict the sample to contiguous county-pairs that straddle a state boundary. In these columns, we check whether the effect of unemployment on neglect still exists in this sub-sample of county-pairs (column (4)), and whether the heterogeneity by benefits duration exists (column (5)). In column (6), we use the contiguous border counties methodology to ask whether extending the duration of benefits causes a reduction in the effect of unemployment on neglect. Data on the duration of benefits is not available for Hawaii. In each case, the dependent variable is the natural logarithm of the number of incidents of neglect per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of neglect across states. We include state-year fixed effects to control for any changes in the definitions of neglect over time within states. Finally, we control for the natural logarithm of the child population and the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year. In column (6), we additionally control for county-pair-year fixed effects.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.4: Private Safety Nets Mitigate the Effect of Unemployment on Neglect

	Heterogeneity by Private Safety Net Coverage				Heterogeneity by Ethnic Group		
	Health Insurance Coverage		Two Employed Parents		White	Black	Hispanic
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	IV	IV	IV	IV	IV	IV	IV
Unemployment Rate	**0.20*****	**0.14****	0.072	**0.19****	0.065	**0.31****	0.22
	(0.066)	**(0.058)**	(0.060)	**(0.091)**	(0.056)	**(0.16)**	(0.15)
Fraction Uninsured Children x Unemployment Rate	-0.17	**0.52****					
	(0.15)	**(0.21)**					
Poverty Rate x Unemployment Rate		**-0.64*****		**-0.38****			
		(0.19)		**(0.16)**			
Fraction Children Two Employed Parents x Unemployment Rate			0.16*	-0.036			
			(0.094)	(0.13)			
State-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes	No	No	No
ln(Ethnic-Group Specific Child Population)	No	No	No	No	Yes	Yes	Yes
Observations	24,181	24,172	24,181	24,172	23,578	23,461	23,565
Counties	2,803	2,802	2,803	2,802	2,736	2,735	2,736
States	46	46	46	46	45	45	45
Mean of outcome	3.05	3.05	3.05	3.05	2.29	-1.18	-1.58
Mean of Unemployment Rate	6.85	6.85	6.85	6.85	6.85	6.86	6.85
Kleibergen-Paap F-stat					52.2	13.7	16.2
SW F-stat for Unemployment Rate	50.5	85.2	56.7	103.6			
SW F-stat for Unemployment Rate x Private Safety Net	101.5	189.9	221.0	269.7			
SW F-stat for Unemployment Rate x Poverty Rate		235.1		296.6			

Notes. Standard errors (in parentheses) are clustered at the state level. In this Table, we ask whether private safety nets mitigate the effect of unemployment on neglect. In columns (1) and (2), we consider the fraction of children who are not covered by health insurance in 2000. In columns (3) and (4), we consider the fraction of children who have two employed parents present in the household with them in 2000. In columns (1) and (3), we interact the unemployment rate with the measure of the private safety net, and interact the Bartik instrument with the measure of the private safety net to create a second instrument. In columns (2) and (4), we additionally control for the interaction between the unemployment rate and the poverty rate in 2000, to ensure the private safety net does not proxy for the interaction with poverty. To do so, we add a third instrument, which is the interaction between the Bartik instrument and the poverty rate in 2000. The poverty rate is measured as a fraction of the population. In columns (5)-(7), we separately count incidents of neglect for White-not-Hispanic, Black-not-Hispanic and Hispanic children for the dependent variables. In each case, the dependent variable is the natural logarithm of the number of incidents of neglect per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of neglect across states. We include state-year fixed effects to control for any changes in the definitions of neglect over time within states. In columns (1)-(4), we control for the natural logarithm of the child population. In columns (5)-(7) we control for the natural logarithm of the ethnic-group specific child population. Finally, we control for the fraction of the overall population that is Black, Hispanic or Other Race. In columns (1)-(4), we weight observations by the child population in the county-year. In columns (5)-(7) we weight observations by the child population of the relevant ethnic group in the county-year. Both the county level unemployment rate and the instrument are not ethnic group specific in any regression. There are fewer observations in the regressions for Black and Hispanic children than White, as some county-years had a total population of zero Black or Hispanic children. The ethnic group of the victim is not observed for Pennsylvania.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: Unemployment Decreases Real Expenditure on Basic Goods

	Food and Beverages		Clothing and Footwear	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
Unemployment Rate	**-0.0095*****	**-0.011*****	**-0.014*****	-0.00040
	(0.0014)	**(0.0031)**	**(0.0045)**	(0.0090)
Year Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes
Observations	400	400	400	400
States	46	46	46	46
Mean of outcome	7.92	7.92	6.97	6.97
Mean of Unemployment Rate	6.35	6.35	6.35	6.35
Kleibergen-Paap F-stat		18.5		18.5

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. In this Table, we ask whether unemployment decreases real expenditure per capita on basic goods that could in turn lead to child neglect. The unit of observation is the state-year. In columns (1) and (2), the dependent variable is the natural logarithm of real expenditure per capita on food and beverages purchased for off-premises consumption (measured in 2012 dollars). In columns (3) and (4), the dependent variable is the natural logarithm of real expenditure on clothing and footwear (measured in 2012 dollars). For the IV regressions, we create a state-level Bartik IV. The specification follows the baseline as closely as possible. We control for state fixed effects, year fixed effects and the fraction of the population that is Black, Hispanic or Other Race. We weight observations by the child population in the state-year. We restrict the sample to the 400 state-years that are included in the baseline regressions.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.6: Unemployment Causes Repeated Neglect

		First Time	Second Time	Third Time	Fourth or More Time
	(1)	(2)	(3)	(4)	(5)
	IV	IV	IV	IV	IV
Unemployment Rate	**0.26*****	**0.23****	**0.52*****	**0.77*****	0.30
	(0.094)	**(0.094)**	**(0.18)**	**(0.26)**	(0.23)
State-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes	Yes
Observations	17,748	17,748	17,748	17,748	17,748
Counties	2,078	2,078	2,078	2,078	2,078
States	38	38	38	38	38
Mean of outcome	2.93	2.70	-0.29	-2.96	-4.69
Mean of Unemployment Rate	6.86	6.86	6.86	6.86	6.86
Kleibergen-Paap F-stat	26.7	26.7	26.7	26.7	26.7

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. This Table contains the results of IV regressions which ask whether unemployment causes repeated neglect. We separately count first time cases, second time cases, third time cases and fourth or more time cases of neglect, using the Child ID variable in the NCANDS dataset. The dependent variable in column (2) is the natural logarithm of the number of incidents of neglect per year, counting only first time cases, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. In column (3), the dependent variable counts only second time cases, in column (4), third time cases, and in column (5), fourth or more time cases. We can only include states that we can verify consistently record the Child ID variable over time, which means we must drop observations from Alabama, Arkansas, Indiana, Minnesota, New Mexico, South Carolina, Texas and Wisconsin. In column (1), we therefore run the baseline regressions on the sub-sample of 38 states who record the Child ID consistently over time, to check if the main result still exists. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of neglect across states. We include state-year fixed effects to control for any changes in the definitions of neglect over time within states. Finally, we control for the natural logarithm of the child population and the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: Alcohol Consumption Does Not Drive the Effects

	Alcohol Consumption						Abuse and Neglect	
	Both		Male		Female		Overall	Neglect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Heavy	Binge	Heavy	Binge	Heavy	Binge		
	IV	IV	IV	IV	IV	IV	IV	IV
Unemployment Rate	-0.063	-0.10	-0.0091	0.042	**-0.12****	**-0.24*****	**0.093***	**0.19*****
	(0.051)	(0.068)	(0.068)	(0.12)	**(0.051)**	**(0.061)**	**(0.048)**	**(0.065)**
Heavy Drinking Prevalence							0.016	0.061
							(0.040)	(0.047)
Binge Drinking Prevalence							-0.044	-0.032
							(0.038)	(0.042)
State-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ln(Child Population)	No	No	No	No	No	No	Yes	Yes
Observations	21,370	24,172	21,370	24,172	21,370	24,172	20,782	21,370
Counties	2,802	2,802	2,802	2,802	2,802	2,802	2,770	2,802
States	45	45	45	45	45	45	44	45
Mean of outcome	7.20	16.6	9.71	23.5	4.77	9.99	3.75	3.04
Mean of Unemployment Rate	7.00	6.85	7.00	6.85	7.00	6.85	7.02	7.00
Kleibergen-Paap F-stat	29.8	33.1	29.8	33.1	29.8	33.1	27.0	27.7

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. In this Table, we test the alcohol mechanism. In columns (1) to (6), we use the same Bartik IV strategy to look at whether unemployment causes an increase in alcohol consumption. In columns (1), (3) and (5) we look at the prevalence of heavy drinking, which is defined as consuming more than one drink per day for women and two drinks per day for men for the last thirty days. In columns (2), (4) and (6), we look at the prevalence of binge drinking, which is defined as consuming more than four drinks in a single day for women or five drinks for men, at least once during the past thirty days. In columns (1) and (2) we look at the prevalence rate across genders, in columns (3) and (4) we look at the male prevalence and in columns (5) and (6) female prevalence. In columns (7) and (8), we then control for the prevalence of heavy and binge drinking across genders in the baseline regression. There are no measures of alcohol for D.C. (and so we lose one state, with one county), and heavy drinking is only measured from 2005 onwards. In columns (7) and (8), the dependent variable is the natural logarithm of the number of incidents of that abuse type per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We control for county fixed effects and state-year fixed effects in all regressions. Finally, we control for the natural logarithm of the child population in columns (7) and (8) and, in all regressions, the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.8: Divorce Does Not Drive the Effects

	Divorce		Maltreatment	
			Overall	Neglect
	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Unemployment Rate	**0.0060******	**0.0018****	**0.097****	**0.19******
	(0.0017)	**(0.00084)**	**(0.049)**	**(0.066)**
Lag Unemployment Rate		**0.0062******		
		(0.0020)		
Divorce Rate			-0.34	-0.38
			(0.37)	(0.57)
State-Year Fixed Effects	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes
ln(Child Population)	No	No	Yes	Yes
Observations	21,354	21,352	20,766	21,354
Counties	2,799	2,799	2,767	2,799
States	46	46	45	46
Mean of outcome	0.12	0.12	3.75	3.04
Mean of Unemployment Rate	7.00	7.00	7.02	7.00
Kleibergen-Paap F-stat	30.0		30.7	31.2
SW F-stat Unemployment Rate		43.2		
SW F-stat Lag Unemployment Rate		40.3		

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. In this Table, we test the divorce mechanism. In column (1), we use the baseline Bartik IV strategy to look at whether unemployment causes a change in the divorce rate. We consider the divorce rate among individuals aged 18 and over. In column (2), we allow for an lagged effect of unemployment on divorce. To do so we add a second instrument, which is simply the lag of the original instrument. The divorce rate is only measured from 2005 and onwards, which is why we do not lose one year of data when we add the lagged unemployment rate. In columns (3) and (4), we then control for the divorce rate in the baseline regression. In columns (3) and (4), the dependent variable is the natural logarithm of the number of incidents of that abuse type per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We control for county fixed effects and state-year fixed effects in all regressions. Finally, we control for the natural logarithm of the child population in columns (3) and (4), and, in all regressions, the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.9: The Main Results Capture an Effect on the Actual Incidence of Maltreatment, Not Reporting

	Control Reporting		Professional/Parent Reporters		Unsubstantiated Reports	
	Overall	Neglect	Overall	Neglect	Overall	Neglect
	(1)	(2)	(3)	(4)	(5)	(6)
	IV	IV	IV	IV	IV	IV
Unemployment Rate	**0.093***	**0.19*****	**0.098***	**0.23*****	0.035	0.079
	(0.048)	**(0.064)**	**(0.054)**	**(0.087)**	(0.070)	(0.085)
Fraction Employed in Schools	0.55	-0.27				
	(0.57)	(1.15)				
Fraction Employed in Social Services	0.88	1.06				
	(2.46)	(2.88)				
Fraction Employed in Health Care	0.27	1.03				
	(0.36)	(0.96)				
Fraction Employed in Police	-2.89	-0.24				
	(4.19)	(3.65)				
Fraction Employed in Clergy	3.70**	0.34				
	(1.62)	(2.86)				
Fraction Employed in Child Care	-0.091	-0.80				
	(1.36)	(1.65)				
State-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes	Yes	Yes
Observations	20,766	21,354	22,686	23,399	23,468	24,181
Counties	2,767	2,799	2,736	2,768	2,771	2,803
States	45	46	43	44	45	46
Mean of outcome	3.75	3.04	3.35	2.53	4.71	4.13
Mean of Unemployment Rate	7.02	7.00	6.85	6.83	6.86	6.85
Kleibergen-Paap F-stat	32.5	33.0	31.2	31.3	34.5	34.9

Notes. Standard errors (in parentheses) are clustered at the state level. In this Table, we ask whether the main results capture an effect on the actual incidence of child maltreatment or reporting behaviour. In columns (1) and (2), we control for the fraction of the working-age population employed in high-reporting sectors. In columns (3) and (4), we only count incidents that are reported by a professional or the victim's parent. We restrict the regressions in columns (3) and (4) to state-years for which more than 80% of reports had a non-missing reporter. In columns (5) and (6), we count unsubstantiated reports. In each case, the dependent variable is the natural logarithm of the relevant number of incidents of that abuse type per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of abuse across states. We include state-year fixed effects to control for any changes in the definitions of abuse over time within states. Finally, we control for the natural logarithm of the child population and the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 1.10: Robustness of Main Results: Alternative Instruments

	Non-Bartik		Bartik					
	Real Exchange Rate		Value Added		Employment		1995 Weight	
	Overall	Neglect	Overall	Neglect	Overall	Neglect	Overall	Neglect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IV	IV	IV	IV	IV	IV	IV	IV
Unemployment Rate	0.028	**0.15***	0.086	**0.24****	0.038	**0.21****	**0.073***	**0.17*****
	(0.065)	**(0.087)**	(0.078)	**(0.12)**	(0.085)	**(0.10)**	**(0.038)**	**(0.053)**
State-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,468	24,181	23,468	24,181	23,468	24,181	23,459	24,172
Counties	2,771	2,803	2,771	2,803	2,771	2,803	2,770	2,802
States	45	46	45	46	45	46	45	46
Mean of outcome	3.77	3.05	3.77	3.05	3.77	3.05	3.77	3.05
Mean of Unemployment Rate	6.86	6.85	6.86	6.85	6.86	6.85	6.86	6.85
Kleibergen-Paap F-stat	39.0	35.8	20.3	22.2	10.2	12.9	41.2	41.1

Notes. Standard errors (in parentheses) are clustered at the state level. In this Table, we check the robustness of the main results to using alternative instruments. In columns (1) and (2), we use the change in the real exchange rate from the previous year, multiplied by the share of the employed working-age population in manufacturing at the county level in 2003. In columns (3)-(8), we use different versions of the Bartik IV. In columns (3) and (4), we use the national-level value added by industry, weighted by the share of the employed working-age population at the county level in 2003. In columns (5) and (6), we use the national-level total number employed by industry, weighted by the share of the employed working-age population at the county level in 2003. In columns (7) and (8), we use the national-level unemployment rate by industry, weighted by the share of the employed working-age population at the county level in 1995. In each case, we present the results from the second stage. The dependent variable is the natural logarithm of the relevant number of incidents of that abuse type per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of abuse across states. We include state-year fixed effects to control for any changes in the definitions of abuse over time within states. Finally, we control for the natural logarithm of the child population and the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix

1.A Additional Tables and Figures

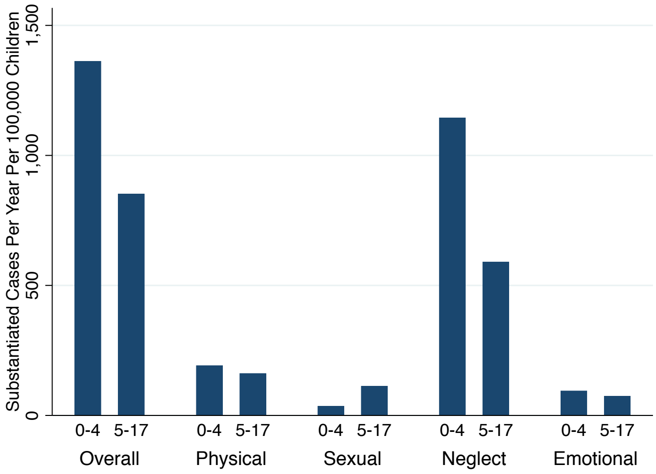


Figure A.1: *Notes.* **Abuse Rates by Child Age.** In this Figure, we plot the weighted mean abuse rate per year per 100,000 children for children aged 0-4 years old and 5-17 years old, for each abuse type. In each county-year, we calculate the abuse rate per 100,000 children by dividing the number of incidents for each age group by the number of children of that age group, using population estimates from the Population and Housing Unit Estimates, before multiplying by 100,000. We then take a weighted mean of the abuse rates across all county-years, where the weights are the child population of the relevant age group in each county-year. We use all available county-years for the entire sample period 2004-12. If a child is abused in multiple ways, we count the case only once in the overall abuse measure, which is why the sum of abuse rates across all individual abuse types exceeds the overall abuse rate.

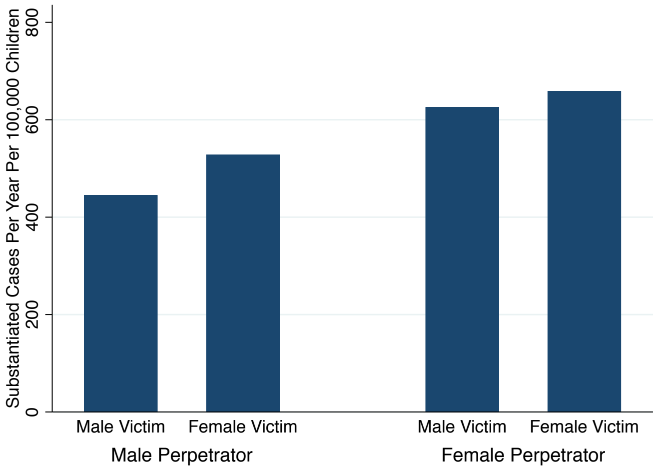


Figure A.2: *Notes.* **Overall Abuse Rate by Perpetrator and Victim Gender.** In this Figure, we plot the weighted mean overall abuse rate per year per 100,000 children for each perpetrator and victim gender combination. In each county-year, we calculate the abuse rate per 100,000 children by dividing the number of incidents of overall abuse in each perpetrator-victim gender group by the number of children of the victim's gender aged 0-17, using population estimates from the Population and Housing Unit Estimates, before multiplying by 100,000. We then take a weighted mean of the abuse rates across all county-years, where the weights are the population of children of the victim's gender aged 0-17 in each county-year. We use all available county-years for the entire sample period 2004-12.

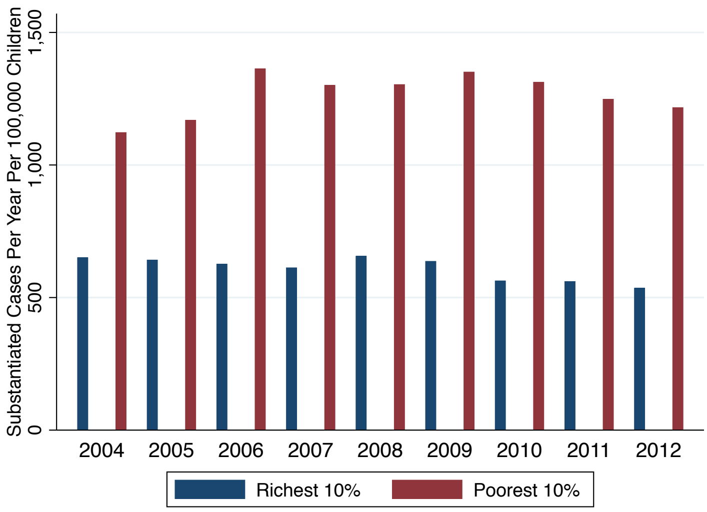


Figure A.3: *Notes.* **Overall Abuse Rate Trends for the Least and Most Poor Counties**. In this Figure, we present the weighted mean overall abuse rate in each year for the initially poorest and least poor 10% of counties in the U.S. We take the 2,803 counties that are included in the final regression analysis, and classify the poorest 10% and least poor 10% in 2003 using the poverty rate taken from the Small Area Income and Poverty Estimates. The poorest 10% of counties are the 273 counties that had a 2003 poverty rate of greater than 20.1%, and the least poor 10% of counties are the 269 counties that had a 2003 poverty rate of less than 7.8%. In each county-year, we calculate the abuse rate per 100,000 children by dividing the number of incidents of overall abuse by the number of children aged 0-17, using population estimates from the Population and Housing Unit Estimates, before multiplying by 100,000. We then take the weighted mean of the abuse rates across all county-years among the least poor or poorest 10% of counties respectively, where the weights are the child population in each county-year. We use all available county-years in any given year to calculate this weighted mean.

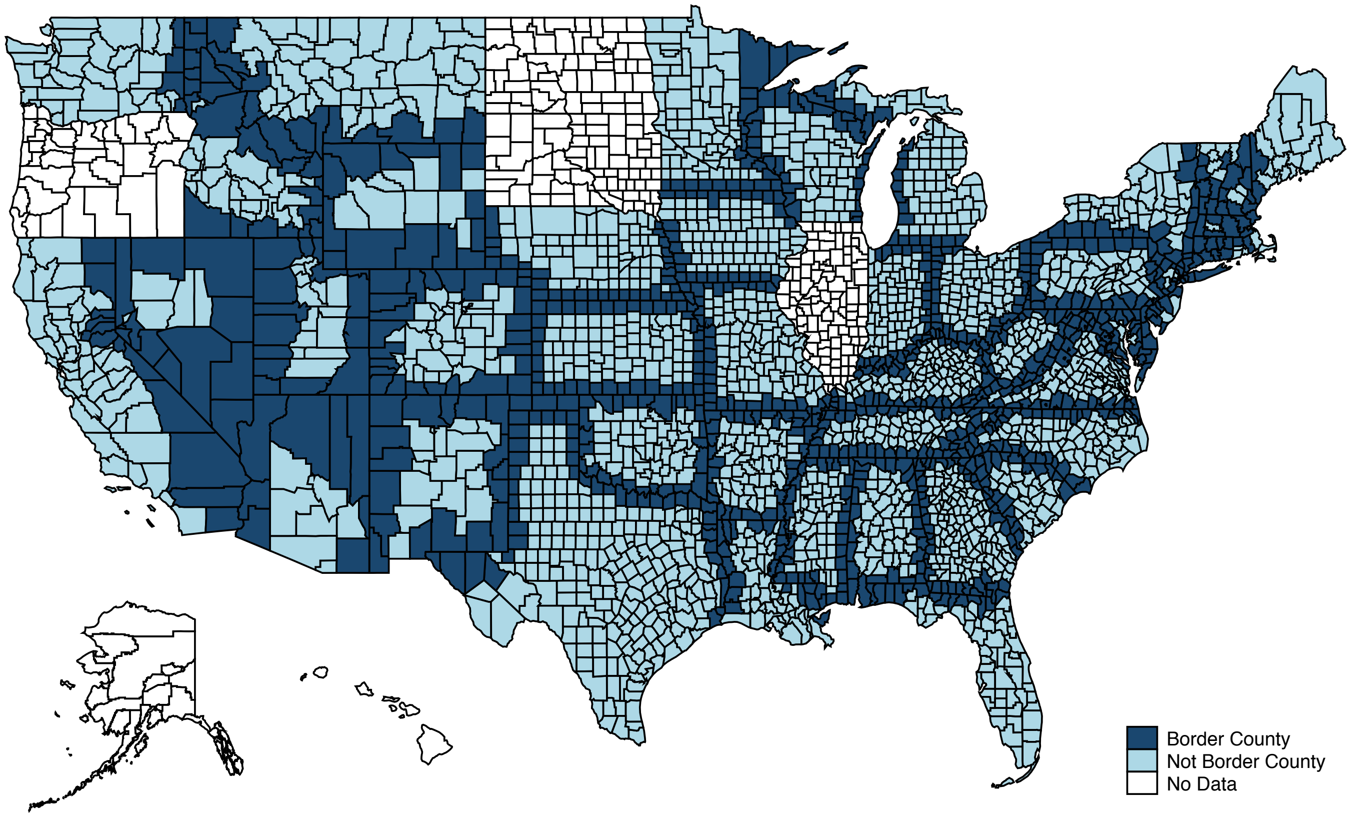


Figure A.4: *Notes.* **Border Counties**. This map highlights the counties that are included in the sample for the regressions that use the contiguous border counties methodology in Table 1.3. These are counties that are part of a contiguous pair which straddles a state border, where both counties are observed in the NCANDS dataset. There are 987 such counties, and they are depicted in dark blue. Alaska and Hawaii are in the bottom left of the figure.

Table A.1: First Stages for the Baseline IV Regressions

	Overall/Emotional	Neglect/Physical/Sexual
	(1)	(2)
	IV	IV
Bartik IV	**0.60*****	**0.63*****
	(0.10)	**(0.11)**
State-Year Fixed Effects	Yes	Yes
County Fixed Effects	Yes	Yes
Controls for Ethnic Group	Yes	Yes
ln(Child Population)	Yes	Yes
Observations	23,468	24,181
Counties	2,771	2,803
States	45	46
Kleibergen-Paap F-stat	34.5	34.9

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. This Table contains the first stage results for the baseline IV estimates in columns (2), (4), (6), (8), and (10) of Table 1.2. In column (1), we present results from the first stage for overall abuse and emotional abuse. There are fewer observations for those two types of abuse because some state-years are missing observations for emotional abuse. In column (2), we present results from the first stage for neglect, physical abuse and sexual abuse. In each case, we regress the unemployment rate on the Bartik IV. We control for county fixed effects, state-year fixed effects, the natural logarithm of the child population and the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: By Victim Age

	Overall		Neglect	
	Age 0-4	Age 5-17	Age 0-4	Age 5-17
	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Unemployment Rate	**0.13****	**0.10****	**0.25*****	0.12
	(0.062)	**(0.046)**	**(0.084)**	(0.11)
State-Year Fixed Effects	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes
ln(Age-Group Specific Child Population)	Yes	Yes	Yes	Yes
Observations	23,468	23,468	24,181	24,181
Counties	2,771	2,771	2,803	2,803
States	45	45	46	46
Mean of outcome	2.57	3.24	2.12	2.27
Mean of Unemployment Rate	6.86	6.86	6.85	6.85
Kleibergen-Paap F-stat	32.4	35.0	32.6	35.5

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. This Table contains the results of the IV regressions which look at the effect of unemployment on the incidence of overall abuse and neglect by victim age. In each case, the dependent variable is the natural logarithm of the number of incidents of abuse of that type per year for a victim of the given age group, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We look at two age groups: children aged 0-4 years and children aged 5-17 years. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of abuse across states. We include state-year fixed effects to control for any changes in the definitions of abuse over time within states. Finally, we control for the natural logarithm of the child population of the given age group and the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population of the given age-group in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: By Perpetrator Gender

	Overall		Neglect	
	Male	Female	Male	Female
	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Unemployment Rate	0.031	**0.18*****	0.13	**0.28*****
	(0.060)	**(0.070)**	(0.092)	**(0.083)**
State-Year Fixed Effects	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes
Observations	20,893	20,893	21,606	21,606
Counties	2,509	2,509	2,541	2,541
States	42	42	43	43
Mean of outcome	2.89	2.97	1.59	2.58
Mean of Unemployment Rate	6.87	6.87	6.85	6.85
Kleibergen-Paap F-stat	29.3	29.3	30.0	30.0

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. This Table contains the results of the IV regressions which look at the effect of unemployment on the incidence of overall abuse and neglect by perpetrator gender. In each case, the dependent variable is the natural logarithm of the number of incidents of abuse of that type per year perpetrated by someone of the given gender, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of abuse across states. We include state-year fixed effects to control for any changes in the definitions of abuse over time within states. Finally, we control for the natural logarithm of the child population and the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: SNAP and EITC: First Stages

	SNAP		EITC	
	Unemployment Rate	SNAP x Unemployment Rate	Unemployment Rate	EITC x Unemployment Rate
	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Bartik IV	0.87***	0.41***	0.75***	1.64*
	(0.14)	(0.16)	(0.12)	(0.98)
SNAP x Bartik IV	-0.36***	-0.50*		
	(0.13)	(0.28)		
EITC x Bartik IV			-0.017**	0.068
			(0.0077)	(0.21)
State-Year Fixed Effects	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes
Observations	21,833	21,833	19,494	19,494
Counties	2,803	2,803	2,803	2,803
States	46	46	46	46
SW F-stat for Unemployment Rate	30.6		4.53	
SW F-stat for Interaction		2.38		0.37

Notes. Standard errors (in parentheses) are clustered at the state level. This Table presents the first stage results for IV regressions which ask whether the Supplemental Nutritional Assistance Program (SNAP) or the state-level Earned Income Tax Credit (EITC) mitigated the effect of unemployment on neglect. In columns (1) and (2), we consider SNAP. We create a variable that tells us what fraction of the twelve months of the year each state had relaxed asset test eligibility requirements for SNAP. We interact this variable with the unemployment rate, and create a second instrument by interacting this variable with the Bartik instrument. In column (1), we present the results from the first stage for the unemployment rate, and in column (2), the results from the first stage for the interaction. In columns (3) and (4), we consider EITC. We create a variable that tells us the percentage of the federal EITC offered by each state in each year in its own state-level EITC. As with SNAP, we interact this variable with the unemployment rate and create a second instrument by interacting this variable with the Bartik instrument. In column (3), we present results from the first stage for the unemployment rate, whilst in column (4), we present results from the first stage for the interaction. Our measure of SNAP is only available for the years 2004-11, whilst our measure of EITC is only available for the years 2004-10. We weight observations by the child population in the county-year in all regressions.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Effect on Neglect for Children Living in Inadequate Versus Adequate Housing

		Inadequate	Adequate
	(1)	(2)	(3)
	IV	IV	IV
Unemployment Rate	**0.19****	-0.11	**0.20****
	(0.076)	(0.15)	**(0.084)**
State-Year Fixed Effects	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes
Observations	11,152	11,152	11,152
Counties	1,514	1,514	1,514
States	23	23	23
Mean of outcome	3.38	0.15	2.96
Mean of Unemployment Rate	7.22	7.22	7.22
Kleibergen-Paap F-stat	21.3	21.3	21.3

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. To calculate p-values for the tests of statistical significance, we compare test statistics to the T(G-1) distribution due to the small number of clusters (where G is the number of clusters). This Table contains the results of IV regressions which look at the effect of unemployment on neglect, allowing for different effects among children living in inadequate and adequate housing. In column (2), we count incidents where the child is described as living in inadequate housing by the Child Protection Services officer. In column (3), we count incidents where the child is described as living in adequate housing by the Child Protection Services officer. We only include state-years for which more than 80% of reports had a non-missing value for this housing measure. We are left with a sub-sample of 23 states. In column (1), we therefore run the baseline regression on that sub-sample of states to check whether the main effects still exist. In each case, the dependent variable is the natural logarithm of the number of incidents of neglect per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of neglect across states. We include state-year fixed effects to control for any changes in the definitions of neglect over time within states. Finally, we control for the natural logarithm of the child population and the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Correlation Between Poverty Rate and Private Safety Net Coverage

	Fraction Uninsured Children	Fraction Children Two Employed Parents
	(1)	(2)
	OLS	OLS
Poverty Rate	**0.72*****	**-1.29*****
	(0.099)	**(0.11)**
Observations	2,802	2,802
States	46	46
Mean of outcome	0.12	0.44
Mean of Poverty Rate	13.4	13.4

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. In this Table, we ask whether the poverty rate in 2000 is unconditionally correlated with each of the two private safety net measures used in Table 1.4. The poverty rate is measured as a fraction of the population. In column (1), the dependent variable is the fraction of children who are not covered by health insurance in 2000. In column (2), the dependent variable is the fraction of children who have two employed parents in 2000. We restrict the sample to the 2,803 counties included in the baseline regressions. For one county the poverty rate in 2000 is missing. We weight observations by the child population in the county in 2000.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Probability of Having Two Employed Parents Present in the Household by Ethnic Group

	Black, Hispanic and White Sample	Black and Hispanic Sample
	(1)	(2)
	OLS	OLS
Black	**-0.23*****	**-0.074*****
	(0.011)	**(0.0097)**
Hispanic	**-0.16*****	
	(0.0098)	
Observations	277,810	74,074
Mean of outcome	0.39	0.25

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. In this Table, we present results from OLS regressions which ask whether there is a difference in the probability that children have two employed parents present in the household by ethnic group. In both regressions, the dependent variable is a dummy variable that takes the value 1 if a child has two employed parents living in the household with them, and 0 otherwise. The unit of observation is the child, and we include children aged 0-17 from the 2003 round of the American Community Survey in the sample. In column (1), we restrict the sample to White, Black and Hispanic children, and compare the probability for Black and Hispanic children to the White base group. In column (2), we restrict the sample to Black and Hispanic children, and compare the probability for Black children to the Hispanic base group. We define Black and White as Black-not-Hispanic and White-not-Hispanic respectively. We weight observations using the person weights provided in the ACS.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Robustness of Main Results: Alternative Choices for the LHS Variable

	Ln(Rate)		IHS Rate		Add 0.01		Add 0.0001		Ln(Number of Children)	
	Overall	Neglect	Overall	Neglect	Overall	Neglect	Overall	Neglect	Overall	Neglect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	IV	IV	IV	IV	IV	IV	IV	IV	IV	IV
Unemployment Rate	**0.096***	**0.21*****	**0.094***	**0.18*****	**0.093***	**0.18*****	**0.099****	**0.23*****	0.11	**0.25*****
	(0.053)	**(0.073)**	**(0.053)**	**(0.066)**	**(0.048)**	**(0.063)**	**(0.050)**	**(0.074)**	(0.067)	**(0.092)**
State-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ln(Child Population)	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,468	24,181	23,468	24,181	23,468	24,181	23,468	24,181	17,403	17,748
Counties	2,771	2,803	2,771	2,803	2,771	2,803	2,771	2,803	2,046	2,078
States	45	46	45	46	45	46	45	46	37	38
Mean of outcome	6.41	5.67	7.15	6.51	3.86	3.23	3.68	2.87	3.66	2.88
Mean of Unemployment Rate	6.86	6.85	6.86	6.85	6.86	6.85	6.86	6.85	6.84	6.86
Kleibergen-Paap F-stat	32.7	33.1	32.7	33.1	34.5	34.9	34.5	34.9	25.2	26.7

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. This Table contains the results of IV regressions which look at the robustness of the effect of unemployment on the incidence of overall abuse and neglect with respect to the choice of the dependent variable. In columns (1) and (2), we take the natural logarithm of the rate of abuse per 100,000 children. We first add 0.001 incidents to the number of abuses before taking the rate, to ensure that no county-years have a rate of zero before taking the natural logarithm. In columns (3) and (4), we take the inverse hyperbolic sine transformation of the rate of abuse per 100,000 children. In columns (1)-(4), we do not control for the natural logarithm of the child population, as the left hand side variable is already based on a rate per 100,000 children. In columns (5) and (6), the dependent variable is the natural logarithm of the number of incidents of overall abuse or neglect, after first adding 0.01. In columns (7) and (8), we again take the natural logarithm of the number of incidents, but now add 0.0001. In columns (9) and (10), we take the natural logarithm of the number of children who are victims of overall abuse and neglect (after adding 0.001), rather than the number of incidents. Only 38 states record a unique child ID variable, and so can be included in these two regressions. In columns (5)-(10), we control for the natural logarithm of the child population. We control for county fixed effects in all regressions. We have therefore implicitly controlled for state fixed effects, and so control for differences in definitions of abuse across states. We include state-year fixed effects to control for any changes in the definitions of abuse over time within states. Finally, we control for the fraction of the overall population that is Black, Hispanic or Other Race. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Robustness of Main Results: Specification Changes

	Control County Trends		Drop State-Year FE		Reduced Form		2003 Child Weight		Total Pop Weight		Cluster County Level	
	Overall	Neglect	Overall	Neglect	Overall	Neglect	Overall	Neglect	Overall	Neglect	Overall	Neglect
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	IV	IV	IV	IV	OLS	OLS	IV	IV	IV	IV	IV	IV
Unemployment Rate	0.042	**0.098****	0.067	**0.13****			**0.095***	**0.20*****	**0.096****	**0.20*****	**0.096***	**0.20*****
	(0.041)	**(0.046)**	(0.052)	**(0.059)**			**(0.049)**	**(0.067)**	**(0.045)**	**(0.063)**	**(0.051)**	**(0.063)**
Bartik IV					**0.054***	**0.13*****						
					(0.028)	**(0.038)**						
State-Year Fixed Effects	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Linear County Trends	Yes	Yes	No	No	No	No	No	No	No	No	No	No
Controls for Ethnic Group	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,468	24,181	23,468	24,181	23,480	24,193	23,468	24,181	23,468	24,181	23,468	24,181
Counties	2,771	2,803	2,771	2,803	2,771	2,803	2,771	2,803	2,771	2,803	2,771	2,803
States	45	46	45	46	45	46	45	46	45	46	45	46
Mean of outcome	3.77	3.05	3.77	3.05	3.77	3.05	3.77	3.05	3.77	3.05	3.77	3.05
Mean of Unemployment Rate	6.86	6.85	6.86	6.85			6.86	6.85	6.86	6.85	6.86	6.85
Kleibergen-Paap F-stat	33.6	30.5	21.0	22.6			32.3	33.0	40.0	40.8	67.4	70.5

Notes. Standard errors (in parentheses) are clustered at the state level in the regressions in columns (1) to (10), and at the county level in columns (11) and (12). In this Table, we test the robustness of the main results for overall abuse and neglect. In columns (1) and (2), we control for linear county trends. In columns (3) and (4), we drop state-year fixed effects and add year fixed effects. In columns (5) and (6), we run reduced form OLS regressions with the instrument as the right hand side variable. In columns (7) and (8), we weight observations using the child population in the county in 2003. In columns (9) and (10), we weight observations using the total population in the county-year. Finally, in columns (11) and (12), we cluster standard errors at the level of the county. In each case, the dependent variable is the natural logarithm of the number of incidents of that abuse type per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. In columns (1)-(6) and (11)-(12), we weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Robustness of Main Results: Sample Changes

	Drop Large Counties		Drop Small Counties	
	Overall	Neglect	Overall	Neglect
	(1)	(2)	(3)	(4)
	IV	IV	IV	IV
Unemployment Rate	**0.12***** **(0.043)**	**0.22***** **(0.065)**	**0.089*** **(0.050)**	**0.20***** **(0.069)**
State-Year Fixed Effects	Yes	Yes	Yes	Yes
County Fixed Effects	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	Yes	Yes
ln(Child Population)	Yes	Yes	Yes	Yes
Observations	23,450	24,163	21,480	22,186
Counties	2,769	2,801	2,545	2,577
States	45	46	45	46
Mean of outcome	3.76	3.05	4.20	3.52
Mean of Unemployment Rate	6.86	6.85	7.02	7.00
Kleibergen-Paap F-stat	35.6	36.1	33.6	34.1

Notes. Standard errors (in parentheses) are clustered at the state level in all regressions. In this Table, we test the robustness of the main results for overall abuse and neglect. In columns (1) and (2), we drop the two counties with a population of more than one million children in 2003 (Harris, Texas; and Los Angeles, California). In columns (3) and (4), we drop the counties in the smallest 10% of counties in the United States in terms of 2003 child population, which are the counties with fewer than 1,232 children. In each case, the dependent variable is the natural logarithm of the number of incidents of that abuse type per year, after first adding 0.001 incidents to every county-year to ensure that no county-years have zero incidents. We weight observations by the child population in the county-year.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Data Appendix

Variable	Data Source	Method
Overall Abuse, Neglect, Physical Abuse, Sexual Abuse, Emotional Abuse	National Child Abuse and Neglect Data System (NCANDS)	We keep only child-reports where at least one allegation of child maltreatment is found to be substantiated. There can be up to four allegations of child maltreatment on any given child-report. To do this, we keep any child-report for which at least one of the variables Mal1Lev, Mal2Lev, Mal3Lev, Mal4Lev is equal to 1 (substantiated), 2 (indicated or reason to suspect), or 3 (alternative response victim). We keep only child-reports for which the child is aged from 0 to 17 years inclusive. Within each county-year we then sum the total number of substantiated incidents of each type of abuse. To do so we use the report date (RptDt) and calendar year, as explained in Section 1.C.1. For the measures of overall abuse, we treat a child-report with more than one substantiated type of abuse as a single case. For example, if a child is both physically and sexually abused, we treat this as only one incident of overall abuse. We create a measure of the overall number of abuses per year as: $A^*_{cst} = A_{cst}/(O_{cst}/D_t)$, where A^*_{cst} is the number of abuses per year for county c in state s in year t, A_{cst} is the number of abuses over the part of the year that we observe, O_t is the number of days in year t that we observe for county c in state s, and D_t is the total number of days in year t. We then take the natural logarithm of the number of abuses per year. For all county-years, we first add 0.001 to the number of abuses before taking the natural logarithm. Further explanation about the creation of the left hand side variables is given in Section 1.C. The measure of neglect includes medical neglect.
Unemployment Rate	Local Area Unemployment Statistics (LAUS)	We use the annual average unemployment rate at a county-level, produced by the Bureau of Labor Statistics. This dataset can be downloaded from: https://www.bls.gov/lau/.
Predicted Unemployment Rate (Instrument)	Quarterly Census of Employment and Wages (QCEW) and Current Population Survey (CPS-BLS)	The weights for the instrument are the fraction of all employed individuals working in each industry at the county-level in 2003. To calculate this, we use the QCEW. We first sum the annual average number employed 'annual_avg_emplvl' across all ownership sectors (government and private), for each of the 20 NAICS industries at a county level in 2003. This tells us the total number employed in each industry in 2003. We then sum these totals across all industries, and then divide the total employed in each industry by that sum to give the fraction of employed individuals in each industry. We calculate national level unemployment rates in each year using the CPS-BLS. We divide the total unemployed by the sum of the total employed and unemployed in each of the 20 NAICS industries. To calculate the instrument we then take a weighted average of these national level unemployment rates across industries using the weights previously described, which capture the initial industrial structure in each county. The industry of an unemployed person is the industry of their last job. Data from the QCEW can be downloaded from: https://www.bls.gov/cew/. National-level unemployment files from the CPS-BLS were sent to us by the BLS.
Fraction Black, Hispanic and Other Race	Population and Housing Unit Estimates (PHUE)	We calculate the fraction of the population who are Black, Hispanic and Other Race. The PHUE contains a breakdown of the total population by race, where an individual can be classified as Black Alone, White Alone, Asian Alone, American Indian Alone, Alaskan Native Alone, Native Hawaiian or Other Pacific Islander Alone or Two or More Races. The PHUE treats being Hispanic as an ethnic group, rather than a race. For the fraction Black, we use Black Alone (non-Hispanic). For the fraction Other Race, we use the sum of American Indian Alone, Asian Alone, Alaskan Native Alone, Native Hawaiian or Other Pacific Islander Alone and Two or More Races (non-Hispanic). For the years 2004-9, we use the 2009 Vintage, and for the years 2010-12 we use the 2014 Vintage. Data from the PHUE can be downloaded from: https://www.census.gov/programs-surveys/popest.html.

Data Appendix: Continued

Variable	Data Source	Method
Child Population	Population and Housing Unit Estimates (PHUE)	We take the total number of boys and girls aged 0-17 from the PHUE. For the years 2004-9, we use the 2009 Vintage, and for the years 2010-12 we use the 2014 Vintage. This is also the dataset we use to create the weight for our regressions. Data from the PHUE can be downloaded from: https://www.census.gov/programs-surveys/popest.html.
Benefit Duration	Fatih Karahan, compiled for the paper Hagedorn et al. (2013).	We use a quarterly dataset of the maximum duration of unemployment benefits an individual can claim at the state level, provided to us by Fatih Karahan at the Federal Reserve Bank of New York. We take an unweighted mean across the four quarters to make the data annual. Hawaii is missing from this dataset.
SNAP	USDA SNAP Policy Database	We use the dummy variable 'bbce' from the SNAP Policy Database, which takes the value 1 if the state uses broad-based categorical eligibility to increase or eliminate the asset test and/or increases the gross income limit for virtually all SNAP applicants. The dataset is provided monthly, and so for each state-year we calculate the fraction of the twelve months of the year for which this dummy variable takes the value 1. This dataset is available for the years 2004-11. It can be downloaded from: https://www.ers.usda.gov/data-products/snap-policy-database/.
EITC	State EITC Based on the Federal EITC, Tax Policy Center	We create a variable at the state-year level which measures the percentage of the federal EITC offered by each state in its own EITC. If a state does not offer its own EITC, this percentage is equal to zero. We treat refundable and non-refundable EITCs equivalently (but check the robustness of results to dropping state-years with a non-refundable EITC). Five states have a non-refundable EITC at some point during the sample period, which are: Delaware, Iowa, Maine, Rhode Island and Virginia. Maryland offers two separate rates, one refundable and one non-refundable. We always use the refundable rate. Minnesota's credit is not expressed as a fraction of the federal EITC, but depending on the family's income the credit can range from 25 to 45% of the federal EITC. We use the highest possible tax credit of 45%. New Jersey's EITC is only available to families with incomes less than $20,000. Wisconsin has three separate rates depending on the number of children a family has, and so we again consider the highest rate, which is 43%. This dataset is available for the years 2004-10. It can be downloaded from: http://www.taxpolicycenter.org/statistics/state-eitc-based-federal-eitc.
Fraction Uninsured Children	Small Area Health Insurance Estimates (SAHIE)	We take the county-level count of the number of children under 18 who are not covered by health insurance in 2000, estimated in the SAHIE. We divide this by the child population from the PHUE in 2000, to create a measure of the fraction of children under 18 who are not covered by health insurance in 2000. The SAHIE for 2000 can be downloaded from: https://www.census.gov/support/USACdataDownloads.html. Details on the methodology used in the SAHIE are available from: https://www.census.gov/did/www/sahie/methods/2000/model.html.
Poverty Rate	Small Area Income and Poverty Estimates (SAIPE)	We take the county-level percentage of all ages in poverty in 2000. We divide by 100 to convert this variable to the fraction of all ages in poverty in 2000. This dataset can be downloaded from: https://www.census.gov/did/www/saipe/data/statecounty/. Details on the methodology used in the SAIPE are available from: https://www.census.gov/did/www/saipe/methods/statecounty/2000county.html.
Expenditure Per Capita on Food and Beverages for Off-Premises Consumption, and Clothing and Footwear	Per Capita Expenditure-By-State Statistics, Bureau of Economic Analysis (BEA)	We have a dataset of per capita expenditure on two sets of basic goods: food and beverages for off-premises consumption, and clothing and footwear. This dataset is at the state-year level. It was sent to us by the BEA. The dataset is in nominal terms, and so we convert to 2012 dollars using the All Items All Urban Consumers (Current Series) Consumer Price Index (U.S. City Average) from the BLS. The CPI can be downloaded from: https://www.bls.gov/cpi/#data.

Data Appendix: Continued

Variable	Data Source	Method
Inadequate and Adequate Housing Neglect	National Child Abuse and Neglect Data System (NCANDS)	The NCANDS dataset includes a caretaker risk factor with the variable name FCHouse. FCHouse is a 'risk factor related to substandard, overcrowded, unsafe or otherwise inadequate housing conditions, including homelessness'. This variable is often not recorded by the Child Protection Services officer. We keep only state-years in which this variable is non-missing for more than 80% of substantiated cases. Among those state-years, there are four states where the FCHouse variable is either always equal to 1 (inadequate housing) or 2 (adequate housing) for at least all but one year in the sample period. These are: Colorado, Virginia, Florida and Indiana. We therefore additionally drop these states. We drop observations from Ohio in 2011 for the same reason. We are left with a final sample of 23 states: Alabama, Arizona, Arkansas, Delaware, D.C., Georgia, Idaho, Iowa, Kentucky, Maine, Minnesota, Mississippi, Missouri, Nevada, New Jersey, New Mexico, Ohio, South Carolina, Tennessee, Texas, Utah, Washington and Wyoming. For these states, we separately count substantiated cases for children living in inadequate and adequate housing at the county-year level, and create the natural logarithm of each, after adding 0.001 incidents as with the main dependent variable.
Fraction Children Two Employed Parents	Census 2000	For the regression analysis in Table 1.4, we use the 'empstat_mom' and 'empstat_pop' variables from the census data in 2000 (available from IPUMS), to calculate the fraction of children, aged 0-17, with an employed mother and father living in the household with them in the PUMA. This measure therefore only counts children with heterosexual parents. To reflect the sampling design of the census, we use the person weights ('perwt') to create this fraction. We then use the 2010 county to 2000 PUMA cross-walk provided by the Missouri Census Data Center (Mable/GeoCorr14, available from: http://mcdc.missouri.edu/websas/geocorr14.html) to crosswalk data to the county-level. For counties that are entirely contained within a single PUMA, we assign to the county the value for the PUMA in which it is contained. For counties that cut across more than one PUMA, we assign a weighted average of the values for the PUMAs that intersect with that county, where the weights are the fraction of the county population in each PUMA. For the analysis in Table A.7, we do exactly the same but using the more recent American Community Survey (2003). In both cases, we wished to use the most recent measure available before the start of the sample period, but the regression analysis in Table 1.4 is at the county level, and so to improve precision we used the census from 2000, since it has a larger sample. Data from both the census and the ACS can be downloaded from: https://usa.ipums.org/usa/.
First Time, Second Time, Third Time, Fourth or More Time Neglect	National Child Abuse and Neglect Data System (NCANDS)	We first keep only the states that we can verify record a unique Child ID variable consistently over time. This variable that allows us to track the same child through the panel. Each state provides mapping files to NCANDS, which outline how each variable is recorded. Alabama, Minnesota and South Carolina do not record a unique Child ID variable, and so are dropped. Indiana does not provide mapping files, whilst Arkansas, New Mexico, Texas and Wisconsin do not provide the Child ID file of the mapping file. For these five states we cannot know whether the child ID variable is recorded consistently over time, and so to be cautious we drop these states too. We are therefore left with 38 states for the analysis. For these states, we do the following. We keep only substantiated cases of neglect. For each child-report, we use the child ID variable to record how many times that child has been the victim of neglect in the past. We can therefore understand whether each child-report is the first, second, third, fourth or more time that the child has been the victim of neglect. We then sum the total number of first time cases, second time cases, third time cases, and fourth or more time cases at the county-year level, and create the natural logarithm of each, after adding 0.001 incidents as with the main dependent variable.

Data Appendix: Continued

Variable	Data Source	Method
Heavy Drinking Prevalence	Behavioural Risk Factor Surveillance System	We take the heavy drinking prevalence rate from the Dwyer-Lindgren et al. (2015) paper. Heavy drinking is defined as consuming more than one drink per day for women and two drinks per day for men for the past thirty days. This measure is only available from 2005 onwards.
Binge Drinking Prevalence	Behavioural Risk Factor Surveillance System	We take the binge drinking prevalence rate from the Dwyer-Lindgren et al. (2015) paper. Binge drinking is defined as consuming more than four drinks in a single day for women and five drinks for men at least once during the past thirty days.
Divorce Rate	American Community Survey (ACS)	We calculate the fraction of the total population aged 18 and over who are divorced. To reflect the sampling design of the ACS, we sum the individual person weights to create the total (weighted) number divorced, and divide this by the total (weighted) number of individuals aged 18 and over. The geographic identifier in the ACS is the PUMA, not the county. We therefore use the 2010 county to 2000 PUMA cross-walk provided by the Missouri Census Data Center (Mable/GeoCorr14, available from: http://mcdc.missouri.edu/websas/geocorr14.html), as explained above in the description for the Fraction Children Two Employed Parents. The ACS only contains PUMA information for the years 2005 to 2012, and so the divorce rate is only defined for those years. Data from the ACS can be downloaded from: https://usa.ipums.org/usa/.
Fraction Employed in Schools, Health Care, Social Services, Police, Clergy, Childcare	American Community Survey (ACS)	For each high-reporting sector, we calculate the fraction of the working age (18-64) population who are employed in each sector. To reflect the sampling design of the ACS, we sum the individual person weights to create the total (weighted) number employed in each sector, and divide this by the total (weighted) number of working age individuals. The geographic identifier in the ACS is the PUMA, not the county. We therefore use the 2010 county to 2000 PUMA cross-walk provided by the Missouri Census Data Center (Mable/GeoCorr14, available from: http://mcdc.missouri.edu/websas/geocorr14.html), as explained above in the description for the Fraction Children Two Employed Parents. The ACS only contains PUMA information for the years 2005 to 2012, and so these variables are only defined for those years. Data from the ACS can be downloaded from: https://usa.ipums.org/usa/.
Substantiated Incidents Reported by Profession-als/Parents	National Child Abuse and Neglect Data System (NCANDS)	We create a measure of the total number of incidents of overall abuse and neglect in the county-year as for the baseline dependent variables, except now only counting incidents reported by a professional, parent, alleged victim or perpetrator. To do this, we use the variable RptSrc in the NCANDS dataset, and keep only child-reports for which RptSrc is equal to 1, 2, 3, 4, 5, 6, 7, 8, 9 or 12. We do not count incidents reported by Other Relative (RptSrc = 10), Friend/Neighbour (RptSrc = 11), Anonymous Reporter (RptSrc = 13), Other (RptSrc = 88) or Unknown (RptSrc = 99). We keep state-years in which more than 80% of substantiated reports have a non-missing reporter (for which RptSrc = 13 or 88 are treated as non-missing).
Unsubstantiated Incidents	National Child Abuse and Neglect Data System (NCANDS)	We create a measure of the total number of unsubstantiated incidents of overall abuse and neglect in the county-year. We use the same approach as for the baseline dependent variables, except that we now count observations for which the relevant MalLev variable is equal to 4, 5, 6, 7, 8 or 88. Again, we consider the disposition with respect to the specific type of maltreatment of interest. A child may be a substantiated victim of one type of maltreatment and an unsubstantiated victim of another type of maltreatment on the same report.

Data Appendix: Continued

Variable	Data Source	Method
Real Exchange Rate Instrument	Real Trade Weighted Dollar Index: Broad (TWEXBPA), Federal Reserve Bank of St. Louis; QCEW	We take the real exchange rate as published by the Federal Reserve Bank of St. Louis. This can be downloaded from: https://fred.stlouisfed.org/series/TWEXBPA. With this, we calculate the change in the real exchange rate on the previous year as: $(RER_t - RER_{t-1})/RER_{t-1}$. We then multiply the change in the real exchange rate by the county-level fraction of the employed working-age population in manufacturing at the start of the sample period, in 2003 (the manufacturing industry weight used to create the original Bartik instrument), from the QCEW.
Real Value Added Bartik Instrument	GDP-by-Industry, Bureau of Economic Analysis (BEA); QCEW	We take a weighted average of the national-level real value added in each of the 20 NAICS industries, where the weights are the fraction of the employed working-age population in each industry in 2003 as in the original Bartik instrument. Real value added by industry is measured in billions of 2009 dollars, and can be downloaded from: https://www.bea.gov/industry/gdpbyind_data.htm.
Employment Bartik Instrument	GDP-by-Industry/ Employment, Bureau of Economic Analysis (BEA); QCEW	We take a weighted average of the national-level total employed in each of the 20 NAICS industries (obtained from Table 6.4D of the NIPA Tables, for each year 2004-12), where the weights are the fraction of the employed working-age population in each industry in 2003 as in the original Bartik instrument. The total employed variable is measured in thousands and can be downloaded from: https://www.bea.gov/industry/more.htm.
Number of Children Maltreated	National Child Abuse and Neglect Data System (NCANDS)	We first keep only the states that we can verify record a unique Child ID variable consistently over time, as explained above for the description of the First Time, Second Time, Third Time and Fourth or More Time variables. We are therefore left with 38 states for the analysis. For these states, we do the following. For each type of abuse or neglect, we keep only substantiated cases. For each child-report, we use the child ID variable to record how many times that child has been the victim of overall abuse or neglect in that year. We then keep only the first time cases within each county-year, and sum these. This gives us a measure of the total number of children who have been the victim of overall abuse or neglect in each county-year. We create the natural logarithm of the number of children after first adding 0.001, as with the main dependent variable.

1.B Safety Nets: Additional Details

1.B.1 Social Safety Nets: SNAP and EITC

The US Department of Agriculture's (USDA) SNAP, previously known as the Food Stamp program, offers financial assistance for food purchases. Like UI, SNAP was also expanded after the recession (Mulligan, 2012). The federal government allowed states to increase or eliminate asset tests for food stamps, thereby widening eligibility. As a consequence, by 2010 half of non-elderly households with an unemployed head or spouse participated in the program (Hagedorn et al., 2013). We create a variable, $SNAP$, that measures the fraction of months in the year for which each state increased or eliminated asset tests, using monthly data from the USDA's SNAP Policy Database for the years 2004-11.

The EITC is a tax credit that is primarily targeted towards families with children (Bitler & Hoynes, 2016). The tax credit is refundable, meaning that any surplus credit over tax liabilities can be claimed as a benefit. Though it is targeted at people in work, unemployed individuals can claim EITC providing they have worked at some point during the tax year. The main EITC is federal, but some states additionally offer their own EITC (Williams, 2017). By the end of the sample period, twenty-three states had their own EITC, which varied in generosity from 3.5 to 45% of the federal EITC.[51] We create a variable $EITC$, which measures the percentage of the federal EITC offered by each state in each year.[52] This dataset comes from the Tax Policy Center, and is available for the years 2004-10.

We ask whether the effect of unemployment on neglect is smaller in states that increase or eliminate asset tests for SNAP, or offer a larger tax credit in the state EITC.

[51] Every state other than Minnesota directly set its state EITC as a percentage of the federal EITC. Depending on the family's income level, the tax credit in Minnesota ranged between 25 and 45% of the federal credit.

[52] For the vast majority of states, the state EITC is refundable like the federal EITC. However, five states had a non-refundable state EITC for at least some years during the sample period. We therefore also create a variable that treats any state-year with a non-refundable EITC as missing and the results are very similar. Further details are in the Data Appendix.

In separate regressions, we interact the unemployment rate with $SNAP$ and $EITC$ respectively, and interact those variables with the Bartik instrument to create a second instrument. However, we cannot learn much about the effect of either program, because the instruments do not identify separate variation in each of the endogenous variables well. The Sanderson-Windmeijer F-statistic on the first stage for the interaction term is just 2.38 and 0.37 for SNAP and EITC respectively. We present the first stage results in Table A.4.

1.B.2 Private Safety Nets and Poverty

We create dependent variables that separately count the number of cases of neglect for children living in inadequate and adequate housing. The indicator of housing conditions is not recorded for every child-report. For the regressions in Table A.5, we restrict the sample to state-years in which this indicator is non-missing for over 80% of substantiated reports. This gives us a sub-sample of twenty-three states, and so in column (1) of Table A.5, we first check that the effect of unemployment on neglect still exists in that sub-sample, which it does. In columns (2) and (3) of Table A.5, we see that the effect of unemployment on neglect occurs for children living in adequate housing. Since we cluster standard errors at the state level, Wald tests may now overreject as we have a small number of clusters. To deal with this, we compare test statistics to the T(G-1) distribution, where G = 23 (the number of clusters), following Cameron & Miller (2015). We cannot use a wild bootstrap, the preferred method of dealing with a small number of clusters, as the properties of the wild bootstrap have not been derived for the IV estimator.

1.C Construction of the Dependent Variables: Additional Details

1.C.1 Organising the Data by Calendar Year and Report Date

The NCANDS data is released annually, and is organised by Federal Fiscal Year (FFY) (running from 1st October to 30th September), and by the investigation disposition date (the date of the outcome of the CPS investigation). For example, the NCANDS dataset for 2012 contains every child-report for which the outcome of the investigation occurred between 1st October 2011 and 30th September 2012. We would ideally like to organise the data by the date of the incident of abuse, but that is unobserved. The closest that we can get to the date of the incident is the date of report, which is also contained in the dataset. We therefore reorganise the data by the date of report and calendar year. This seems straightforward. However, an issue arises because seventeen states are not observed in at least one year during the sample period. The problem is that one missing year of NCANDS data by the federal fiscal year and investigation disposition date does not translate into only one missing year by calendar year and report date. To see this, take the example of Indiana, as demonstrated in Figure C.1. Indiana is missing the FFY 2012. Our dataset for this state then does not include any incident whose investigation was concluded between 1st October 2011 and 30th September 2012, as indicated by the solid cross. Now suppose that an incident is reported on 15th September 2011. Whilst this incident is reported within a 'non-missing' FFY of the dataset, if the investigation was concluded more than 15 days after the report was made, then the investigation disposition date falls in a missing FFY and this incident will be missing from the data. To deal with this, for each missing FFY of the data, we extend the missing dates to twelve months before the start of the missing FFY, as demonstrated by the dashed cross in the Figure. Over 99% of reports reach an investigation disposition within twelve months of the report date, and so doing this we can claim to capture over 99% of all cases of child

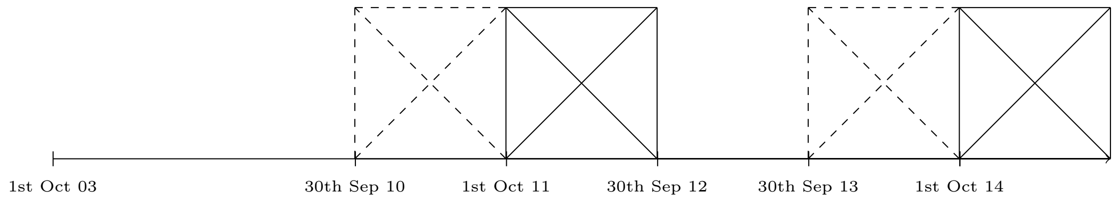


Figure C.1: Missing Years by Report Date for Indiana. The solid crosses indicate the missing periods of data for Indiana by investigation disposition date (the Federal Fiscal Year 2012, and from 1st October 2014 onwards). To organise the data by report date, we treat both the solid and dashed time periods as missing. In other words, we extend the missing period by twelve months before the start of the missing FFY by investigation disposition date. To see the intuition: for the first missing year in the Figure, we know that more than 99% of all incidents reported before 30th September 2010 will have had their investigation disposition before the start of the missing year (1st October 2011), and will therefore appear in the dataset.

abuse in the final sample period.

As can be seen in Figure C.1, for some state-years we then only observe reports for part of the calendar year. For example, for Indiana in 2010 we only observe reports from 1st January until 30th September 2010. To deal with this, we firstly restrict the sample period to 2004 to 2012 (when the majority of states have a complete year's worth of data). Secondly, for the states with missing years, we calculate the number of abuses per year as: $A^*_{cst} = A_{cst}/(O_{cst}/D_t)$, where A^*_{cst} is the number of abuses per year for county c in state s in year t, A_{cst} is the number of abuses over the part of the year that we observe, O_t is the number of days in year t that we observe for county c in state s, and D_t is the total number of days in year t. After creating the measure of the number of abuses per year in this way, we take the natural logarithm transformation as explained in Section 1.3.2.1.

1.C.2 Dealing with Missing Counties

The county of report is typically the county in which the victim resides. However, in a small number of states (for example Utah), it is the county where the office investigating the report of child abuse lies. Even in these states, the two are generally the same because the county-level office tasked with investigating the report is located in the county in which the child resides. However, it is possible that in these states some counties may not have a CPS office, in which case the county is missing from the dataset because there is no office located there, rather than because there were truly no incidents of child abuse. To deal with this, we assume that the former explanation is true only if a county is missing from the dataset for every type of abuse for every year of the dataset, which is the case for 54 counties. We therefore treat these 54 counties as missing from the dataset throughout, and treat any other county that does not appear in the dataset in a particular year for a particular abuse type as having zero incidents of that abuse type in that year.

Chapter 2

The Gender of the First Child and Parental Labour Supply in the UK

with Sonia Bhalotra (University of Essex)

Abstract

Fathers work less after having a first born son rather than a daughter in the UK in the short-term. Fathers are on average 4.5 percentage points less likely to be in employment. On the intensive margin, fathers work on average 2.2 fewer hours per week in the first year of the child's life. We exploit the randomness of the gender of the first child to identify these effects. The effect for degree educated fathers is consistent with the child production function mechanism. Degree educated fathers and their partners both take time off work to increase total parental inputs received by sons, because boys' cognitive and non-cognitive development is less advanced than that of girls at a young age. The effect for non-degree educated fathers is consistent with the new man mechanism. Non-degree educated fathers take time off work to reduce the burden of childcare on their partners, enabling their partners to work. We show that child birthweight, pure son preference or the higher costs of raising daughters than sons do not drive the effects.

JEL: J13, J16, J22

Keywords: child gender, labour supply, parental time, UK.

We thank the participants in the Family Research Group workshop at the University of Essex, the Gorman workshop at the University of Oxford, and the Understanding Society Scientific Conference 2017 organised by the University of Essex.

2.1 Introduction

In this paper, we ask: what is the effect of the gender of the first child on parental labour supply in the UK in the short-term? New parents must choose how to allocate their time between work outside the home and childcare. According to an emerging literature in economics on early childhood inputs, this choice can have important implications for the child's later life outcomes (for a review, see Heckman & Mosso (2014)). Mild shocks to parental investments in pre-school years can have large effects on educational attainment, employment, earnings and health in adulthood (Heckman et al., 2013; Almond et al., 2017). Indeed, even pre-natal inputs can have impacts on adult outcomes (Almond & Currie, 2011). An understudied aspect of the early life environment, which we investigate in this paper, is whether a child's gender affects their parents' choice between work and childcare.

Our research question is also interesting for UK policymakers. The government introduced Shared Parental Leave in April 2015, to try to encourage a greater role for fathers in childcare than has traditionally been the case. This scheme allows parents to divide fifty-two weeks of parental leave between them however they choose, which replaces a previous system where fathers were eligible for just two weeks of statutory paternity leave. Shared Parental Leave was introduced at the end of our sample period. Nonetheless, by understanding how fathers make their choice between work and childcare, we can get a better understanding of whether and how these policies might work.

The existing literature on the effect of child gender on parental labour supply is limited to a handful of papers that focus on the United States and, for one paper, West Germany (Lundberg & Rose, 2002; Choi et al., 2005; Lundberg, 2005; Wulff-Pabilonia & Ward-Batts, 2007). We provide the first evidence on this question for the United Kingdom. We also consider the more recent period from 2009 to 2016. The existing literature focuses on the 1970s, 1980s and 1990s, when norms about men's involvement in childcare may have been different to today. A wider literature shows that child gender matters for a range of other parental behaviours in developed countries. For example, child gender affects the probability of marriage, divorce, and custody arrangements in the event of divorce (Lundberg & Rose, 2003; Dahl & Moretti, 2008), voting preferences and risk taking behaviours (Powdthavee et al., 2009; Powdthavee & Oswald, 2010), and even happiness (Kohler et al., 2005).

We find that fathers work less after having a first born son rather than a first born daughter in the short-term. Fathers are on average 4.5 percentage points less likely to be in employment. On the intensive margin, fathers work on average 2.2 fewer hours per week in the first year of the child's life. The gender of first child effect for fathers is large. It is approximately one quarter the size of the effect of having a child (of either gender) for mothers.

Our results concur with more recent papers in the existing literature, which find that fathers work less after a first born son rather than a daughter, and that the effects differ by socio-economic status (Lundberg, 2005; Wulff-Pabilonia & Ward-Batts, 2007). Our results contradict earlier papers in the literature which suggest that fathers work more after having a first born son rather than a daughter (Lundberg & Rose, 2002; Choi et al., 2005).

Our identification strategy relies on the assumption that the gender of the first born child is exogenous conditional on the decision to have a child. We identify the effects from changes in behaviour within person before and after the birth of their first child, by running fixed effects regressions that exploit the panel data structure of the UK Household Longitudinal Study (UKHLS). The UKHLS is the largest existing longitudinal study of UK households, and has been ongoing for six waves over the period from 2009 to 2016. We have a sample of 48,064 observations from 13,540 individuals aged 16-64 who start the sample period having never had a child. Of these, we have 6,848 observations from 1,320 parents (9.7% of individuals in the sample) who have their first child at some point during the sample period. We demonstrate that our results are robust. We show that fathers do not selectively leave the sample based on the gender of the first child, and so attrition is not driving the results. We also present placebo tests to show that the gender of the first child has no effect before that gender is known to the father.

The effects for degree and non-degree educated fathers require different explanations. The effect for degree educated fathers is consistent with the child production function mechanism. According to the child production function mechanism, boys' initial cognitive and non-cognitive development is less advanced than that of girls, and so parents take time off work to devote more parental inputs to boys. Consistent with this, we find that the partners of degree educated fathers also take more time off work after a first born son rather than a daughter, increasing total parental time at home. The fathers themselves invest their additional time at home in the son, and parental time invested in children is positively correlated with development.

The effect for non-degree educated fathers is consistent with the new man mechanism. According to the new man mechanism, men are starting to engage in childcare to reduce the burden on their partners and facilitate their partners' return to work. However, men's first step to engaging in childcare is with boys, which may be due to same-sex parenting norms or because they feel more comfortable looking after a son. Consistent with this, we find that the partners of non-degree educated fathers are more likely to work after a first born son rather than a daughter, though there is a lag between the effects for the fathers and their partners. Further, the effect for the non-degree educated fathers is driven by those with more progressive attitudes about women's role in the labour market.

Combining the effects for the fathers and their partners, we find that a first born son of a degree educated father could receive up to 702 additional hours of parental time than a first born daughter before his fourth birthday, on average, depending on how much of the additional time that parents take off work is spent engaged with the child. By contrast, a first born son of a non-degree educated father receives almost no difference in parental time compared to a first born daughter, on average.

We find little evidence for three alternative explanations of the results. First, according to the child cost mechanism, daughters are more costly to raise than sons, and so parents work more to increase savings after a first born daughter rather than a son. The transitory nature of the gender of first child effects on fathers' labour supply is not consistent with this explanation. Second, we ask whether the effects are driven by child birthweight. Boys have a higher average birthweight than girls, and so deplete the mother more during the birth and may require more breastfeeding soon after birth. Fathers may then take more time off work after a first born son rather than a daughter to support their more physically exhausted partners. We investigate this by controlling for child birthweight, and find no evidence that this explains our results. Third, we ask whether the effects are driven by pure son preference. The gender of the first child has no effect on subsequent fertility or marital decisions. These results, and the heterogeneity in the effect by ethnic group, are not consistent with the pure son preference explanation.

In Section 2.2, we explain the identification strategy and describe the UK Household Longitudinal Study dataset. In Section 2.3, we outline the main results and present robustness tests. In Section 2.4, we investigate the possible mechanisms for the gender of first child effect on fathers' labour supply. We conclude in Section 2.5.

2.2 Identification Strategy and Data

2.2.1 Identification Strategy: The Gender of the First Child as an Exogenous Shock

We wish to understand the effect of child gender on parental labour supply decisions. To do so, we face three main identification concerns. The first comes from the family stopping rule, which is the decision of when to stop having children after the first child. A naive approach might use the number of girls or boys and the total number of children as right hand side variables, arguing that conditional on family size, the gender mix of children is random. Whilst the decision to have a child is not random, the gender of the child plausibly is because parents cannot directly control the child's gender. However, if the reason an individual stops having children is related to the gender of their existing children for unobservable reasons that are also correlated with labour supply, then the gender mix of children will be correlated with labour supply even conditional on the total number of children. For example, suppose that a man subscribes to the traditional social norm that men should be the breadwinners of the household. Ceteris paribus, this man is more likely to be employed. Further, he believes that a son is more valuable to him than a daughter, as he expects a son to provide for him financially in later life. He is then more likely to keep having children until he gets a son, and so we are more likely to observe this man with a succession of daughters, conditional on the total number of children he has. To deal with the bias this creates, we focus on the gender of the *first born* child, following, for example Choi et al. (2005); Lundberg (2005); Powdthavee et al. (2009).

Our second concern is that adoptive and step parents may choose to be the parent of a child for reasons that are related to that child's gender. Since the gender of adopted and step children is much less plausibly random, we focus on first born biological children only throughout the paper.

Thirdly, we were concerned about sex-selective abortion. However, in the UK sex-selective abortion is not permitted.[1] Dubuc & Coleman (2007) demonstrate some evidence that sex-selective abortion is practised by Indian-born mothers in the UK. This is unlikely to be a big concern for our estimates, since Indian-born parents make up just 2.6% of parents in our sample. Further, we present estimates which exclude those parents.[2] Given these responses to the identification concerns, we estimate the following equation separately for mothers and fathers:

$$Y_{it} = \beta_1 (Child_{it} \times Boy_i) + \beta_2 Child_{it} + X'_{it}\phi + year'_{it}\pi + \alpha_i + \epsilon_{it} \tag{2.1}$$

[1] According to the 1967 Abortion Act, abortion is legal if two doctors agree that continuing the pregnancy and having the child would be worse for the mother's physical or mental health than having an abortion. See: www.legislation.gov.uk.

[2] If the Indian-born parents in our sample do practise sex-selective abortion, then we will observe them with children (specifically girls) less often, and so they will under-contribute to the estimate of β_1. In that event, we should not interpret the estimates as representative of Indian-born parents.

Here, Y_{it} is the outcome of interest for individual i in wave t. $Child_{it}$ is a dummy variable taking value 0 if individual i has never had a biological child and value 1 if individual i has ever had at least one biological child (of any gender), by wave t. Boy_i is a dummy variable taking value 1 if individual i's first born child is a son, and 0 if it is a daughter. In X'_{it}, we control only for the age of individual i in the baseline regression, and we control for calendar year dummy variables through the vector $year'_{it}$, with 2009 as the base year.[3] We control for time invariant unobservables, such as an individual's education level, through the individual fixed effects, α_i.

We are interested in three pieces of information from this equation. The main coefficient of interest is the coefficient on the interaction term, β_1. This coefficient tells us the effect of having a first born son rather than a first born daughter, conditional on the decision to have a child. If our key identifying assumption holds, namely that the gender of the first born child is random conditional on the decision to have a child, then we can interpret β_1 causally. We refer to this as the gender of first child effect. We include observations from after a parent has their second and additional children, and so β_1 captures the total effect of the gender of the first child, including any effects that operate through subsequent fertility choices.

This equation also provides us with information about the effect of having a first child, which we refer to as the fertility effect. The fertility effect breaks down into two, depending on whether the first child turns out to be a daughter or a son. The effect of having a first child who is female is given by β_2. The effect of having a first child who is male is given by $\beta_1 + \beta_2$. We cannot interpret these fertility effects causally, since time-varying unobservables may be correlated both with the decision to have a child and labour supply decisions. Controlling for individual fixed effects does at least allow us to control for any time invariant unobservables that are correlated both with the decision to have a child and labour supply. We are nonetheless interested in the fertility effects

[3] Since a given wave is collected across multiple calendar years, these year dummy variables can vary across individuals i within a wave t, and so have the subscript it. We include dummy variables for each year from 2010 to 2016.

since they provide a useful benchmark with which to understand the size of the gender of first child effects.

We cluster standard errors at the level of the primary sampling unit. We therefore allow for correlation in the error term over time within individuals, since primary sampling units are time invariant. In addition, we allow for correlation across neighbours who live in the same primary sampling unit, which is approximately the size of a single UK postcode.[4] We weight observations so that we can interpret our estimates as being representative of the UK population, given that the UKHLS systematically over-samples some demographic groups (for example ethnic minorities).[5]

We also use an event study specification, which allows us to estimate a separate gender of first child effect in each year relative to the birth of the first child. The $\beta_{1,j}$ coefficients tell us the gender of first child effects after the birth of the child, and the $\beta_{1,p}$ coefficients tell us the effects before the birth of the child.

$$\begin{aligned} Y_{it} = &\sum_{p=2}^{4} \beta_{1,p}(RC_{p,it}^{PRE} \times Boy_i) + \sum_{p=2}^{4} \beta_{2,p} RC_{p,it}^{PRE} \\ &+ \sum_{j=0}^{4} \beta_{1,j}(RC_{j,it}^{POST} \times Boy_i) + \sum_{j=0}^{4} \beta_{2,j} RC_{j,it}^{POST} \\ &+ X_{it}'\phi + year_{it}'\pi + \alpha_i + \epsilon_{it} \end{aligned} \tag{2.2}$$

Here, we create two sets of impulse dummy variables. The dummy variables $RC_{p,it}^{PRE}$ take the value 1 if wave t is p waves before individual i has their first child (e.g. a dummy for the period two waves before birth, three waves before birth, and so on). The dummy variables $RC_{j,it}^{POST}$ take the value 1 if individual i has had their first child by wave t, and the child is aged j years old (e.g. child is 0, child is 1, and so on). We take the wave before the birth of the first child, ($p = 1$), as the base year. In the event study figures

[4]In later robustness checks, we cluster standard errors at the level of the individual, and the results are virtually identical.

[5]Specifically, we use the 'a_indinus_xw' weight for almost all regressions, with which we intend to make our estimates representative of the UK population in the first wave of the sample period (conducted 2009-11). We use the 'a_indscus_xw' weight for any regression where the dependent variable is taken from the self-completion questionnaire.

presented later in the paper, we plot the estimates of $\beta_{1,p}$ and $\beta_{1,j}$ in the same figure.[6]

2.2.2 The UK Household Longitudinal Study

We use data from the UK Household Longitudinal Study (UKHLS), produced by the Institute of Social and Economic Research at the University of Essex. The UKHLS is a longitudinal study of UK households, covering six waves. Interviews for each wave were conducted over a 24-month period, with interviews for the first wave in 2009-11, and the last wave in 2014-16. Crucially for our purposes, the UKHLS asks every adult about their fertility history. This presents a major advantage for us over other UK datasets such as the Census or the Labour Force Survey. Having fertility histories allows us to know the gender of the *first born* child, not just the oldest child currently living in the household, which is essential for our identification strategy as explained in Section 2.2.1.

The UKHLS asks each adult whether they have ever had any biological children before the start of the sample period. We restrict our sample to individuals aged 16-64 who have never had a biological child before the sample period begins.[7] We have a final sample of 48,064 observations from 13,540 individuals.[8] Of these, we have 6,848 observations from 1,320 parents (9.7% of all individuals in the sample) who have their first born child at some point during the sample period.[9] We include individuals whether they are married,

[6]Fathers have their first child at different points during the sample period. A father who has his first child in the sixth wave will be observed five times without his child and only once with, whilst a father who has his first child in the second wave will be observed once without his child and five times with. Estimates of the gender of first child effect further away from the base year are therefore driven by fewer fathers and so may be unrepresentative. We therefore treat estimates further from the base year with caution, and drop observations from five waves before the birth of the first child, and from when the child is aged five and older. This is why $p, j < 5$.

[7]We then know that the first child to appear in their household with them during the sample period is their first born child.

[8]Every individual in our final sample was observed in the first wave of the sample period, such that the sampling weight we use in the regression analysis (which is defined only for the first wave of the dataset) is non-missing. These are individuals from the GPS and EMB samples of the UKHLS, who were not Temporary Sampling Members. TSMs do not have a sampling weight defined even if they do appear in the first wave, and are not intended to be in the UKHLS dataset. They are individuals who live in an ethnic minority boost household who are not of the ethnic group that is intended to be over-sampled, and so are not followed up in later rounds of the survey unless they continue to live in that EMB household.

[9]We drop twins from the baseline regressions. However, in later regressions we include the additional 205 observations from 40 parents with twin first births, and the results are very similar. To do so, we

cohabiting or single. For 99% of the observations in which the individual has had a child, that child is aged 4 or less.[10] The Data Appendix table contains detailed information on how we constructed the dataset.

2.2.2.1 Outcome Variables

We focus on two main outcome variables, which capture the quantity of work an individual undertakes. The first is a dummy variable that takes the value 1 if an individual is employed, and 0 if they are not. To construct this variable, we follow the Office for National Statistics (ONS) definition of employment.[11] An individual is employed if they either did paid work last week, or if they did not do paid work last week but have a paid job.[12] An individual is also employed if they are on a government training scheme, or are doing unpaid work for a family business.

The second measures the weekly hours worked by an individual. To create this variable, we sum the regular weekly hours an individual works in a normal week and the number of overtime hours they work in a normal week. These weekly hours variables are not defined for the self-employed in the UKHLS. Therefore, our weekly hours variable takes the value 0 if an individual is neither in paid nor self-employment, a positive value if they are in paid employment, and missing values for the self-employed.[13] We later create two alternative measures of weekly hours worked. The first includes the self-employed (using a separate variable for the self-employed from the UKHLS), and the second also adds hours worked in a second job.[14]

---

redefine the variable *Boy* to take value 1 if any of the first born children is a boy. By twins here we mean any case where there is more than one oldest child of the same age. We do not observe birth dates of children in the UKHLS, and so we cannot distinguish twins from a case where a mother gives birth twice within the same year.

[10]The sample period is six waves long, but the first wave is a time period where all individuals have no children by definition of our sample construction. A child can be at most aged 7, but this is very rare.

[11]See www.ons.gov.uk.

[12]This definition includes the self-employed.

[13]In the full-sample, in 9.6% of observations men are self-employed, and in 4.5% of observations women are self-employed.

[14]Hours worked in a second job are reported per month, and so we first divide that number by four before adding to the existing measure of weekly hours.

2.2.2.2 Treatment: The Gender of the First Born Child

For both the variables $Child_{it}$ and Boy_i, we focus on first born biological children only. For the robustness checks that include twins, we redefine Boy_i to take the value 1 if at least one of the first born children is a son.

2.2.2.3 Summary Statistics

In Table 2.1, we present means and standard deviations of key variables. In every case, we use the first wave of the dataset for all individuals. Panel A demonstrates that the sample is predominantly urban, white and UK born. Panel B demonstrates that the sex ratio of the first born children is approximately balanced, with 49.3% boys and 50.7% girls. This is consistent with our identifying assumption that the gender of the first born child is random. We have 576 fathers and 744 mothers in the final sample of 1,320 parents.[15]

2.3 Results

2.3.1 Fathers Work Less After Having a First Born Son Rather than a Daughter

The results in column (1) of Table 2.2 demonstrate that fathers are on average 4.5 percentage points less likely to be employed after having a first born son rather than a daughter. This effect is statistically significant at the 5% level. The results in column (2) demonstrate that fathers work on average 3.1 fewer hours per week after having a first born son rather than a daughter, an effect that is statistically significant at the 1% level. There is a change in fathers' labour supply after having a boy but not after having a girl. In both columns (1) and (2), the coefficient β_2 is not statistically significant, but the sum of the coefficients $\beta_1 + \beta_2$ is, with p-values 0.003 and 0.000 from the associated

[15]The UKHLS is not well suited to tracking couples over time, only individuals. Often one partner is interviewed whilst the other is not. In addition, there is no couple ID variable, and the household ID changes in each wave. This is one reason that we run regressions for mothers and fathers separately.

Wald tests. We only interpret the estimate of β_1 causally, not the fertility effects.

By contrast, the results in columns (5) and (6) demonstrate that there is no gender of first child effect for mothers. For both employment and weekly hours, the point estimates of β_1 are small and statistically insignificant. However, we cannot claim those nulls are precise. For example, we can only rule out a decrease in weekly hours of 0.28 standard deviations or more,[16] or an increase by 0.18 standard deviations or more, after a first born son rather than a daughter (using the lower and upper bounds of the 95% confidence interval).[17] There is a fertility effect on mothers' labour supply. After having either a first born son or daughter, mothers are 16 percentage points less likely to be employed and work on average 12.7 fewer hours per week, both effects that are statistically significant at the 1% level.

The fertility effects for mothers provide an interesting benchmark which shows that the gender of first child effects for fathers are large. The gender of the first child has an effect on fathers' employment probability and weekly hours that is one quarter the size of the effect of having a child (of either gender) for mothers. The fertility effects we estimate for mothers are towards the higher end of existing estimates in the literature. Depending on the identification strategy used,[18] the literature suggests that an additional child reduces the probability that women are employed or participate in the labour force by in the region of ten percentage points (Angrist & Evans, 1998; Lundborg et al., 2014; Silles, 2016). If the fertility effects that we estimate are biased due to time varying unobservables, such that the true effects are smaller in magnitude, then the gender of first child effect for fathers would look even larger comparatively.

The effect on fathers' weekly hours estimated in column (2) is the combination of the effect on the extensive margin, estimated in column (1), and the effect on weekly

[16] For these calculations we use the within person variation in the dependent variable, which is the source of identifying variation in these fixed effects estimates.

[17] We can only rule out a decrease in the probability of employment after a first born son rather than a daughter of 0.28 standard deviations or more, and an increase of 0.17 standard deviations or more, using the lower and upper bounds of the 95% confidence interval.

[18] Empirical approaches to understanding fertility effects in the existing literature include instrumental variables approaches that use twins, the gender mix of the first two children, and IVF treatment success as instruments.

hours for men who are employed throughout the sample period: the intensive margin. We investigate the latter effect in columns (3) and (4). In column (4), we allow for a separate effect when the child is aged 0 from when the child is older than 0. To do so, we create two new variables: *Child is* 0, and *Child older than* 0. The former is a dummy variable taking the value 1 if you have ever had a child by that wave, and your first born child is aged 0; whilst the latter is a dummy variable taking the value 1 if you have ever had a child by that wave, and your first born child is older than 0. We interact each with the variable *Boy*. We find that, on the intensive margin, fathers work 2.2 fewer hours per week in the first year of a child's life after having a first born son rather than a daughter. There is no fertility effect on weekly hours in the first year after having a first born daughter, but there is a fertility effect in the first year after having a first born son, with a p-value of 0.000 from the associated Wald test. In columns (7) and (8), we demonstrate that there is also no gender of first child effect for mothers on the intensive margin.

The results from the event study specification described in equation (2.2) tell the same story. In Figure 2.1, we plot the point estimates and 95% confidence intervals for the effects on fathers' weekly hours in each year relative to the birth of the first child (the accompanying regression results are presented in Table A.1). In the first year of the child's life, fathers work 4.9 fewer hours per week after a first born son rather than a daughter, and there is no statistically significant effect in any other year. The aforementioned 3.1 hour per week effect estimated in column (2) of Table 2.2 is an average of the effects across the first and later years of the child's life.

We use the results from the event study to calculate the total effect on fathers' weekly hours in the first year of the child's life in Section 2.B.1. Fathers work 225 fewer hours after having a first born son rather than a daughter in the first year. This effect is large in the context of existing estimates in the literature. It is the same sign as the effect in Lundberg (2005), and within their range of estimates (100 to 300 hours). The effect is 2.3 times the size of the effect estimated in Choi et al. (2005), and 3.5 times the size of the effect estimated in Lundberg & Rose (2002), though the effects in those papers (for West Germany and the U.S. respectively) were in the opposite direction to the effects we find here.

2.3.2 Other Results

The gender of first child effect on the extensive margin for fathers, estimated in column (1) of Table 2.2, is in fact driven by effects on two separate extensive margins. In column (1) of Table A.2, we demonstrate that fathers who are initially employed are less likely to remain employed after having a first born son rather than a daughter. In column (2), we cannot rule out that fathers who are initially not employed are less likely to become employed after having a first born son rather than a daughter, given the point estimate for β_1 is large and negative, and has an associated p-value of 0.127.

We cannot conclude whether the effect on weekly hours is driven by a change in hours within fathers' existing jobs or a change in jobs themselves, as job changes are not directly measured in the UKHLS. The effects that we estimate for the best available proxies are not precise (see columns (1)-(3) of Table A.3).[19] The gender of first child effect on the intensive margin for fathers is predominantly driven by a reduction in regular hours rather than overtime hours. The effect on regular hours in the first year of the child's life, estimated in column (4) of Table A.3, is statistically significant, whilst the effect on overtime hours, in column (5), is not. Further, the effect on regular hours is double the size of the effect on overtime hours in terms of standard deviations of the dependent variable. There is even a small positive effect on overtime hours on the intensive margin when the child is older than zero, with men perhaps making up some of the lost work hours from the first year of the child's life.

[19]The UKHLS does not measure whether an individual changes jobs. We define three impulse dummy variables as proxies. The first takes the value 1 if an individual changed occupation in the present wave compared to the previous wave. The second takes the value 1 if an individual changed industry in the present wave compared to the previous wave. The third combines the first two and takes the value 1 if either of the first or the second impulse dummy variables takes the value 1.

2.3.3 Robustness of the Gender of First Child Effect for Fathers

2.3.3.1 Fathers' Attrition

We are concerned that the attrition of fathers from the sample could be related to the gender of the first child for unobservable reasons that are correlated with labour supply decisions. To test this, we directly ask whether the gender of the first child predicts the probability that the father leaves the sample. We create a new dependent variable, which is a dummy variable that takes the value 1 if the man is missing from the sample in that wave, and the value 0 if he is not.

We find that the attrition of fathers is not related to the gender of the first child. The coefficient on $Child \times Boy$ is statistically insignificant, with a p-value of 0.870, as demonstrated in column (1) of Table 2.3. Further, the point estimate suggests that a first born son reduces the probability that a father attrits from the sample by just 0.4 percentage points, which is small (fourteen percent of possible observations for the fathers are missing). The coefficient on $Child$ is statistically significant and negative, indicating that men are less likely to leave the sample after having a child. This effect might occur because childless men are less rooted geographically, and so more difficult to track over time.

2.3.3.2 Placebo Tests

The gender of the first child should not affect a father's labour supply decision before he knows the gender. In columns (2) and (3) of Table 2.3, we run placebo tests. We create an impulse dummy variable that takes the value 1 for the time period two waves before the birth of the first child. We interact this with the variable Boy, and add both to the baseline regression. If the interview one wave before the birth of the first child took place after the father's partner's second ultrasound scan, which is typically the scan where the gender of the child is identified, then the gender of first child could be known to the father and have anticipatory effects on his labour supply decisions in that wave.

It is for this reason that we look two waves before the birth, which is a time period when no father could have known the child's gender. We find no statistically significant gender of first child effect on either employment or weekly hours two waves before the birth of the first child.

To investigate this further, we consider the gender of first child effects two, three and four waves before the birth of the first child using the event study specification in Figure 2.1.[20] This Figure demonstrates the absence of any pre-birth trend in the gender of first child effect.

2.3.3.3 Other Robustness Tests

2.3.3.3.1 Adding the Sixth Wave of the Dataset Demonstrates Parameter Stability

In an earlier draft of this paper we could only use the first five waves of the UKHLS, as the sixth wave had not been publicly released. After five waves, we had a sample of 512 fathers. The sixth wave gave us an additional observation for the existing sample of fathers, and 64 men became new fathers. After adding the sixth wave and running the regressions again, we found that the estimates of the gender of first child effect on weekly hours, and on both the extensive and intensive margins, changed very little in size or statistical significance. This process demonstrates the stability of our parameter estimates, and the earlier results are available upon request.

2.3.3.3.2 Changes to the Specification, Sample and Variable Definitions

In columns (4) and (5) of Table 2.3, we drop the ethnic minority boost sample. In columns (6) and (7), we include twin first births, and in columns (8) and (9), we cluster standard errors at the individual level. In all cases, the gender of first child effects on both employment and weekly hours change little from the main results, and remain statistically significant at the 5% and 1% levels respectively.

[20] We plot the corresponding event study for women in Figure A.1.

In Table A.4, we control for the total number of children a man has ever had, we control for quarter rather than year fixed effects, we add wave fixed effects, we drop observations from the wave in which a father adds his second and additional children, and we drop outlying observations where the father reports working more than 75 hours in a week. In all cases the gender of first child effects on employment and weekly hours are robust. In Table A.5, we drop individual fixed effects and estimate the regression using pooled OLS models, as well as a Probit model in the case of employment, and a Tobit model in the case of weekly hours.[21] We use a trimmed least squares estimator to deal with censoring from below in an individual fixed effects model for weekly hours (Honore, 1992). We drop the use of sampling weights. Finally, we use two different definitions of weekly hours; the first including observations from self-employed men, and the second including hours worked in second jobs. In all cases the gender of first child effect remains statistically significant and very similar in size to the effects estimated in the main results. When we drop individual fixed effects, the coefficient on *Child* becomes positive and statistically significant, as we would expect if the time invariant unobservable characteristics that make men more employable also make them more likely to have children.

2.3.3.3.3 Balance Tests

Finally, in Table A.6, we present balance tests that demonstrate support for our identifying assumption. If the gender of the first child is random, conditional on the decision to have a child, then we expect the gender of the first child to be uncorrelated with demographic characteristics of fathers in the first wave of the dataset, before their first child is born. We present weighted means and the p-values for tests of differences in those weighted means. There is no statistically significant difference in the average age, weekly hours worked, probability of employment, monthly wages, ethnicity, education level (achieved a degree or at most GCSE qualifications), urban or rural residence or

[21] When we drop individual fixed effects, we add dummy variables that control for an individual's ethnic group. Specifically, we have dummies for *Black*, *Asian* and *Other Race*, relative to the White base category.

whether or not the man identifies with a religion, between the fathers who go on to have a son versus a daughter. Fathers who go on to have a son are less likely to initially be married or foreign born. However, initial marital status and place of birth are time invariant, and so we control for these imbalances in our regressions through the individual fixed effects. The effects that we estimate throughout the paper are identified using changes in behaviour within person, before and after the birth of the child, and so cannot be driven by these two imbalances.

The initial demographic characteristics of mothers are also balanced in general. A specific concern here is that child gender is affected by maternal energy intake prior to conception (Mathews et al., 2008; Mazumder & Seeskin, 2015). Employed mothers or mothers who work longer hours might be more likely to skip breakfast, and so have a daughter. However, mothers' initial employment probability and weekly hours worked are balanced across child gender.

2.3.4 Marriage and the Interpretation of the Fertility Effects

There may be pre-existing trends in labour supply before an individual has their first child. Marriage might influence both an individual's labour supply and fertility decisions. Whilst we do not attempt to identify causal effects of fertility, it is still interesting to understand whether the fertility effects we estimate are driven by changes in behaviour before and after marriage. On average, the married mothers and fathers in our sample have their first child soon after getting married, and so behavioural changes after marriage could be conflated with the effects of fertility. The modal lag between first marriage and first birth for both mothers and fathers is two years, as demonstrated in Figure A.2.

To investigate this, we use a new sample that consists of any individual who starts the sample period single and subsequently gets married.[22] We create a dummy variable

[22] This is a different sample to the sample that we use for the main results. We include any individual in the UKHLS who starts the sample period single, and gets married at some point (in a heterosexual marriage) during the sample period. We do not include observations where the individual has subsequently separated or divorced their spouse. We have a sample of 257 men and 332 women.

that takes the value 1 if an individual is married. We regress employment and weekly hours on the marriage dummy, and control for individual fixed effects, year fixed effects and the age of the individual as in the baseline regressions. We therefore identify the effects of marriage from changes in behaviour within person.

In columns (1)-(3) of Table A.7, we find no evidence that marriage affects the labour supply decisions of men. This is confirmed in the event study in Figure A.3. In columns (4)-(6) of Table A.7 and the left hand panel of Figure A.4, we find some evidence that women are less likely to be employed after getting married. The right hand panel of Figure A.5 demonstrates that this effect is driven by non-degree educated women. This suggests that part of the negative effect of fertility on labour supply that we estimate for women in columns (5) and (6) of Table 2.2 could be driven by marriage. However, the size of this effect on employment is only one third the size of the fertility effect.

2.4 Mechanisms: What Drives the Gender of First Child Effect on Fathers' Labour Supply?

In this Section, we discuss possible mechanisms for the effect of the gender of the first child on fathers' labour supply. In Table 2.4, we split the sample of men by whether they are degree educated.[23] In columns (1)-(3), we see that the effect on the extensive margin is driven primarily by degree educated fathers, whilst in columns (4)-(5), we see that the effect on the intensive margin is driven primarily by non-degree educated fathers. This difference suggests that the effect for each group might require a different explanation. This difference is true for a range of other indicators of socio-economic status too, such as marital status, partner's education, and own and partner's occupation and wages, as

[23]Amongst the 576 fathers, 41.3% are degree educated and 58.7% are non-degree educated.

shown in Tables A.8 and A.9.[24]

In Section 2.4.1, we show that the effect for degree educated fathers is consistent with the child production function mechanism, which is as follows. At a young age, boys have lower non-cognitive skills (e.g., Bertrand & Pan (2013)), perform worse in measures of educational attainment (Rutter et al., 2004; Ready et al., 2005), more often experience behavioural disorders such as ADHD and delayed speech, and are more disruptive (Szatmari et al., 1989; Halpern, 1997; Else-Quest et al., 2006; DiPrete & Jennings, 2012). These differences can emerge in infancy, even before a child reaches 12 months of age, and several authors have proposed biological explanations such as exposure to foetal testosterone (e.g., Lutchmaya et al. (2001, 2002)). Given this initial gender gap in child development, parents must devote more inputs to boys to achieve the same level of development as a girl,[25] which requires that they take more time off work.[26]

In Section 2.4.2, we show that the effect for non-degree educated fathers is consistent

[24]For the effect on employment, we test whether the difference is statistically significant. To do so, we run the following regressions:

$$\begin{aligned} Y_{it} = \theta_1(Child_{it} \times Boy_i) + \theta_2 Child_{it} + \lambda_1(H_i \times Child_{it} \times Boy_i) \\ + \lambda_2(H_i \times Child_{it}) + X_{it}'\varphi + year_{it}'\delta + \eta_i + \varepsilon_{it} \end{aligned} \tag{2.3}$$

Here, H_i, is a time-invariant dummy variable capturing the demographic characteristic of interest. We can then interpret θ_1 as the gender of first child effect for the group for whom $H_i = 0$, and the sum, $\theta_1 + \lambda_1$ as the gender of first child effect for the group for whom $H_i = 1$. Then, λ_1 tells us the difference in the gender of first child effect by that demographic characteristic. The differences are typically not statistically significant, which is likely due to low statistical power given the sub-samples of fathers are small once we split them by socio-economic status. For the case of partner's occupation, the difference is close to statistical significance, with a p-value of 0.110 associated with the coefficient on $Partner\ Non\text{-}Professional \times Child \times Boy$. We do not test the statistical significance of differences on the intensive margin, because the sub-samples are even smaller.

[25]To formalise things a little, we can think about a simple linear child development production function: $Development_i = \omega_i + \gamma Parent\ Inputs_i + u_i$. According to the child production function mechanism, ω_i is smaller for boys than girls, on average. A related literature demonstrates that boys' cognitive and non-cognitive development is more sensitive to parental inputs than girls (Bertrand & Pan, 2013; Autor et al., 2015; Fan et al., 2017), which suggests that γ is larger for boys than girls. However, these papers focus specifically on missing parental inputs from absent fathers and claim that boys are differentially sensitive because they lose a same-sex role model. In this paper, we consider two parent households too, and so it is not clear that γ should be larger for boys than girls. Indeed, we later find no evidence that γ is larger for boys than girls, in columns (4)-(6) of Tables 2.7 and A.11, and columns (5)-(8) of Table A.12 and (3)-(4) of Table A.13.

[26]A related channel to the child production function mechanism is excess male morbidity in early childhood. Fathers may take more time off work when their children are ill, and the medical literature has demonstrated that boys are more likely to have health problems in early childhood than girls (e.g., Wells (2000)).

with the new man mechanism. The new man mechanism is as follows. Men are starting to engage in childcare to reduce the burden on their partners and allow their partners to take less time off work after childbirth. However, men's first step towards engaging in childcare is to do so with sons. This may be due to same-sex parenting norms or simply because men feel more comfortable looking after a son.

We set these two explanations against three alternatives in Section 2.4.3. We show there is little evidence supporting the child cost mechanism in Section 2.4.3.1. We rule out that child birthweight drives the effects in Section 2.4.3.2, and show there is no evidence of pure son preference in Section 2.4.3.3.

2.4.1 The Effect for Degree Educated Fathers is Consistent with the Child Production Function Mechanism

2.4.1.1 Partners of Degree Educated Fathers Also Work Less After a First Born Son Rather Than a Daughter

To distinguish the child production function from the new man mechanism, we need to understand how the partners of the fathers respond to the birth of his first child. According to the child production function mechanism, a father takes time off work to increase the parental time received by his son. We do not expect to see an offsetting effect on his partner's labour supply. In fact, we might expect his partner to also take more time off work after a son rather than a daughter, so that total parental time received by the son increases further. By contrast, according to the new man mechanism, a father takes time off work to reduce the burden of childcare on his partner and enable his partner to take less time off work. Then we expect to see an offsetting effect on his partner's labour supply.

To test this, we create a sample of the men's partners. This is a different sample from the main sample of women in the baseline regressions.[27] For a childless man, we

[27]There are three reasons why these samples are different. The main sample of women additionally contains women who are not partnered, women who are partnered to men who are not interviewed in

match his partner in the first wave of the dataset. For a father, we match his partner in the wave before the birth of his first child. Of the 7,078 men, 2,392 initially have a partner, of whom 2,126 are interviewed in the UKHLS in that wave. This includes 474 of the 576 fathers. We use the event study specification described in equation (2.2) to estimate a gender of first child effect in each year relative to the year before the birth of the father's first child. We plot these effects on weekly hours for the partners of the degree educated fathers in the right hand panel of Figure 2.2, where the left hand panel plots the corresponding figure for the degree educated fathers themselves.[28] We present the regression results underlying the left and right hand panels of Figure 2.2 in columns (1) and (2) of Table 2.5 respectively. We can match partners to 202 degree educated fathers.

The partners of degree educated fathers work less after a first born son rather than a daughter, consistent with the child production function mechanism. When the child is aged 2 and 3, the partners of degree educated fathers work on average 4.7 and 6.4 fewer hours per week after a son rather than a daughter respectively, effects that are both statistically significant at the 10% level. The effect for the partners is more delayed than for the degree educated fathers themselves, for whom the effect is concentrated in the first year of the child's life. In Section 2.B.2, we aggregate the effects for the degree educated fathers and their partners. A first born son of a degree educated father could receive up to an additional 194 hours with his father in the first year of his life, and an additional 214 and 293 hours with his mother when aged 2 and 3, on average, relative to a first born daughter. This amounts to an additional 702 hours of parental time before the child's fourth birthday. In Figure A.6, we plot the effects for partners on the extensive margin, which demonstrate a similar pattern.

As a robustness check, we ask whether this sample of the partners of degree educated

the UKHLS, and women who are partnered to men who are interviewed in the UKHLS but who do not qualify for the final sample, as those men do not have their first born child during the sample period (the child is her first born, but not his).

[28] We plot the event studies for the extensive and intensive margin separately for the degree educated fathers in Figure A.12.

men is representative of degree educated women. There is a high rate of positive assortative mating in our sample, since seventy-four percent of the partners of fathers have the same degree status as him. We look at the gender of first child effect on weekly hours for the main sample of degree educated women in Figure A.7. We find that degree educated women also work less after a first born son rather than a daughter when the child is aged 3. This result is consistent with Ichino et al. (2014), who find that mothers in the UK work less after a first born son rather than a daughter, though they do not distinguish by mothers' education.

It is perhaps unsurprising that the child production function mechanism is consistent with the behaviour of degree and not non-degree educated parents.[29] Degree educated parents might be more aware of their child's development (e.g. Lundberg (2005) claims that as parents become better off, the desire for child quality increases), and so are more likely to learn that their son's development is lagging behind that of girls in his cohort. In addition, degree educated parents are more likely to have the financial security for both parents to take time off work.

To interpret this as a successful investment in sons' development, we need to check whether the reduction in hours worked led to a cut in expenditure on first born sons relative to daughters, which could damage their development. This is not necessary to understand whether the child production function mechanism drives the effect. Unfortunately, the UKHLS does not measure expenditure. However, it is plausible that the households of degree educated fathers are better able to shelter expenditure from the short-term income loss than the households of non-degree educated fathers, through greater savings and family financial support, or better paid paternity and maternity leave.[30]

[29]The existing literature suggests that highly educated parents invest more time in the development of their children, and the gap with less educated parents is increasing (Guryan et al., 2008; Phillips, 2011; Autor & Wasserman, 2013; Cunha et al., 2013; Lundberg et al., 2016).

[30]Child benefit payments also help shelter expenditure, and were £20.30 per week for most of the sample period.

2.4.1.2 Fathers Spend the Additional Time at Home Engaged with the Child

If the child production function mechanism drives the effect for degree educated fathers, then we expect these men to invest their additional time off work in the child. For the children in our sample, the UKHLS contains a single measure of parental time with the child. It asks how often each parent spends time with their child engaged in leisure activities or outings outside the home, such as going to the zoo, movies, sports or picnics. This question is asked of parents in the first, third and fifth waves of the dataset, and is answered on a six-point scale. We take the first observation for each parent after the birth of their first child.[31] We define two dummy variables. The first takes the value 1 if the parent engages in leisure activities with the child at least several times per week (the second highest response). The second takes the value 1 if the parent engages in leisure activities with the child almost every day (the highest response). By definition, we cannot identify effects from changes before and after having a child, and so we run cross-sectional regressions.

The results in columns (1) and (2) of Table 2.6 demonstrate that for the first definition, fathers are significantly more likely to engage in leisure with the child after a first born son rather than a daughter. This is consistent with evidence in the existing literature that fathers invest more time in childcare activities with sons than daughters (Yeung et al., 2001; Lundberg et al., 2007; Price, 2008; Mammen, 2011; Nicoletti & Tonei, 2015). In column (5), we demonstrate that this is also true for the degree educated fathers specifically.[32]

[31] This variable is non-missing for 449 out of 576 fathers.

[32] The effect is not statistically significant for the sub-samples of degree and non-degree educated fathers separately, likely due to low statistical power given the sample size. However, the size of the point estimates in both cases is very close to that for the full sample of fathers. In Table A.10, we demonstrate that there is no such effect in the full sample of mothers.

2.4.1.3 Time Engaged with Children is Positively Correlated with Development

As further supporting evidence, we ask whether parental time is positively correlated with child development. The UKHLS asks parents of three-year-old children about their child's cognitive, socio-emotional and physical development. For cognitive development, it contains 18 indicators; for socio-emotional development, 8 indicators; and for physical development, 4 indicators. Further details about these indicators are given in the Data Appendix table. These questions are asked from the third to the sixth wave. We run cross-sectional regressions, pooling observations across the third and fifth waves, the two waves in which both the parental time and child development questions are asked. We do not restrict the sample to parents who have their first born child during the sample period. We control for the gender of the child, wave fixed effects, the parent's education level and ethnic group.[33] For each area of child development, we take the first common factor across the indicators as the dependent variable, where a more positive value indicates a higher level of development.

In columns (1)-(3) of Table 2.7, we find that fathers' time engaged in leisure activities with the child is positively and statistically significantly correlated with both cognitive and socio-emotional development. A father engaging in leisure activities with his child at least several times per week is associated with a 0.24 and 0.41 standard deviation higher level of cognitive and socio-emotional development respectively. In Table A.11, we show that mothers' engagement in leisure activities with the child is positively and significantly correlated with cognitive, socio-emotional and physical development too.

We repeat this exercise using data from the second wave of the Millennium Cohort Study (MCS). The MCS contains more detailed information about parental time with children, and contains objective measures of children's cognitive development by giving them tests. Interviews were carried out five years before the start of our sample period, in 2004-5. Again, all children in the sample are aged three years old, and we do not

[33] We control for the parent's ethnic group because the child's ethnic group is not observable.

restrict the sample to first born children. In Table A.12, we look at children's cognitive development, as measured by their performance in a vocabulary test and a school readiness test. In Table A.13, we look at children's socio-emotional development, as measured by the first common factor from a factor analysis of thirty five indicators. We create two independent variables to capture parental engagement. Both are the first common factor from a factor analysis of a set of indicators of parental engagement in activities such as taking the child to the library, helping the child to learn the alphabet or to count. The two independent variables differ by the threshold we use to classify a high level of parental engagement, and more details are in the Data Appendix table. We control for the gender and ethnic group of the child and the parent's education level. In columns (1)-(4) of Table A.12, and in columns (1)-(2) of Table A.13, we find that again parental time engaged with the child is positively and statistically significantly correlated with the child's cognitive and socio-emotional development.

We allow for heterogeneity in the correlation between parental time and child development by the gender of the child and the parent. We do not find systematic evidence that the correlation between parental time and child development differs by child gender across columns (4)-(6) of Tables 2.7 and A.11, and columns (5)-(8) of Table A.12 and (3)-(4) of Table A.13. The pattern of heterogeneity by the parent's gender is mixed. In the UKHLS, the correlation between fathers' engagement and socio-emotional development is significantly greater than for mothers (column (8) of Table 2.7). However, in the MCS, the correlation between mothers' engagement and cognitive development is significantly greater than for fathers (columns (9)-(12) of Table A.12).

Finally, in Table A.14, we test the premise of the child production function mechanism that boys achieve lower levels of development than girls at young ages. There is indeed a statistically significant and substantial gender gap in children's cognitive and socio-emotional development in both the UKHLS and the MCS. Boys perform worse than girls, with the gender gap in the region of one fifth to one third of a standard deviation.

2.4.2 The Effect for Non-Degree Educated Fathers is Consistent with the New Man Mechanism

2.4.2.1 Partners of Non-Degree Educated Fathers Work More After a First Born Son Rather Than a Daughter

As described in Section 2.4.1.1, if the new man mechanism drives the gender of first child effect, then the partners of fathers will be more likely to work after a first born son rather than a daughter. A father takes time off work to reduce the burden of childcare on his partner, which facilitates his partner's earlier return to work. In Figure 2.3, we present event studies of the gender of first child effects on weekly hours for the sample of non-degree educated fathers and their partners. The regression results in columns (3) and (4) of Table 2.5 correspond to the left and right hand panels of Figure 2.3 respectively. We can match 272 of these partners.

The partners of non-degree educated fathers work on average 5 hours more per week after a first born son rather than a daughter when the child is aged 2, an effect that is statistically significant at the 10% level. In Figure A.8, we plot the effects on the extensive margin. The effect on the extensive margin when the child is aged 2 is statistically significant at the 5% level. The partners of non-degree educated fathers get back to work sooner after a first born son rather than a daughter on average. In absolute terms, in the year before the birth of the first child, 80% of these partners are employed for both the group who go on to have a boy or a girl. When the first child is aged 2, 82% with a first born boy are employed, compared to only 66% with a first born girl.

The effect for the partners of non-degree educated fathers is almost identical in size to the effect for the fathers themselves, but in the opposite direction. In the first year of the child's life, non-degree educated fathers work 5.1 fewer hours per week after a first born son rather than a daughter.[34] Therefore, there is almost no difference in total parental

[34] We plot the event studies for the extensive and intensive margin separately for the nondegree educated fathers in Figure A.13.

time at home after a first born son rather than a daughter on average, as demonstrated in Section 2.B.2. Across the first few years of the child's life, there is a substitution of father for mother inputs for first born sons relative to daughters, consistent with the new man mechanism.

However, the difference in the timing of the effect for the non-degree educated fathers and their partners is a reason to be cautious with the new man interpretation. The effect for the fathers occurs in the first year of the child's life, whilst the effect for the partners occurs when the child is aged 2. This lag could be partly explained by the process of job search, if some of the partners change jobs. Reducing the burden of childcare may free up time for the partner's job search in the first year, but the effect on her work hours occurs with a lag because the conclusion of the job search process takes time. Nonetheless, we would be more convinced about the new man mechanism if the effects for these fathers and their partners occurred simultaneously.

As a robustness check, we ask whether this sample of the partners of non-degree educated men is representative of non-degree educated women. We plot event study figures for the main sample of non-degree educated women. In the left hand panel of Figure A.9, we find that non-degree educated women are statistically significantly more likely to be employed after a first born son rather than a daughter at the 10% level when the child is aged 3.

2.4.2.2 Amongst Non-Degree Educated Fathers the Effect is Driven by those with More Progressive Attitudes Towards Women's Role in the Labour Market

To be consistent with the new man mechanism, the effect should be driven by fathers who believe that men should be involved in childcare and women should be able to return to work after having a child. We refer to this as a 'progressive' attitude about women's role in the labour market. We use the term progressive only insofar as it opposes a traditional attitude that women should look after the home whilst men work.

We allow for heterogeneity in the gender of first child effect by two indicators of the progressiveness of a man's attitude towards women's role in the labour market. The first indicator measures a man's source of news and current affairs. We use the Leveson Report to categorise news sources (Leveson, 2012). The Leveson Inquiry was a public inquiry into the cultures, practises and ethics of the British press, and its final report names five news sources that publish particularly damaging portrayals of women.[35] Readership of four of those sources is measured in the UKHLS. We create a dummy variable, *Non-Progressive News*, that takes the value 1 if a man receives news from either the newspaper or website version of The Sun, The Daily Mail, The Daily Express or The Daily Star. We claim that men who select into reading these sources have less-progressive attitudes towards women's role in the labour market than those who do not, ceteris paribus. We validate that claim statistically in Section 2.C. This question is asked in the third wave of the dataset only. We include all childless men, but restrict the sample of fathers to those who have their first child from the fourth wave onwards, to ensure that their choice of news source does not respond to the gender of the child. As such, this variable is defined for 230 out of 576 fathers,[36] of whom 24% read the less-progressive news source.

The second indicator measures a man's initial contribution to household chores. We create a measure of the fraction of six household chores that a man does mostly himself, *Fraction Chores*, which are: grocery shopping, cooking, cleaning, washing/ironing, gardening and DIY.[37] We claim that a man who initially does a higher fraction of household chores has a more progressive attitude towards women's role in the labour market versus

[35]The Sun, The Daily Mail, The Daily Express, The Daily Star and The Sport are named for their sexualisation and demeaning of women, and/or trivialisation of violence against women, and/or attitudes towards transgender individuals, in a chapter on the Representation of Women and Minorities.

[36]Only 239 out of 576 fathers have their first child in waves four, five or six and are interviewed in the UKHLS in wave 3, such that the news source indicator could be non-missing. Nine fathers do not respond to the question. We are therefore not concerned about non-random non-response to the news source question. Since both indicators of progressiveness are only defined for a sub-sample of fathers, we also first check that the gender of first child effects still exist in those sub-samples. The results in columns (3)-(4) and (7)-(8) of Table 2.8 demonstrate that the effects on both weekly hours and employment do still exist in both cases.

[37]Amongst the men in our sample who have a partner but no child by the second wave of the dataset, 15.6% buy groceries, 20.7% cook, 15.5% clean, 11.1% do the washing and ironing, 35.2% do the gardening and 68.7% do the DIY, 'mostly himself'.

the home. These questions are asked in the second, fourth and sixth wave of the dataset. We take the responses in the second wave for childless men, and for fathers, we take the responses in the most recent wave in which the questions were asked before the birth of their first child. These questions are only asked of men who have a partner, and we can only include fathers who have their first child from the third wave onwards. This variable is defined for 298 out of 576 fathers.[38]

In the combined sample of non-degree and degree educated men in Table 2.8, we find that the gender of first child effect is driven by fathers with more progressive attitudes towards women's role for both indicators of progressiveness, consistent with the new man mechanism. In both cases this heterogeneity is statistically significant for weekly hours. Fathers who receive news from a progressive source work on average 6.5 fewer hours per week after a first born son rather than a daughter, as demonstrated in column (2). The Wald test for the sum of the coefficients on $Child \times Boy$ and $Non\text{-}Progressive\ \ News \times Child \times Boy$ indicates that the gender of first child effect in fact operates in the opposite direction in absolute terms for fathers who get their news from a non-progressive source. These fathers work 5.4 more hours per week on average after a first born son rather than a daughter. Fathers who initially did one additional household chore out of six work 1.7 fewer hours per week after a first born son rather than a daughter, as demonstrated in column (6). This heterogeneity is also statistically significant on the extensive margin for the news source measure, as demonstrated in column (1).

In Table 2.9, we ask whether this heterogeneity exists for both the degree and non-degree educated fathers. In columns (1), (2), (4) and (5), we split the sample of men by initial degree status. We find that the heterogeneity by progressiveness is statistically significant for only the non-degree educated fathers for both indicators. In columns (3) and (6), we test whether the differences across degree and non-degree educated fathers are statistically significant. To do so, we interact each of the four key independent

[38]The missing values are as follows: 169 fathers have their first child in the second wave, 52 fathers do not have a partner in the wave before the birth of their first child (and so are not asked these questions), and 57 are not interviewed in the wave in which the questions are asked before the birth of their first child.

variables with the dummy variable for the degree status of the man. The coefficients of interest are the coefficients on *Degree* $\times$ *Non-Progressive News* $\times$ *Child* $\times$ *Boy* and *Degree* $\times$ *Fraction Chores* $\times$ *News* $\times$ *Child* $\times$ *Boy*. These tell us the difference in the difference in the gender of first child effect between progressive and non-progressive fathers, between the degree and non-degree educated. In both cases, the difference in difference is statistically significant, at the 10% level and 5% level respectively for the news source and fraction of household chores indicators. Again, it is the non-degree educated fathers for whom the evidence is consistent with the new man mechanism.

2.4.3 Alternative Explanations

2.4.3.1 The Child Cost Mechanism

The child cost mechanism is as follows. Some papers claim that daughters are more expensive to raise than sons (e.g., Dahl & Moretti (2008)). Whilst we are not aware of evidence for a difference in costs in early life, it is plausible that parents might expect daughters to be more financially dependent on them in later life due to the gender wage gap (e.g., Blau & Kahn (2000); Goldin (2014); Olivetti & Petrongolo (2016); Adda et al. (2017)). Parents therefore need to save more after having a daughter rather than a son, and so they might work more. To test this mechanism, we would ideally look at whether savings increase more after having a first born daughter rather than a son. However, we do not observe savings.[39] We therefore look for indirect evidence.

If this is the explanation for the differential costs by child gender, then the additional costs of raising a daughter fall long after she is born. Therefore, the incentives for parents to work more, to save more, shift permanently. In that case, we expect the gender of first child effect on fathers' labour supply to be persistent. However, in Figure 2.1, we saw that the gender of first child effect on weekly hours is concentrated in the first year of the

[39] Savings are measured in only three waves in the UKHLS, and so we do not have a sufficiently large sample size to compare changes in behaviour before and after birth between those with sons and daughters.

child's life only. In Figure 2.4, we look separately at the extensive and intensive margin. For both, we see that the effect is concentrated in the first year of the child's life. This is not consistent with the child cost mechanism.

Is the timing of the effects in Figures 2.1 and 2.4 real, or is it driven by compositional effects? Fathers have their first child in different waves across the sample period, and so the pattern in Figures 2.1 and 2.4 might arise because the only fathers for whom there is an effect are those who are only observed with their child at age 0. To check this, we restrict the sample to fathers who are observed at least once when the child is aged 0 and once when the child is older than 0, and estimate a separate gender of first child effect when the child is 0 and older than 0.[40] The results in columns (1), (2) and (3) of Table A.15 demonstrate that the effect on weekly hours, and the effects on the extensive and intensive margin separately, are all indeed transitory and occur in the first year of the child's life.

An alternative indirect approach to assess the child cost mechanism is to ask whether it is fathers with first born sons or fathers with first born daughters who change their behaviour before and after the birth of their first child. Either could drive the gender of first child effect. However, according to the child cost mechanism, the behavioural change should occur for fathers with daughters, who work more after the birth of the child. By contrast, according to the child production function and new man mechanisms, the behavioural change should occur for fathers with sons. In Figure A.10, we split the sample of fathers into those with sons and those with daughters, and present event studies for the effect of having a child on weekly hours.[41] For fathers with sons, there is clear evidence of a negative effect of having a child on weekly hours. There is no evidence of a pre-trend, and weekly hours decrease in the first year of the son's life. For fathers with daughters, the evidence is less clear. Whilst fathers work significantly more in the first year of the daughter's life, there is some evidence of a pre-birth trend in weekly hours.

[40] We include all childless men in these regressions too.

[41] The effects plotted in this figure are not causal. The figure plots the fertility effects separately for fathers with sons and daughters, and as previously discussed the decision to have a child is not random.

The point estimate in the first year of the daughter's life does not deviate much from that trend. We cannot rule out the possibility that some of the effect is driven by a behavioural change for fathers with daughters. Nonetheless, the evidence for the child cost mechanism is limited.

2.4.3.2 Child Birthweight

Boys are heavier than girls at birth (Pedersen, 1980; De Zegher et al., 1999; Voldner et al., 2009). In our sample, boys weigh on average 3.7 ounces more than girls at birth, a difference that is statistically significant at the 10% level.[42] Child birthweight might matter for parental labour supply through the physical depletion of the mother. The birth of a heavier child places a greater physical strain on the mother. It is also plausible that a heavier child requires more feeding immediately after birth, which again places a greater physical strain on the mother. If fathers take time off work to support their partners when their partners are physically exhausted from a high birthweight child, then the effect we have estimated might simply capture the effect of child birthweight, which is correlated with child gender. The timing of the effects on both the intensive and extensive margins in the first year of the child's life is consistent with the time at which the physical strain faced by the mother is greatest.

To test this mechanism, we directly control for the birthweight of the first child. We create a time invariant variable that measures the birthweight of the first child in ounces, and interact this with the variable *Child*. If child birthweight drives the gender of first child effect, then the coefficient on $Child \times Boy$ should become statistically insignificant and close to zero in absolute magnitude, whilst the coefficient on $Child \times Birthweight$ should be negative and significant. For the intensive margin regression, we create two interaction terms: $Child\ is\ 0 \times Birthweight$ and $Child\ older\ than\ 0 \times Birthweight$, and so allow for child birthweight to drive the effect in the first year after birth. For these tests to be valid, we require enough overlap in the distribution of birthweight for

[42] Of the 419 first children of the sample of fathers for whom we observe birthweight, 203 are boys and 216 girls. The girls weigh on average 118.0 ounces at birth, and the boys 121.7.

girls and boys, which we demonstrate is the case in Figure A.11.[43]

We observe the birthweight of the first child for 419 of the 576 fathers in the sample.[44] Therefore, in columns (1) to (3) of Table 2.10, we first demonstrate that the gender of first child effects still exist in this sub-sample. In columns (4) to (6), we then add the controls for birthweight. Doing so has almost no effect on the size or statistical significance of the gender of first child effects on employment, weekly hours, or weekly hours on the intensive margin, and the coefficient on $Child \times Birthweight$ itself is always small and statistically insignificant. Child birthweight does not drive the gender of first child effects on either the extensive or intensive margins.

2.4.3.3 Pure Son Preference

We next ask whether pure son preference drives the effects. By pure son preference, we mean that fathers derive higher utility from spending time with a son rather than a daughter, for whatever reason. We test this in three ways. In the first two tests, presented in Section 2.4.3.3.1, we ask whether having a first born son affects family structure; specifically, subsequent fertility or the relationship status of parents. This is the standard approach to test for the existence of son preference used in the existing literature (Baker & Milligan, 2013). In the third test, we ask which group of fathers by ethnic group drives the effect, as explained in Section 2.4.3.3.2.

2.4.3.3.1 The Gender of the First Child and Family Structure

If a father prefers spending time with sons then he is less likely to have an additional child, or birth spacing will be longer, after having a first born son rather than a daughter, because he already has the child that he wants. Further, the value of marriage is greater to him if his first born child is a son rather than a daughter, since in the event of a break

[43]If the distributions did not overlap, we would be unable to distinguish the effect of child birthweight from child gender.

[44]Only women are asked for the birthweight of their newborn children, and so we have to match the information from the man's partner. If she was not interviewed in the wave immediately following the birth of this child, then this information is missing. Further, not all mothers provide a birthweight of the child.

up or divorce the mother typically has custody rights to the children. We therefore expect that a man with son preference is more likely to be married after a first born son rather than a daughter (Dahl & Moretti, 2008).

In columns (1) and (2) of Table 2.11, we look at the effect of the gender of the first child on subsequent fertility, for men and women respectively. The dependent variable is an impulse dummy variable taking the value 1 if an individual has an additional child in that wave (after the first child). The coefficient on $Child \times Boy$ in column (1) is statistically insignificant (p-value 0.448) and small. Using the lower bound of the 95% confidence interval, we can rule out a decrease in the probability that a man has an additional child by 2.3 percentage points or more as a consequence of having a first born son rather than a daughter. The same is true for mothers, as demonstrated in column (2).

In column (3), we test whether having a first born son rather than a daughter increases the probability that a father is married.[45] The coefficient on $Child \times Boy$ is positive, but small and not statistically significant, with p-value 0.913. However, at a 95% significance level, we cannot rule out a 4.5% increase in the probability of being married, which is not a small effect. A similar pattern is true for mothers, in column (4). The coefficient is again positive but not statistically significant, and we cannot rule out a 4.2% increase in the probability of being married. The effects on marital status are not precise nulls.

In column (5), we restrict the sample to women who are initially not single,[46] and ask whether having a first born son reduces the probability that they become single.[47] The dependent variable is a dummy variable taking value 1 if the woman is single. The

[45] This combines the effect of the gender of first child on the probability that an initially not married father becomes married, and that an initially married father remains married. However, in our sample just one father (and seven mothers) gets divorced after having their first child. By contrast, thirty-six fathers (and thirty-five mothers) get married after having their first child. So the regressions in columns (3) and (4) are really providing information on whether the gender of first child affects the probability that an initially unmarried parent gets married.

[46] We make this sample restriction because we are interested in whether the father of the child stays with the mother, not whether the gender of the child affects the probability that another man starts a relationship with the mother.

[47] We do not investigate the probability that fathers become single, as so few fathers in our dataset are single. For just 0.2% of observations of a father with a child, that father is single, compared to 15.4% of observations of mothers with a child.

coefficient on $Child \times Boy$ is negative, but statistically insignificant. Further, the point estimate is less than one tenth the size of the coefficient on $Child$, which is statistically significant and suggests that women are less likely to become single after they have had a child of either gender.

One caution with these tests is that we only observe parents for at most five time periods with their child. We therefore cannot know if the gender of first child affects subsequent relationship or fertility decisions beyond those immediate few years after birth. Another caution is that these are tests for quite an extreme level of son preference. A father might prefer spending time with a son, just not to the extent that it affects his subsequent relationship or fertility decisions. We therefore present a third test of pure son preference in Section 2.4.3.3.2.

2.4.3.3.2 The Effects are Driven by White Fathers

If the gender of first child effect is concentrated amongst Asian fathers, we might be more willing to believe that pure son preference drives the effect, given the claim that son preference is particularly pronounced amongst Asian parents, even after they emigrate (e.g. Almond et al. (2009)). Given our sample size, we separate out all ethnic minority fathers from White fathers in columns (6) to (9) of Table 2.11.[48] We find that the effects on employment and weekly hours are driven by the sub-sample of White fathers, which again does not suggest that pure son preference explains the results. For non-White fathers, the coefficient on $Child \times Boy$ is small, statistically insignificant and even positive, for both employment and weekly hours.

2.5 Conclusion

We investigated the effect of the gender of the first child on parental labour supply in the UK in the short-term. We find that fathers are on average 4.5 percentage points less

[48] In our sample, 26.9% of men are non-White. This fraction is high because the UKHLS systematically oversamples ethnic minorities using an Ethnic Minority Boost sample.

likely to be in employment after a first born son rather than a daughter. On the intensive margin, fathers work on average 2.2 fewer hours per week in the first year of the child's life. The gender of first child effect for fathers is approximately one quarter the size of the effect of having a child of either gender for mothers.

Degree educated fathers take more time off work after a first born son rather than a daughter to invest more parental inputs in the son. Their partners also take more time off work, which may increase total parental inputs further. This is consistent with the child production function mechanism, according to which parents devote more inputs to sons than daughters because boys' initial cognitive and non-cognitive development is less advanced than that of girls. Non-degree educated fathers also take more time off work after a first born son rather than a daughter. Consistent with the new man mechanism, their partners work more after a first born son rather than a daughter, and the effect is driven by the fathers with more progressive attitudes towards women's role in the labour market. However, the difference in the timing of the effects for these fathers and their partners is a reason for caution with the new man interpretation.

Combining the effects for the fathers and their partners, we find that on average a first born son of a degree educated father could receive up to 702 additional hours of parental time before his fourth birthday than a first born daughter, depending on how much of the additional time parents take off work is spent engaged with their child. By contrast, on average a first born son of a non-degree educated father receives almost no difference in parental time compared to a first born daughter.

Our results have implications for an emerging literature on the gradient in the gender gap in educational attainment by socio-economic status. In recent decades, the gender gap in graduation from secondary and tertiary education has reversed, with girls now outperforming boys (Goldin et al., 2006; DiPrete & Jennings, 2012). This reversal has been most pronounced amongst low socio-economic status households (Autor & Wasserman, 2013; Autor et al., 2015). Explanations typically focus on low parental inputs in low socio-economic status households, arguing that boys' development is particularly sensitive to absentee fathers (e.g. Bertrand & Pan (2013)).[49] The results in our paper suggest that the behaviour of high socio-economic status parents may also contribute to the explanation. In households with degree educated fathers, total parental time off work is greater after a first born son rather than a daughter. It is in these households that boys are less likely to fall behind relative to girls at a young age, which could have implications for later educational attainment.

As future rounds of the UKHLS become available, we can test whether the differences in parental time received at a young age have an effect on later childhood outcomes for boys versus girls. We can also understand whether there are any long-term effects on parents' career paths. In addition, once we observe multiple rounds of the data since April 2015, we can evaluate whether parents' choices between childcare and work have been affected by the introduction of Shared Parental Leave.

[49] Sixteen percent of children born to the mothers in our sample of women are born to single mothers. Looking by socio-economic status, 26% of children born to non-degree educated mothers are born to single mothers, compared to just 3% of children born to degree educated mothers.

2.6 Tables and Figures

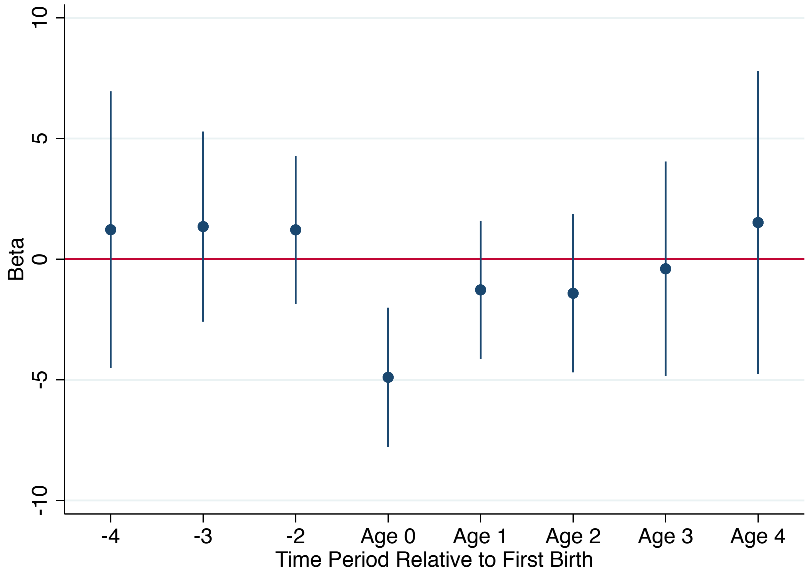


Figure 2.1: *Notes.* **Gender of First Child Effect on Weekly Hours for Men**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on weekly hours. We create two sets of impulse dummy variables. The dummy variables in the first set each take the value 1 in a single time period before the birth of the first child. The dummy variables in the second set each take the value 1 for a single age of the child. In combination, we therefore have an indicator for each time period relative to the birth of the first child. We interact each of these with the variable Boy, such that we estimate a gender of first child effect in each wave relative to the birth of the first child. The base year is the wave before the birth of the first child. For the regression underlying the figure, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the man's age, and we cluster standard errors at the level of the primary sampling unit. Fathers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

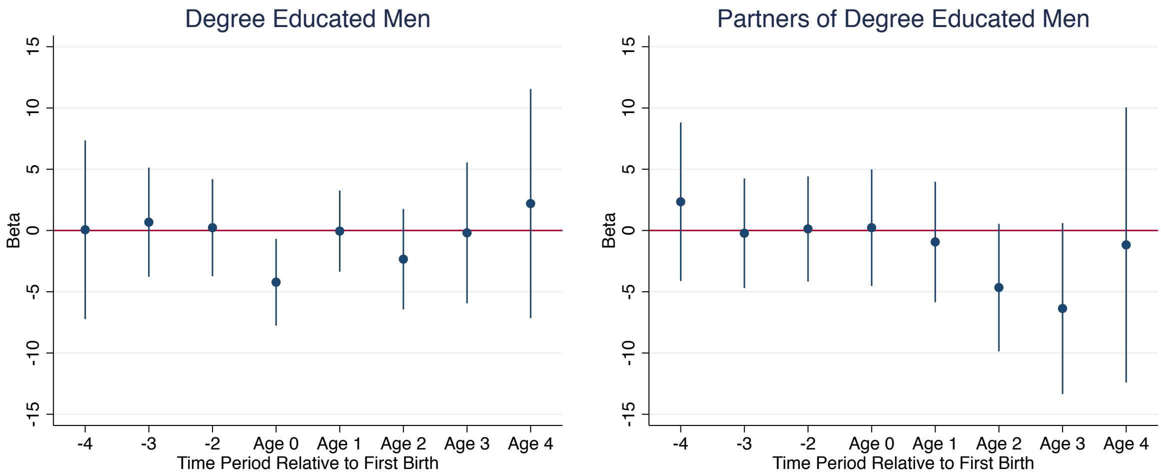


Figure 2.2: *Notes.* **Gender of First Child Effect on Weekly Hours for Degree Educated Men and their Partners**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on weekly hours. The base year is the wave before the birth of the first child. The left hand panel plots estimates for the weekly hours worked by degree educated men, and the right hand panel for weekly hours worked by their partners. Of the 576 fathers in the full sample, 505 have a partner in the wave before the birth of their first child, of whom 474 are interviewed in the UKHLS in that wave and so can be matched. Of those, 202 are partners of degree educated fathers. For the regressions underlying both figures, we weight observations using the a_indinus_xw weight, we control for individual fixed effects and year fixed effects, and we cluster standard errors at the level of the primary sampling unit. For the regressions underlying the figure in the left hand panel, we control for the man's age, and in the right hand panel, we control for the partner's age. Fathers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

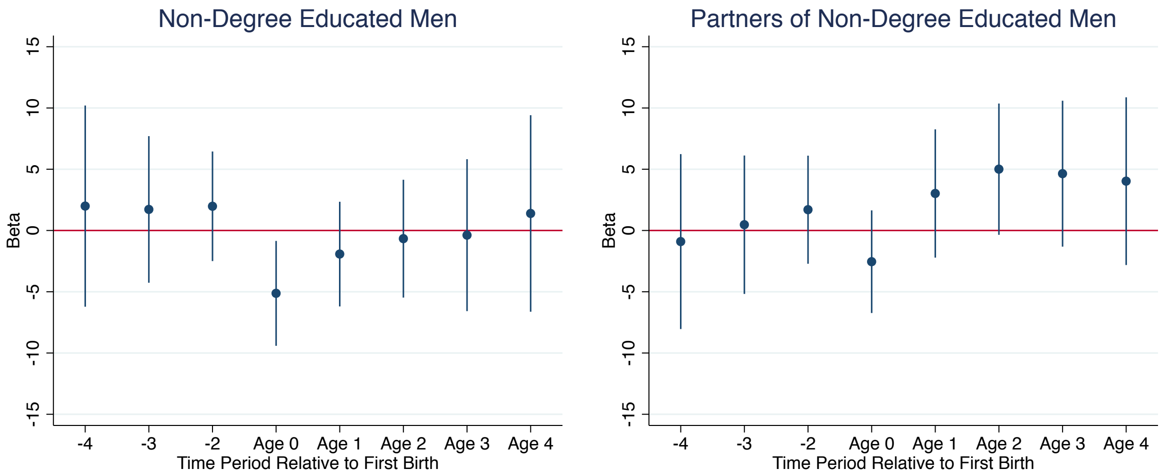


Figure 2.3: *Notes.* **Gender of First Child Effect on Weekly Hours for Non-Degree Educated Men and their Partners**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on weekly hours. The base year is the wave before the birth of the first child. The left hand panel plots estimates for the weekly hours worked by non-degree educated men, and the right hand panel for weekly hours worked by their partners. Of the 576 fathers in the full sample, 505 have a partner in the wave before the birth of their first child, of whom 474 are interviewed in the UKHLS in that wave and so can be matched. Of those, 272 are partners of non-degree educated fathers. For the regressions underlying both figures, we weight observations using the a_indinus_xw weight, we control for individual fixed effects and year fixed effects, and we cluster standard errors at the level of the primary sampling unit. For the regressions underlying the figure in the left hand panel, we control for the man's age, and in the right hand panel, we control for the partner's age. Fathers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

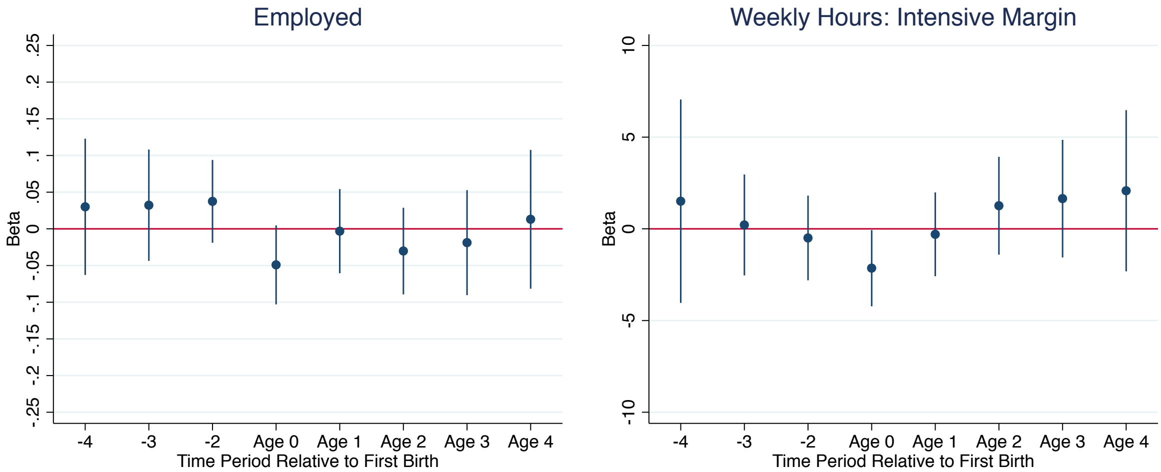


Figure 2.4: *Notes.* **Gender of First Child Effect on the Extensive and Intensive Margin for Men**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on the extensive and intensive margins separately. The base year is the wave before the birth of the first child. For the regressions underlying both figures, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the man's age, and we cluster standard errors at the level of the primary sampling unit. The left hand panel plots the results for the extensive margin, where the dependent variable is the employment dummy. The right hand panel plots the results for the intensive margin, where the dependent variable is weekly hours. Fathers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

Table 2.1: Summary Statistics

Panel A.			
	Men	Women	Total
Age	31.545	30.447	31.021
	(13.026)	(13.012)	(13.030)
Married	0.188	0.210	0.198
	(0.391)	(0.407)	(0.399)
Weekly Hours	23.397	22.336	22.878
	(21.677)	(20.063)	(20.909)
Employed	0.643	0.652	0.647
	(0.479)	(0.476)	(0.478)
Religious	0.445	0.491	0.467
	(0.497)	(0.500)	(0.499)
Urban	0.834	0.831	0.833
	(0.372)	(0.375)	(0.373)
Foreign Born	0.209	0.184	0.197
	(0.407)	(0.387)	(0.398)
Ethnic Minority	0.269	0.245	0.258
	(0.444)	(0.430)	(0.437)
Degree	0.281	0.313	0.296
	(0.449)	(0.464)	(0.457)
GCSE only	0.362	0.319	0.341
	(0.481)	(0.466)	(0.474)
Observations	7,078	6,462	13,540

Panel B.			
	Gender of Child		
	Son	Daughter	Total
Proportion	0.493	0.507	1
Total	651	669	1,320
	Gender of Parent		
	Father	Mother	Total
Proportion	0.436	0.564	1
Total	576	744	1,320

Notes. In Panel A, we present unweighted means and standard deviations for a set of demographic characteristics for the men and women in the sample. Standard deviations are presented in parentheses. We consider only the men and women who have a non-missing and non-zero value of the sampling weight a_indinus_xw, as these are the men and women who make up the sample for almost all of the regression analysis. We calculate means and standard deviations using observations from the first wave of the dataset. In Panel B, we present the gender of the children and parents in the sample.

Table 2.2: Gender of First Child and Parents' Labour Supply

	Men				Women			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Employed	Weekly Hours	Weekly Hours Intensive Margin	Weekly Hours Intensive Margin	Employed	Weekly Hours	Weekly Hours Intensive Margin	Weekly Hours Intensive Margin
Child x Boy	**-0.045****	**-3.13*****	-0.50		-0.014	-0.52	0.62	
	(0.021)	**(1.12)**	(0.74)		(0.028)	(1.13)	(1.05)	
Child	-0.0055	-0.51	-1.10**		-0.16***	-12.7***	-9.32***	
	(0.015)	(0.83)	(0.53)		(0.018)	(0.76)	(0.68)	
Child is 0 x Boy				**-2.16****				0.32
				(0.90)				(1.22)
Child is 0				-0.073				-5.86***
				(0.69)				(0.75)
Child older than 0 x Boy				0.70				0.17
				(0.95)				(1.23)
Child older than 0				-1.80**				-12.2***
				(0.70)				(0.85)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,353	21,830	11,604	11,604	23,711	22,452	11,815	11,815
Individuals	7,078	6,625	3,544	3,544	6,464	6,285	3,318	3,318
Mean of outcome	0.71	26.8	40.9	40.9	0.71	24.6	37.0	37.0
p-value Wald Test: Child + (Child x Boy)	0.003	0.000	0.005		0.000	0.000	0.000	
p-value Wald Test: Child is 0 + (Child is 0 x Boy)				0.000				0.000

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the main results of the paper, and asks what is the effect of the gender of the first born child on parents' labour supply. Columns (1) - (4) look at changes in men's labour supply, whilst columns (5) to (8) look at changes in women's labour supply. The dependent variable in columns (1) and (5) is a dummy variable which takes the value 1 if the individual is employed and the value 0 if not. The dependent variable in columns (2), (3), (4), (6), (7) and (8) is weekly hours worked, which has missing values for the self-employed. Columns (1) and (2) focus on the full sample of men, columns (5) and (6) on the full sample of women. The regressions in columns (3) and (4) include only men who are employed in every wave in which they are observed in the dataset: the intensive margin. Columns (7) and (8) look at the intensive margin for women. We allow for a difference in the gender of child effect when the child is aged 0 and when the child is older than 0 on the intensive margin in columns (4) and (8). Observations in all regressions are weighted by the a_indinus_xw weight. In all regressions, we control for individual fixed effects, year fixed effects and the age of the individual. In columns (1), (2), (3), (5), (6), (7), the coefficient on Child x Boy tells us the effect of having a son rather than a daughter, conditional on having a child. In columns (4) and (8), the coefficient on Child is 0 x Boy tells us the gender of first child effect when the child is 0, and the coefficient on Child older than 0 x Boy when the child is older than 0.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3: Robustness of Gender of First Child Effect on Fathers

	Attrition	Placebo Tests		Drop EMB		Include Twins		Cluster Individual Level	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Attrited	Employed	Weekly Hours	Employed	Weekly Hours	Employed	Weekly Hours	Employed	Weekly Hours
Child x Boy	**-0.0041**	-0.038*	-2.84**	**-0.048****	**-3.28*****	**-0.045****	**-3.20*****	**-0.045****	**-3.13*****
	(0.025)	(0.023)	(1.17)	**(0.022)**	**(1.16)**	**(0.021)**	**(1.11)**	**(0.021)**	**(1.12)**
Child	-0.25***	-0.011	-0.45	-0.0032	-0.50	-0.0043	-0.45	-0.0055	-0.51
	(0.017)	(0.015)	(0.86)	(0.015)	(0.85)	(0.015)	(0.83)	(0.015)	(0.82)
Two waves before birth x Boy		**0.026**	**0.97**						
		(0.026)	**(1.26)**						
Two waves before birth		-0.019	0.23						
		(0.018)	(0.86)						
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	42,462	24,353	21,830	20,849	18,595	24,443	21,908	24,353	21,830
Individuals	7,077	7,078	6,625	5,855	5,457	7,096	6,641	7,078	6,625
Mean of outcome	0.43	0.71	26.8	0.73	27.9	0.71	26.9	0.71	26.8

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level in columns (1)-(7). Standard errors are clustered at the individual level in columns (8) and (9). Observations in all regressions are weighted by the a_indinus_xw weight. In this Table, we perform robustness tests for the main results for men in Table 2.2. In column (1), we test whether the gender of the first child affects the probability that the father leaves the UKHLS sample. The dependent variable in this regression is a dummy variable that takes value 1 if the man is not present in the sample in that wave. In columns (2) and (3), we run placebo tests. The gender of first child should not affect labour supply decisions before that gender is known (before the child is born). We allow for a separate gender of first child effect two waves before a child is born (a time period when no father could have known the gender of his first child), by defining an impulse dummy variable for the time period two waves before the birth of his first child. In columns (4) and (5), we drop the ethnic minority boost sample. In columns (6) and (7), we include twin births or other cases where a parent has multiple first born children of the same age. To do so, we redefine the variable Boy to take value 1 if at least one of the first born children was a son. In columns (8) and (9), we cluster standard errors at the individual level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.4: Heterogeneity of Gender of First Child Effect on Fathers by Own Degree Status

	Employed			Weekly Hours Intensive Margin	
	(1)	(2)	(3)	(4)	(5)
	Degree	No Degree	Sig Diff?	Degree	No Degree
Child x Boy	**-0.064****	-0.028	-0.028		
	(0.025)	(0.031)	(0.031)		
Child	0.0043	-0.0072	0.00099		
	(0.020)	(0.020)	(0.020)		
Degree x Child x Boy			-0.040		
			(0.040)		
Degree x Child			-0.016		
			(0.027)		
Child is 0 x Boy				-0.89	**-3.17****
				(1.05)	**(1.40)**
Child is 0				-0.53	0.26
				(0.86)	(1.03)
Child older than 0 x Boy				**2.09****	-0.41
				(1.06)	(1.50)
Child older than 0				-2.69***	-1.03
				(0.79)	(1.10)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes
Observations	6,980	17,356	24,336	4,321	7,276
Individuals	1,989	5,083	7,072	1,295	2,247
Mean of outcome	0.85	0.66	0.71	42.1	40.3

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. Observations are weighted using the a_indinus_xw weight in all regressions. In this Table, we present results demonstrating the heterogeneity of the gender of first child effect on men's labour supply by whether or not the man has a degree. In columns (1)-(3), the dependent variable is the employment dummy variable, and in columns (4) and (5), weekly hours worked. In columns (4) and (5), we restrict the sample to men who are employed in every wave in which they are observed in the sample: the intensive margin. In columns (1) and (4), we restrict the sample to men who initially have a degree, and in columns (2) and (5), to men who initially do not have a degree. By initially, we mean the first wave of the sample period for those who never have a child, and the wave before the birth of the first child for those who do.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.5: Event Studies for Men and their Partners by Men's Degree Status

	Degree		Non-Degree	
	(1)	(2)	(3)	(4)
	Men	Partners	Men	Partners
Child is 0 x Boy	**-4.22****	0.23	**-5.13****	-2.55
	(1.80)	(2.42)	**(2.18)**	(2.13)
Child is 0	0.85	-7.42***	1.73	-5.70***
	(1.42)	(1.81)	(1.26)	(1.34)
Child is 1 x Boy	-0.049	-0.94	-1.93	3.02
	(1.69)	(2.50)	(2.18)	(2.67)
Child is 1	-1.97	-12.0***	1.09	-11.7***
	(1.48)	(1.84)	(1.55)	(1.80)
Child is 2 x Boy	-2.34	**-4.66***	-0.67	**5.01***
	(2.09)	**(2.65)**	(2.45)	**(2.73)**
Child is 2	-1.05	-10.1***	0.79	-11.0***
	(1.72)	(1.83)	(1.63)	(1.79)
Child is 3 x Boy	-0.19	**-6.37***	-0.38	4.64
	(2.93)	**(3.55)**	(3.16)	(3.03)
Child is 3	-0.64	-8.82***	-1.93	-12.1***
	(2.07)	(2.25)	(2.17)	(2.43)
Child is 4 x Boy	2.20	-1.18	1.39	4.03
	(4.77)	(5.72)	(4.09)	(3.49)
Child is 4	-6.82*	-13.9***	-3.94	-9.80***
	(4.13)	(4.13)	(3.08)	(2.27)
Year Fixed Effects	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes
Observations	6,071	2,656	15,680	4,436
Individuals	1,849	714	4,770	1,274
Mean of outcome	34.3	31.1	23.9	26.7

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the results from the event studies for weekly hours for men and their partners, by degree status of the men. We create an impulse dummy variable for the time periods four, three and two waves before the birth of the first child, and when the child is aged 0, 1, 2, 3 and 4. We interact each with the variable Boy. We therefore estimate a separate gender of first child effect in each of the first five years of the child's life. We drop observations from five waves before the birth of the first child as well as from when the child is aged 5 and older (there are very few such observations, and as such we cannot estimate effects in those years). The gender of first child effects are estimated relative to a base year, which is the year before the birth of the first child. In all regressions, the dependent variable is weekly hours. In columns (1) and (2), we present results from the regressions underlying the left and right hand panels respectively of Figure 2.2, where we use the sample of degree educated men and their partners. In columns (3) and (4), we present results from the regressions underlying the left and right hand panels respectively of Figure 2.3, where we use the sample of non-degree educated men and their partners. Observations in all regressions are weighted using the a_indinus_xw weight, and we control for individual fixed effects, year fixed effects and the age of the individual (or the partner's age in columns (2) and (4)).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: Fathers' Leisure Time with Child

	All Fathers				Degree	Non-Degree
	Several Times Per Week		Almost Every Day		Several Times Per Week	
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Probit	OLS	Probit	OLS	OLS
Boy	**0.086***	**0.22***	0.049	0.22	**0.089**	**0.084**
	(0.052)	**(0.13)**	(0.038)	(0.16)	**(0.086)**	**(0.066)**
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	449	449	449	449	185	264
Mean of outcome	0.50	0.50	0.17	0.17	0.48	0.52

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the results from regressions which ask whether the gender of the first child affects the time that fathers spend engaged in leisure activities with that child. The UKHLS asks how often you and your child spend time together on leisure activities or outings outside the home, such as going to the park or the zoo, movies, sports or picnics. This question is asked in waves 1, 3 and 5 of adults with children. By definition, we cannot look at changes in behaviour before and after having a child, and so we now run cross-sectional regressions on the sample of fathers (dropping childless adults, for whom this question is not asked). We take the observation in the earliest wave after the birth of the first child for each individual. This variable is non-missing for 449 out of 576 fathers. Answers are given on a six point scale, from which we create two dummy variables. The first, used in columns (1), (2), (5) and (6), is a dummy variable that takes the value 1 if the parent spends leisure time with their child at least several times per week (the second highest category on the six-point scale). The second, used in columns (3) and (4) takes the value 1 if the parent spends leisure time with their child almost every day (the highest category on the six-point scale). Observations in all regressions are weighted by the a_indinus_xw weight. In all regressions, we control for year fixed effects and the age of the father.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.7: Parents' Leisure Time and Child Development: UKHLS

	Father Leisure Time			Difference by Child Gender?			Difference by Parent Gender?		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Cognitive	Socio-Emotional	Physical	Cognitive	Socio-Emotional	Physical	Cognitive	Socio-Emotional	Physical
Father Leisure	**0.23*****	**0.32*****	0.026	0.19**	0.29***	0.073	**0.16****	**0.30*****	0.016
	(0.069)	**(0.065)**	(0.061)	(0.081)	(0.089)	(0.078)	**(0.069)**	**(0.068)**	(0.062)
Boy	-0.36***	-0.21***	-0.090	-0.39***	-0.23**	-0.055	-0.37***	-0.20***	-0.097*
	(0.075)	(0.068)	(0.057)	(0.11)	(0.092)	(0.074)	(0.074)	(0.069)	(0.059)
Father Leisure x Boy				0.079	0.075	-0.098			
				(0.15)	(0.13)	(0.12)			
Mother Leisure							**0.16****	0.054	0.038
							(0.081)	(0.072)	(0.065)
Wave Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent Education Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent Ethnic Group Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	548	558	558	548	558	558	547	557	557
Mean of outcome	-0.037	0.0090	-0.031	-0.037	0.0090	-0.031	-0.038	0.0087	-0.030
Standard Deviation of outcome	0.95	0.79	0.69	0.95	0.79	0.69	0.95	0.79	0.69
p-value Wald Test: Father Leisure = Mother Leisure							0.99	**0.023**	0.82

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the results from OLS regressions that ask whether parents' leisure time is correlated with child development. We look at three areas of child development. In columns (1), (4), and (7), we look at cognitive skills; in columns (2), (5), and (8), we look at socio-emotional skills, and in columns (3), (6), and (9), we look at physical skills. The dependent variable in each case is the first common factor from a factor analysis across a range of indicators of the given area of child development. For every indicator underlying these three common factors, parents are asked to judge their child's development, and the questions are asked only of parents with a child aged 3 years old. These regressions are not restricted to first born children only. The common factor for cognitive skills is based on 18 indicators, the common factor for socio-emotional skills is based on 8 indicators, and the common factor for physical skills is based on 4 indicators. The key independent variable is a dummy variable that takes the value 1 if the parent engages in leisure activity with the child at least several times per week. The dependent variables are measure in waves 3 to 6 of the UKHLS, whilst parent leisure time is measured in waves 3 and 5, and so we pool observations from waves 3 and 5. In columns (4)-(6), we allow for a difference in the correlation between fathers' time and child development across boys and girls. In columns (7)-(9), we compare the correlation between fathers' and mothers' leisure time and child development. In each regression we control for wave fixed effects and the gender of the child. In columns (1)-(6) we control for the ethnic group of the father (dummies for Black, Asian and Other race, relative to white base group) and his education level (dummies for degree, other higher, A-level, GCSE, other qualification, relative to the base group of no qualification), and in columns (7)-(9) we control for the ethnic group and education level of both the father and mother. Observations in all regressions are weighted using the a_indscus_xw weight, as answers are taken from the self-completion questionnaire. In columns (7)-(9), we present Wald tests for a difference in the coefficient on father and mother leisure time.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: Gender of First Child Effect on Father Labour Supply: Heterogeneity by Progressiveness

	News Source				Contribution to Household Chores			
	Allow for Heterogeneity		Still Effects in Sub-sample?		Allow for Heterogeneity		Still Effects in Sub-sample?	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Employed	Weekly Hours	Employed	Weekly Hours	Employed	Weekly Hours	Employed	Weekly Hours
Child x Boy	**-0.13*****	**-6.53*****	**-0.079****	**-3.84****	-0.029	0.46	**-0.038***	**-2.67****
	(0.036)	**(2.03)**	**(0.031)**	**(1.71)**	(0.043)	(2.11)	**(0.023)**	**(1.25)**
Child	0.0034	-0.63	-0.0086	-1.20	0.053*	0.60	0.050***	2.16**
	(0.022)	(1.48)	(0.018)	(1.23)	(0.027)	(1.44)	(0.016)	(0.95)
Non-Progressive News x Child x Boy	**0.23*****	**11.9*****						
	(0.060)	**(3.21)**						
Non-Progressive News x Child	-0.049	-2.77						
	(0.034)	(2.09)						
Fraction Chores x Child x Boy					-0.030	**-10.2***		
					(0.10)	**(5.90)**		
Fraction Chores x Child					-0.012	4.96		
					(0.053)	(4.26)		
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	16,794	15,023	16,794	15,023	7,461	6,375	7,461	6,375
Individuals	3,306	3,122	3,306	3,122	1,558	1,415	1,558	1,415
Mean of outcome	0.72	27.1	0.72	27.1	0.86	35.5	0.86	35.5
p-value Wald Test: (Child x Boy) + (Non-Progressive News x Child x Boy)	**0.035**	**0.030**						

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the results from regressions which allow for a heterogeneous gender of first child effect on fathers' labour supply by two indicators of the progressiveness of his attitude towards women's role in the labour market. The first indicator of progressiveness measures men's source of news. We create a dummy variable that takes the value 1 if a man receives information about news or current affairs from either the print newspaper or website of The Daily Express, The Daily Mail, The Sun or The Daily Star. The dummy variable takes the value 0 if none of their news sources are from these four companies. Men who receive their news and current affairs from the aforementioned four sources are assumed to have a less progressive attitude. These news source questions are only asked in wave 3. We include all childless men who were interviewed in wave 3 in the regressions. We restrict the sample of fathers to those who had their first child in waves 4, 5 or 6 of the sample, to ensure that their news source does not respond to having a child. We have information on news source from 230 of the 576 fathers. The second measure of men's attitudes towards women's role in the labour market is the initial contribution that men make to household chores. There are six questions that ask who does the following household chores: grocery shopping, cooking, cleaning, washing/ironing, gardening and DIY. We create a dummy variable in each case that takes the value 1 if the man reports doing the task mostly himself. We use these to create a variable that measures the fraction of the six household chores that the man does mostly himself. A man who is more engaged in household chores is assumed to have a more progressive attitude. This question is asked in the UKHLS in waves 2, 4 and 6. For men who never have a child we take their answer from the second wave of the dataset. For men who do have a child, we take their answer from the most recent available wave in which the question was asked before the birth of their first child. Only men with partners are asked this question. We have information on engagement in household chores from 298 out of 576 fathers. In columns (1), (3), (5) and (7), the dependent variable is the dummy for employment status, and in columns (2), (4), (6) and (8), the measure of weekly hours worked. In columns (3), (4), (7) and (8), we re-run the baseline regressions on the sample for whom each measure of progressiveness is defined, to check whether the main effects exist in this sub-sample. Observations in all regressions are weighted by the a_indinus_xw weight. In all regressions, we control for individual fixed effects, year fixed effects and the age of the man.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.9: Non-Degree Educated Fathers Drive Heterogeneity by Progressiveness

	News Source			Contribution to Household Chores		
	(1)	(2)	(3)	(4)	(5)	(6)
	Degree	Non-Degree	Sig Diff?	Degree	Non-Degree	Sig Diff?
Child x Boy	-1.67	-10.7***	-10.7***	-2.88	5.33*	5.20*
	(2.08)	(3.41)	(3.41)	(2.77)	(3.12)	(3.10)
Child	-1.91	0.54	1.00	-1.55	2.10	1.44
	(1.75)	(2.17)	(2.17)	(2.45)	(1.88)	(1.84)
Non-Progressive News x Child x Boy	6.14	**17.2*** **	17.4***			
	(4.15)	**(4.71)**	(4.69)			
Non-Progessive News x Child	0.80	-5.53*	-5.78*			
	(2.46)	(3.19)	(3.18)			
Degree x Child x Boy			8.62**			-8.05*
			(4.00)			(4.13)
Degree x Child			-3.85			-1.99
			(2.73)			(2.98)
Degree x Non-Progressive News x Child x Boy			**-10.8***			
			(6.20)			
Degree x Non-Progressive News x Child			6.60*			
			(3.97)			
Fraction of Chores x Child x Boy				0.63	**-25.0*** **	-24.6***
				(8.48)	**(8.78)**	(8.79)
Fraction of Chores x Child				6.29	4.12	4.28
				(7.69)	(4.79)	(4.98)
Degree x Fraction of Chores x Child x Boy						**24.2***
						(12.2)
Degree x Fraction of Chores x Child						2.28
						(9.27)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4,212	10,799	15,011	2,614	3,761	6,375
Individuals	858	2,262	3,120	561	854	1,415
Mean of outcome	34.6	24.2	27.1	37.7	34.1	35.5

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. In this Table, we ask who drives the heterogeneity in the gender of first child effect by the progressiveness of attitudes towards women's role in the labour market observed in Table 2.8. In all regressions, the dependent variable is weekly hours worked. In columns (1) and (4), we use the sub-sample of degree educated men, and in columns (2), and (5), the sub-sample of non-degree educated men. In columns (3) and (6), we interact each of the four key independent variables in (1), (2), (4) and (5) respectively with a dummy for whether the man is degree educated. If the heterogeneity in the effect by progressiveness is driven by non-degree educated men, then we expect the coefficient on Degree x Non-Progressive News x Child x Boy to be negative, and the coefficient on Degree x Fraction of Chores x Child x Boy to be positive. The coefficient on Degree x Non-Progressive News x Child x Boy is the difference in the difference in the gender of first child effect between the non-progressive and progressive, between the degree educated and the non-degree educated. There are 105 degree educated fathers for whom the news source dummy is observed and 125 non-degree educated fathers. The coefficient on Degree x Fraction of Chores x Child x Boy is the difference in the difference in the gender of first child effect between those who do and do not contribute to household chores, between the degree educated and the non-degree educated. There are 145 degree educated fathers for whom the news source dummy is observed and 153 non-degree educated fathers. Observations in all regressions are weighted by the a_indinus_xw weight. In all regressions, we control for year fixed effects, individual fixed effects and the age of the individual.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.10: Testing the Birthweight Mechanism: Men's Labour Supply

	Baseline Regression			Control for Birthweight		
	(1)	(2)	(3)	(4)	(5)	(6)
	Employed	Weekly Hours	Weekly Hours Intensive Margin	Employed	Weekly Hours	Weekly Hours Intensive Margin
Child x Boy	**-0.063***** **(0.023)**	**-3.67***** **(1.22)**		**-0.065***** **(0.023)**	**-3.75***** **(1.21)**	
Child x Birthweight				0.00055 (0.00048)	0.021 (0.039)	
Child	0.00096 (0.016)	-0.36 (0.91)		-0.065 (0.059)	-2.89 (4.85)	
Child is 0 x Boy			**-1.75*** **(0.97)**			**-1.69*** **(0.94)**
Child is 0 x Birthweight						-0.026 (0.033)
Child is 0			-0.065 (0.76)			3.06 (4.38)
Child older than 0 x Boy			0.44 (1.09)			0.55 (1.02)
Child older than 0 x Birthweight						-0.028 (0.046)
Child older than 0			-1.87** (0.79)			1.54 (5.92)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,618	21,212	11,186	23,618	21,212	11,186
Individuals	6,921	6,480	3,447	6,921	6,480	3,447
Mean of outcome	0.71	26.6	40.8	0.71	26.6	40.8

Notes. Standard errors (in parentheses) clustered at the primary sampling unit level. Observations in all regressions are weighted by the a_indinus_xw weight. This Table contains the results from the test of the birthweight mechanism for the gender of first child effect on fathers' labour supply. Of the 576 fathers, we observe the birthweight of the first child for 419. For this reason, in columns (1)-(3), we re-run the main regressions for men from Table 2.2 on a restricted sample which drops fathers for whom child birthweight is missing, to check if the main results still exist in this restricted sample. The gender of first child effects do still exist on both the extensive and intensive margin. In columns (4)-(6), we then test the birthweight mechanism. We include an interaction term, Child x Birthweight, to control for the additional strain imposed on the mother from a higher birthweight child. If the gender of first child effect only operates through birthweight, the coefficient on Child x Boy should fall in absolute magnitude towards zero, and the coefficient on Child x Birthweight should be negative and significant.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.11: Testing the Pure Son Preference Mechanism

	Family Structure					Ethnic Group			
	Fertility		Relationship Status			White	Minority	White	Minority
	Men	Women	Men	Women	Women	Men			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Add Child	Add Child	Married	Married	Single	Employed	Employed	Weekly Hours	Weekly Hours
Child x Boy	-0.0062	-0.00062	0.0024	0.0082	-0.0087	**-0.050****	0.020	**-3.49*****	0.50
	(0.0081)	(0.0070)	(0.022)	(0.017)	(0.021)	**(0.023)**	(0.063)	**(1.17)**	(3.52)
Child	1.00***	0.97***	0.049***	0.028***	-0.093***	-0.0028	-0.042	-0.46	-0.82
	(0.011)	(0.0074)	(0.016)	(0.011)	(0.014)	(0.015)	(0.047)	(0.86)	(2.71)
Time Since Previous Birth	0.0049***	0.0045***							
	(0.00070)	(0.00060)							
Second Parity	-0.96***	-0.94***							
	(0.0056)	(0.0043)							
Third Parity	-0.33***	-0.32***							
	(0.019)	(0.017)							
Fourth Parity	-0.25*	-0.24**							
	(0.13)	(0.10)							
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,353	23,711	24,345	23,684	9,959	18,944	5,387	16,823	4,987
Individuals	7,078	6,464	7,074	6,458	2,456	5,166	1,902	4,792	1,824
Mean of outcome	0.007	0.009	0.23	0.25	0.055	0.74	0.62	28.5	21.3

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. Observations in all regressions are weighted by the a_indinus_xw weight. In this Table, we test the pure son preference mechanism in three ways. First, in columns (1) and (2), we ask whether the gender of the first child affects the probability an individual has additional children. The dependent variable is an impulse dummy variable that takes value 1 if the individual has a subsequent child (after the first child) in that wave. We control for time since previous birth (starting from age 15) and dummy variables for the birth parity of the individual. Second, in columns (3), (4) and (5), we ask whether the gender of the first child affects the relationship status of the parents. In columns (3) and (4), we ask whether the gender of first child affects the probability that an individual is married for men and women respectively. This combines whether the gender of first child affects the probability that an initially not married individual becomes married, and whether an initially married individual remains married. However, just one father and seven mothers in the sample got divorced during the sample period, so this regression really provides information about whether the gender of first child affects the probability that individuals get married. In the sample, 36 fathers and 35 mothers got married after having their first child. The dependent variable is a dummy variable taking value 1 if the individual is married. In column (5), we restrict the sample to women who are initially not single, and ask whether the gender of first child affects the probability that they become single (where by initially we mean the first wave of the data for the women who never have a child and the wave before they have a child for the women who do). The dependent variable is a dummy variable taking value 1 if the woman is single. We do this only for women and not men as there are only three observations where a man has a child and is single in our sample. One caution with the tests in columns (1) to (5) is that we only observe parents for at most five time periods with their child, and so we can only understand whether the gender of first child affects family structure in those immediate few years after birth. In columns (6), (7), (8) and (9), we ask which group of fathers by ethnic group drive the main effect. We separate out the sub-sample of white men in columns (6) and (8), from ethnic minority men in columns (7) and (9).

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix

2.A Additional Tables and Figures

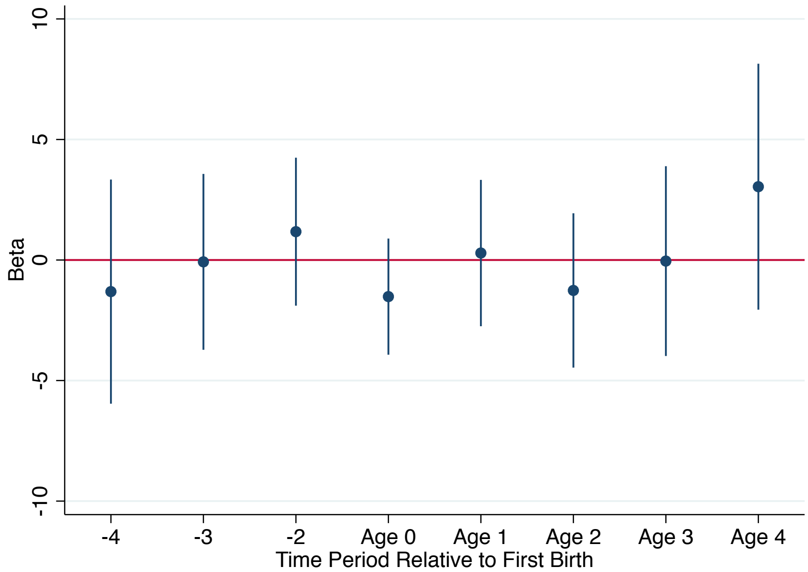


Figure A.1: *Notes.* **Gender of First Child Effect on Weekly Hours for Women**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on weekly hours. The base year is the wave before the birth of the first child. For the regression underlying the figure, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the woman's age, and we cluster standard errors at the level of the primary sampling unit. Mothers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

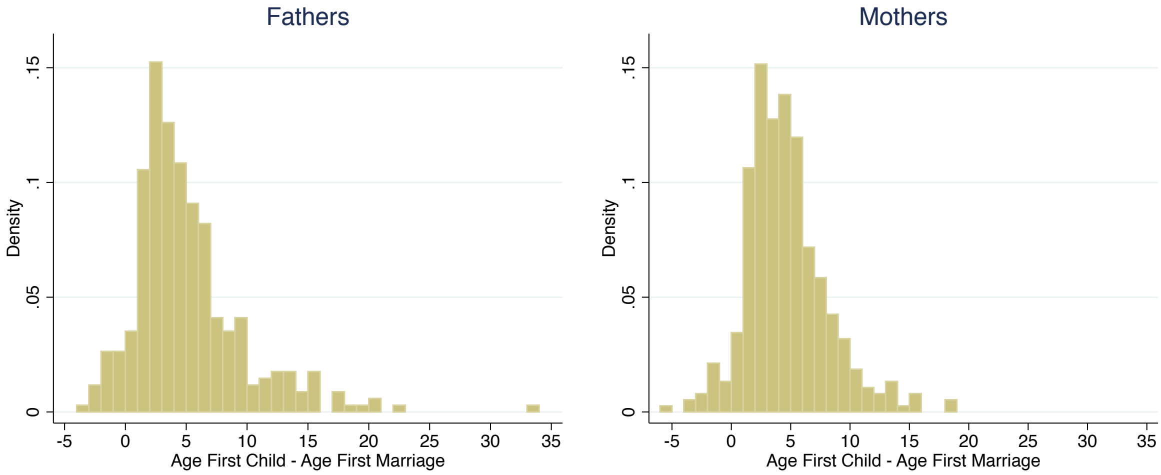


Figure A.2: *Notes.* **Distribution of the Difference Between Age at First Child and Age at First Marriage**. This Figure presents histograms of the length of time between age at the birth of the first child and age at first marriage for the fathers and mothers in the main sample. Of the 576 fathers, 342 have been married by the end of the sample period. Of the 744 mothers, 377 have been married by the end of the sample period. For both fathers and mothers, having a child two years after marriage is the mode. The mean and standard deviation are: 4.6 and 4.6 years for fathers, and 4.1 and 3.4 years for mothers.

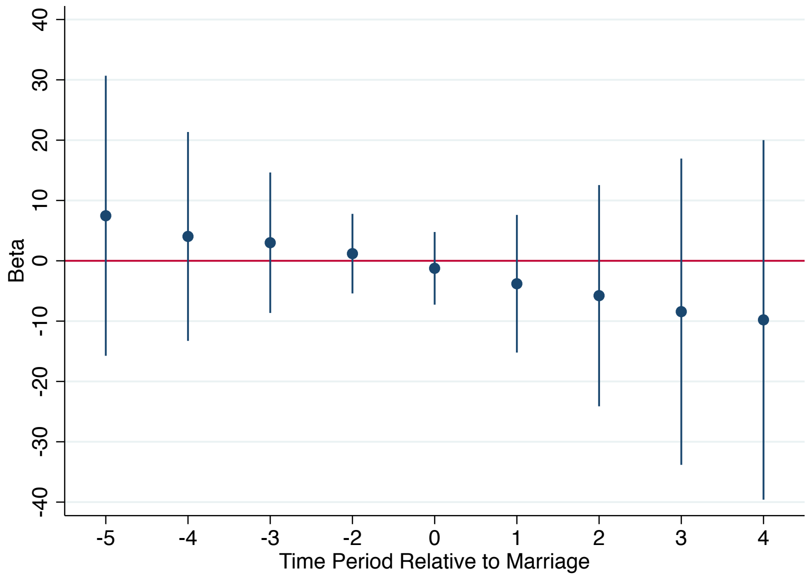


Figure A.3: *Notes.* **Effect of Marriage on Weekly Hours for Men**. This Figure presents the point estimates and 95% confidence intervals for an event study of the effect of marriage on weekly hours for men. We create a set of impulse dummy variables, each of which takes the value 1 for a single time period relative to the wave in which the man gets married. The base year is the wave before the man gets married. For the regression underlying the figure, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the man's age, and we cluster standard errors at the level of the primary sampling unit. We restrict the sample to all men in the UKHLS (regardless of how many children he has) who start the sample period single and get married at some point during the sample period. Men get married at different points during the sample period. Estimates further from the base year are therefore more susceptible to compositional effects and should be treated with caution.

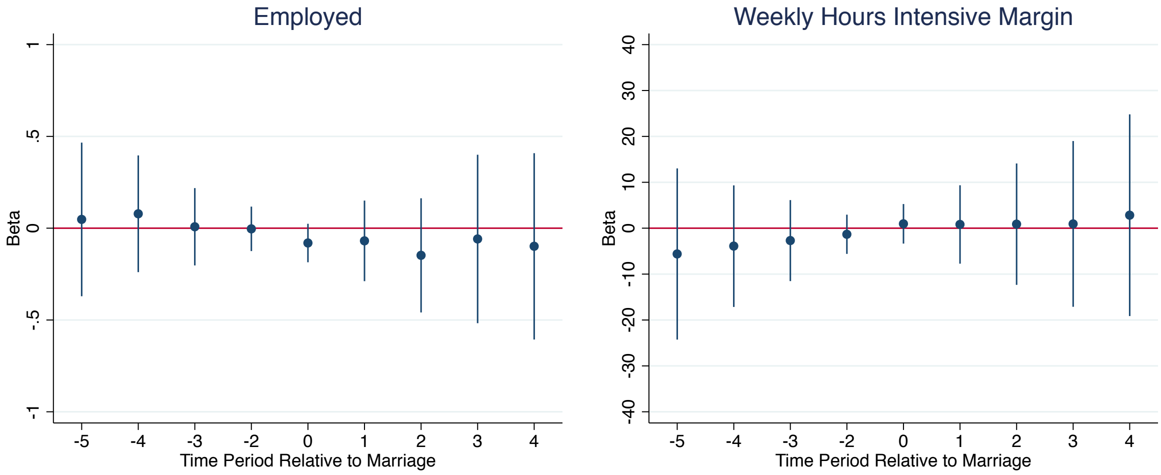


Figure A.4: *Notes.* **Effect of Marriage on Extensive and Intensive Margin for Women**. This Figure presents the point estimates and 95% confidence intervals for an event study of the effect of marriage on the extensive and intensive margin for women. We create a set of impulse dummy variables, each of which takes the value 1 for a single time period relative to the wave in which the woman gets married. The base year is the wave before the woman gets married. For the regressions underlying both figures, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the woman's age, and we cluster standard errors at the level of the primary sampling unit. We restrict the sample to all women in the UKHLS (regardless of how many children she has) who start the sample period single and get married at some point during the sample period. The left hand panel plots the effect on the extensive margin, where the dependent variable is the employment dummy. The right hand panel plots the effect on the intensive margin, where the dependent variable is weekly hours. Women get married at different points during the sample period. Estimates further from the base year are therefore more susceptible to compositional effects and should be treated with caution.

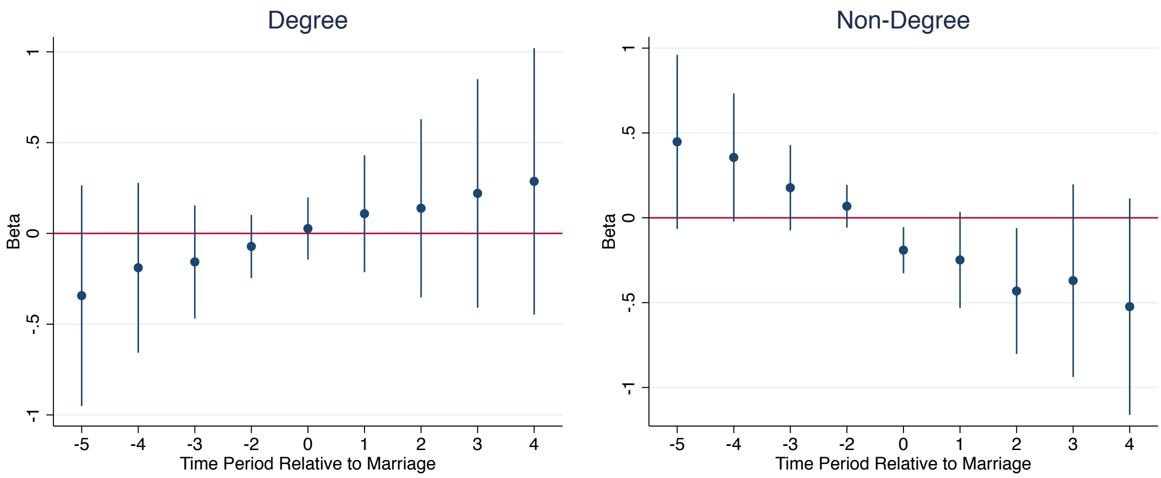


Figure A.5: *Notes.* **Effect of Marriage on Extensive Margin for Degree and Non-Degree Educated Women**. This Figure presents the point estimates and 95% confidence intervals for an event study of the effect of marriage on the extensive margin for degree and non-degree educated women. We create a set of impulse dummy variables, each of which takes the value 1 for a single time period relative to the wave in which the woman gets married. The base year is the wave before the woman gets married. For the regressions underlying both figures, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the woman's age, and we cluster standard errors at the level of the primary sampling unit. We restrict the sample to women in the UKHLS (regardless of how many children she has) who start the sample period single and get married at some point during the sample period. The left hand panel plots the effect on the extensive margin for degree educated women, and the right hand panel for non-degree educated women. In both cases, the dependent variable is the employment dummy. Women get married at different points during the sample period. Estimates further from the base year are therefore more susceptible to compositional effects and should be treated with caution.

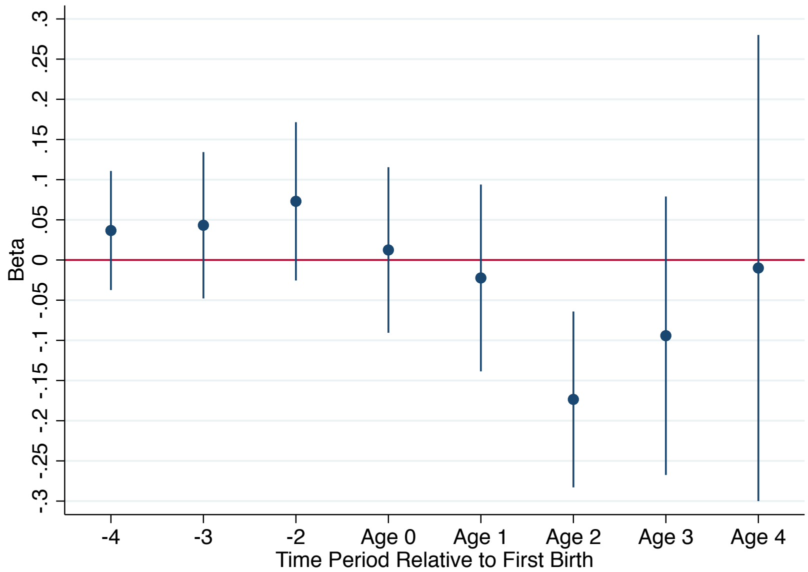


Figure A.6: *Notes.* **Gender of First Child Effect on Extensive Margin for Partners of Degree Educated Men**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on the extensive margin for the partners of degree educated men. The dependent variable is the employment dummy. The base year is the wave before the birth of the first child. For the regression underlying the figure, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects, and the partner's age, and we cluster standard errors at the level of the primary sampling unit. Fathers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

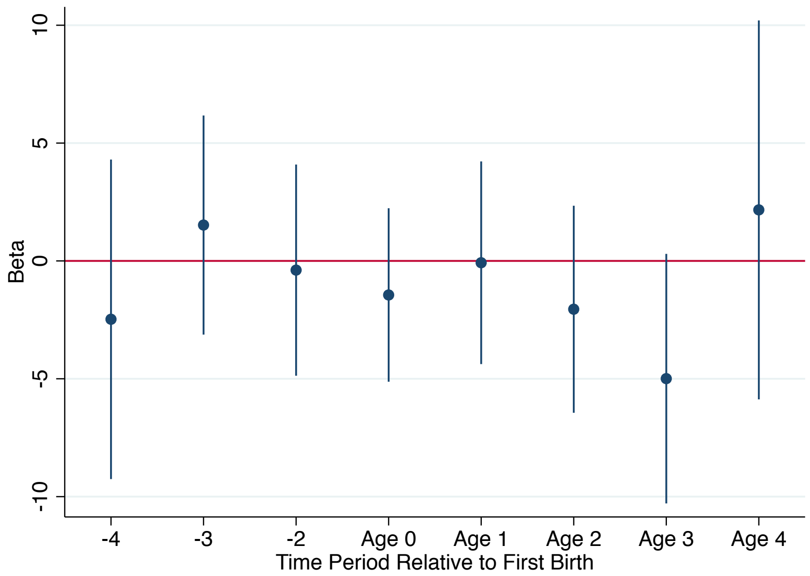


Figure A.7: *Notes.* **Gender of First Child Effect on Weekly Hours for Degree Educated Women**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on weekly hours for degree educated women. The base year is the wave before the birth of the first child. For the regression underlying the figure, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the woman's age, and we cluster standard errors at the level of the primary sampling unit. Mothers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

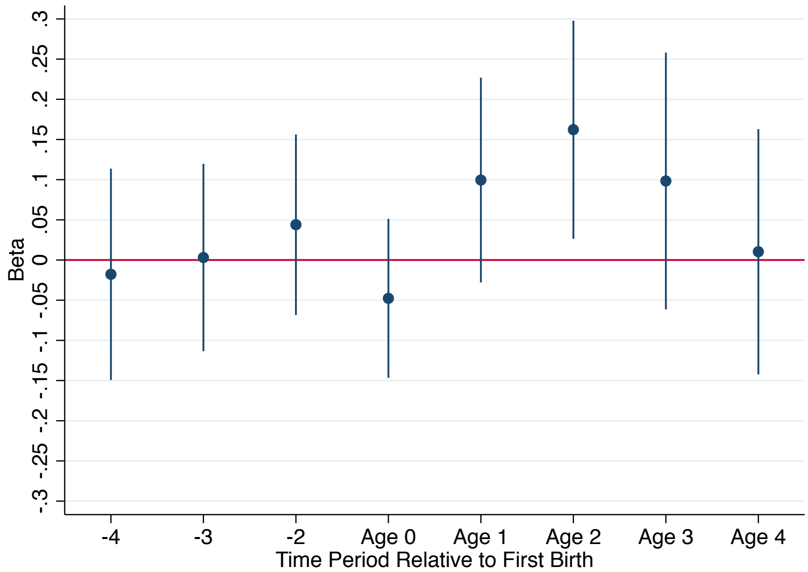


Figure A.8: *Notes.* **Gender of First Child Effect on Extensive Margin for Partners of Non-Degree Educated Men**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on the extensive margin for the partners of non-degree educated men. The dependent variable is the employment dummy. The base year is the wave before the birth of the first child. For the regression underlying the figure, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects, and the partner's age, and we cluster standard errors at the level of the primary sampling unit. Fathers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

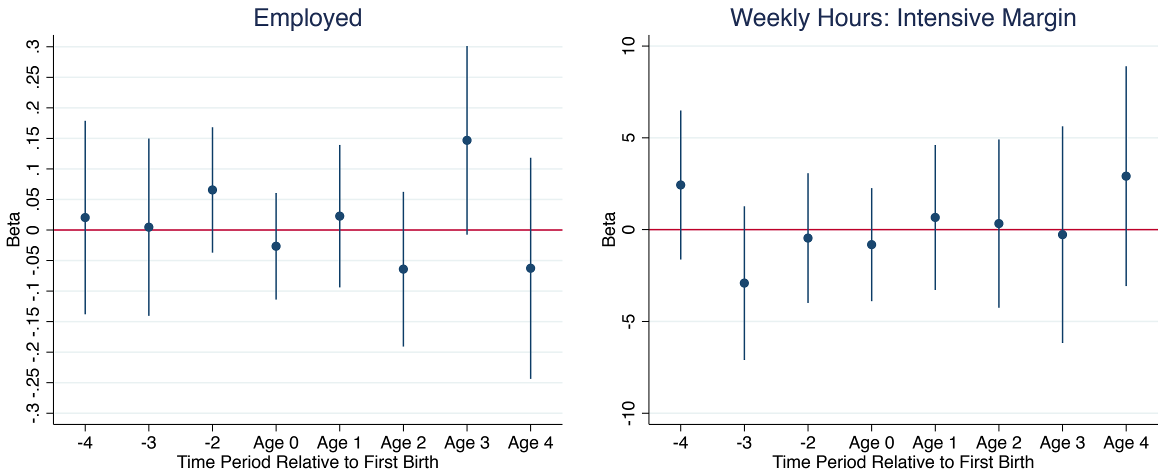


Figure A.9: *Notes.* **Gender of First Child Effect on the Extensive and Intensive Margin for Non-Degree Educated Women**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on the extensive and intensive margins separately for non-degree educated women. The base year is the wave before the birth of the first child. For the regressions underlying both figures, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the woman's age, and we cluster standard errors at the level of the primary sampling unit. The left hand panel plots the results for the extensive margin, where the dependent variable is the employment dummy. The right hand panel plots the results for the intensive margin, where the dependent variable is weekly hours. Mothers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

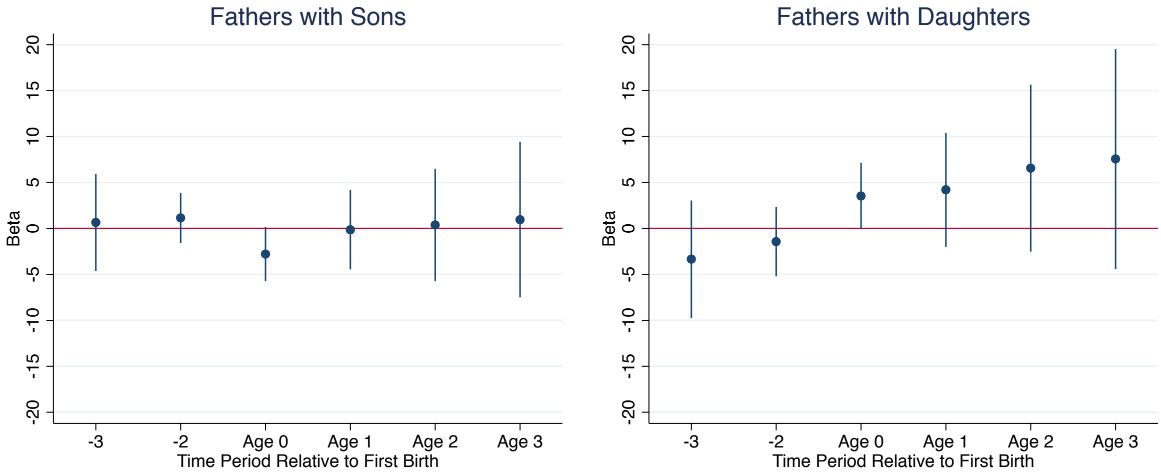


Figure A.10: *Notes.* **Effect of Having a Child on Weekly Hours for Fathers with Sons versus Daughters**. This Figure presents the point estimates and 95% confidence intervals for an event study of the effect of having a child on weekly hours worked, separating the sample of fathers into those with sons (left hand panel) and daughters (right hand panel). This Figure demonstrates that the gender of first child effect is in part driven by a behavioural change by men with sons and by men with daughters, before and after the birth of their first child. We restrict the sample to men who have a child at some point during the sample period. The base year is the wave before the birth of the first child. For the regressions underlying both figures, we weight observations using the a_indinus_xw weight, we control for individual fixed effects and year fixed effects, and we cluster standard errors at the level of the primary sampling unit. Fathers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from four waves or more before the birth of the first child, and from when the child is aged 4 or older.

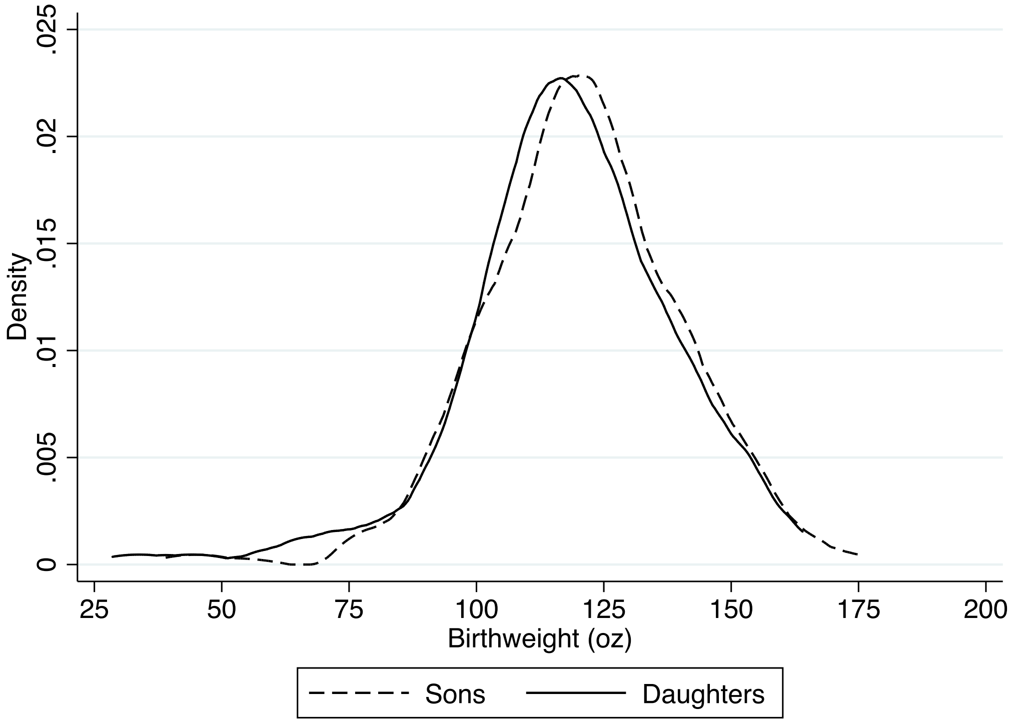


Figure A.11: *Notes.* **Birthweight of Sons and Daughters.** In this Figure, we present kernel density plots to demonstrate the distribution of birthweight for the sons and daughters of the fathers for whom we observe child birthweight. Boys weigh more on average at birth than girls. To test the child birthweight mechanism in Table 2.10, we require that the distributions of birthweight for boys and girls overlap. Otherwise, it is not possible to distinguish the effect of child birthweight from child gender. This Figure demonstrates that indeed the distributions overlap, and so provides justification for the tests presented in columns (4) to (6) of Table 2.10. We observe the birthweight of 216 girls and 203 boys, who are included in those regressions. One boy is claimed to have a birthweight of 317 ounces, which is 1.8 times the birthweight of the next heaviest boy. We have excluded this outlier from the kernel density plot. We include that boy's father in the regressions presented in Table 2.10, but the results are almost identical when he is excluded.

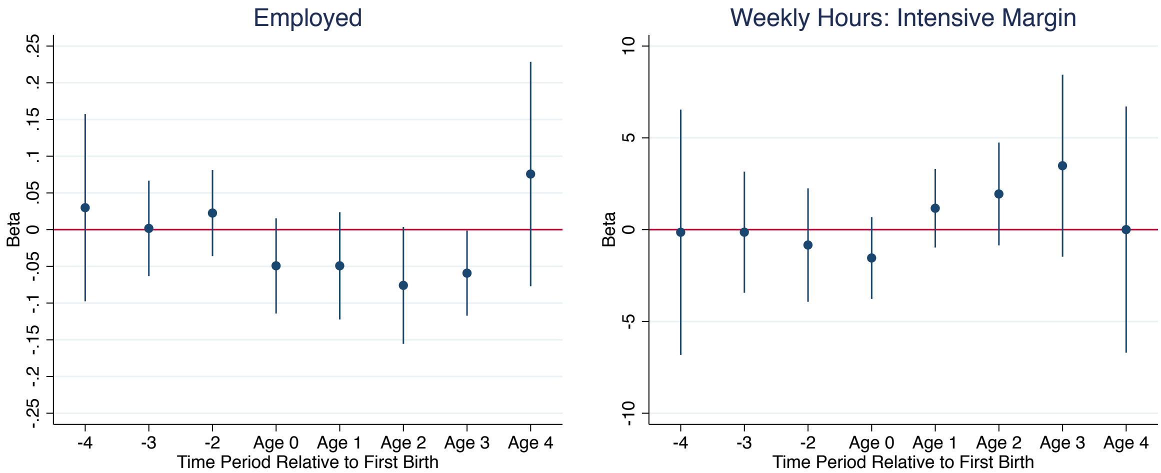


Figure A.12: *Notes.* **Gender of First Child Effect on the Extensive and Intensive Margin for Degree Educated Men**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on the extensive and intensive margins separately for degree educated men. The base year is the wave before the birth of the first child. For the regressions underlying both figures, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the man's age, and we cluster standard errors at the level of the primary sampling unit. The left hand panel plots the results for the extensive margin, where the dependent variable is the employment dummy. The right hand panel plots the results for the intensive margin, where the dependent variable is weekly hours. Fathers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

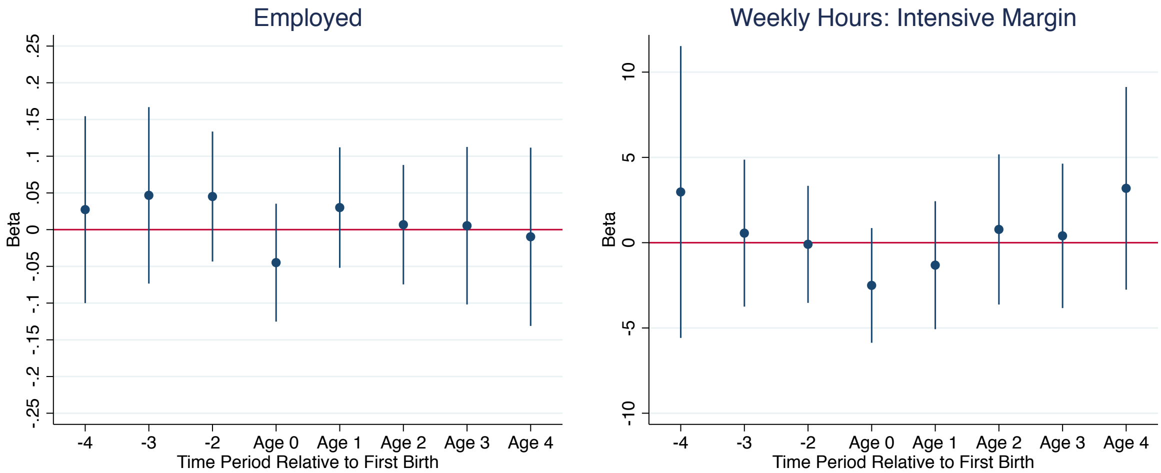


Figure A.13: *Notes.* **Gender of First Child Effect on the Extensive and Intensive Margin for Non-Degree Educated Men**. This Figure presents the point estimates and 95% confidence intervals for an event study of the gender of first child effect on the extensive and intensive margins separately for non-degree educated men. The base year is the wave before the birth of the first child. For the regressions underlying both figures, we weight observations using the a_indinus_xw weight, we control for individual fixed effects, year fixed effects and the man's age, and we cluster standard errors at the level of the primary sampling unit. The left hand panel plots the results for the extensive margin, where the dependent variable is the employment dummy. The right hand panel plots the results for the intensive margin, where the dependent variable is weekly hours. Fathers have their first child at different points during the sample period, and so estimates further from the base year are more susceptible to compositional effects. For that reason, we drop observations from five waves before the birth of the first child, and from when the child is aged 5 or older.

Table A.1: Regression Results Underlying Figure 2.1

	(1) Weekly Hours
Child is 0 x Boy	**-4.90***** **(1.47)**
Child is 0	1.36 (0.95)
Child is 1 x Boy	-1.27 (1.46)
Child is 1	-0.29 (1.11)
Child is 2 x Boy	-1.42 (1.67)
Child is 2	-0.28 (1.18)
Child is 3 x Boy	-0.40 (2.27)
Child is 3	-1.90 (1.54)
Child is 4 x Boy	1.52 (3.21)
Child is 4	-5.14** (2.46)
Year Fixed Effects	Yes
Individual Fixed Effects	Yes
Control for Age	Yes
Observations	21,767
Individuals	6,624
Mean of outcome	26.8

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the results from the regression underlying the event study in Figure 2.1. We create an impulse dummy variable for the time periods four, three and two waves before the birth of the first child, and when the child is aged 0, 1, 2, 3 and 4. We interact each with the variable Boy. We therefore estimate a separate gender of first child effect in each of the first five years of the child's life. We drop observations from five waves before the birth of the first child as well as from when the child is aged 5 and older (there are very few such observations, and as such we cannot estimate effects in those years). The gender of first child effects are estimated relative to a base year, which is the year before the birth of the first child. The dependent variable is weekly hours. Observations are weighted using the a_indinus_xw weight, and we control for individual fixed effects, year fixed effects and the age of the individual.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.2: Gender of First Child Effect for Fathers on Both Extensive Margins

	(1)	(2)
	Initially Employed	Initially Not Employed
Child x Boy	**-0.035****	**-0.21**
	(0.017)	**(0.14)**
Child	0.054***	0.24**
	(0.0091)	(0.11)
Year Fixed Effects	Yes	Yes
Individual Fixed Effects	Yes	Yes
Control for Age	Yes	Yes
Observations	16,348	8,005
Individuals	4,566	2,512
Mean of outcome	0.94	0.25

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. Observations in all regressions are weighted by the a_indinus_xw weight. In this Table, we break down the overall effect on employment for fathers (the extensive margin), into two separate extensive margins. In column (1), we restrict the sample to men who are initially employed and ask whether the gender of first child affects the probability that they remain employed. In column (2), we restrict the sample to men who are initially not employed, and ask whether the gender of first child affects the probability that they become employed. By initially, we mean the first wave of the sample for men who never have a child, and the wave before the birth of the first child for men who do go on to have a child. The dependent variable in both regressions is the employment dummy variable, taking value 1 if the individual is employed and 0 otherwise.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Job Changes and Regular Versus Overtime Hours

	Job Change			Regular Versus Overtime	
	(1)	(2)	(3)	(4)	(5)
	Change Industry	Change Occupation	Change Either	Regular Hours	Overtime
Child x Boy	0.020 (0.030)	0.031 (0.040)	0.012 (0.041)		
Child	-0.0095 (0.025)	-0.0044 (0.029)	0.0064 (0.030)		
Child is 0 x Boy				**-1.41**** **(0.69)**	-0.76 (0.59)
Child is 0				-0.052 (0.49)	-0.035 (0.45)
Child older than 0 x Boy				-0.46 (0.71)	**1.22*** **(0.71)**
Child older than 0				-0.59 (0.54)	-1.24** (0.51)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes
Observations	11,420	11,448	11,564	11,649	11,618
Individuals	3,591	3,597	3,606	3,550	3,547
Mean of outcome	0.089	0.13	0.15	36.8	4.16

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. Observations in all regressions are weighted by the a_indinus_xw weight. In columns (1)-(3), we attempt to understand whether the change in hours worked for fathers is driven by a change in job or a change in hours worked within their existing jobs. In column (1), we use a dependent variable that is an impulse dummy variable taking the value 1 if the man changed industry compared to the previous wave. In column (2), we rather use as the dependent variable an impulse dummy variable that takes value 1 if he changes occupation. In column (3), we create a third impulse dummy variable which takes value 1 if either of the dependent variables in column (1) or (2) takes value 1, and use this as the dependent variable. In columns (4) and (5), we attempt to understand whether the gender of first child effect comes from a change in regular hours or overtime hours. In column (4), the dependent variable is the regular weekly hours of the man, whilst in column (5) it is the normal weekly overtime hours of the man. In columns (4) and (5), we restrict the sample to men who are employed in every wave in which they are observed in the sample: the intensive margin. Since we have evidence that the effects on the intensive margin are concentrated in the first year of the child's life, we allow for a separate effect when the child is aged 0 compared to older than 0.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Additional Robustness Checks: Add Control Variables, Drop Later Births and Drop Outliers

	Control Total Children		Quarter FE		Wave FE		Drop Later Births		Drop Outliers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Employed	Weekly Hours	Employed	Weekly Hours	Employed	Weekly Hours	Employed	Weekly Hours	Weekly Hours
Child x Boy	**-0.046****	**-3.21*****	**-0.045****	**-3.08*****	**-0.044****	**-3.06*****	**-0.047****	**-3.84*****	**-3.29*****
	(0.021)	**(1.12)**	**(0.021)**	**(1.12)**	**(0.021)**	**(1.12)**	**(0.022)**	**(1.13)**	**(1.10)**
Child	0.043**	2.09	-0.0057	-0.51	-0.0063	-0.56	0.00055	0.036	-0.35
	(0.019)	(1.42)	(0.015)	(0.83)	(0.015)	(0.83)	(0.014)	(0.82)	(0.82)
Total Children	-0.043***	-2.28**							
	(0.014)	(1.15)							
Year Fixed Effects	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Yes
Quarter Fixed Effects	No	No	Yes	Yes	No	No	No	No	No
Wave Fixed Effects	No	No	No	No	Yes	Yes	No	No	No
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,353	21,830	24,353	21,830	24,353	21,830	24,040	21,569	21,691
Individuals	7,078	6,625	7,078	6,625	7,078	6,625	7,078	6,622	6,610
Mean of outcome	0.71	26.8	0.71	26.8	0.71	26.8	0.71	26.7	26.4

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. Observations in all regressions are weighted by the a_indinus_xw weight. In this Table we perform further robustness checks for the main results for fathers' employment and weekly hours worked. In columns (1) and (2), we additionally control for the total number of children the individual has ever had. In columns (3) and (4), we control for quarter fixed effects rather than year fixed effects as in the baseline. In columns (5) and (6), we control for wave fixed effects in addition to the calendar year fixed effects. Finally, in columns (7) and (8), we drop observations from the wave in which an individual adds a second child onwards. There are three fewer individuals in the regression in column (8) than the baseline. The reason is that these three individuals were self-employed in years when they did not have more than one child, so that they are only included in the weekly hours regression when we include years in which they have more than one child. In column (9), we drop individuals who work more than 75 hours per week. Otherwise the regressions are as in the baseline.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Additional Robustness Checks: Specification Changes and Alternative Weekly Hours Definitions

	Employed			Weekly Hours					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Pooled OLS	Probit	Unweighted	Pooled OLS	Tobit	Unweighted	Trimmed LS	Include Self-Employed	Include Second Job
Child x Boy	**-0.044****	**-0.36****	**-0.039***	**-2.98****	**-3.72****	**-2.57****	**-2.71****	**-2.95*****	**-3.00*****
	(0.019)	**(0.16)**	**(0.021)**	**(1.27)**	**(1.53)**	**(1.12)**	**(1.20)**	**(1.12)**	**(1.14)**
Child	0.20***	1.02***	-0.0070	12.6***	16.1***	-0.39	-1.47*	-0.63	-0.72
	(0.013)	(0.13)	(0.015)	(0.84)	(1.03)	(0.83)	(0.88)	(0.76)	(0.77)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	No	No	Yes	No	No	Yes	No	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls for Ethnic Group	Yes	Yes	No	Yes	Yes	No	No	No	No
Observations	24,331	24,331	24,353	21,810	21,810	21,830	21,830	24,149	24,149
Individuals	7,068	7,068	7,078	6,616	6,616	6,625	6,625	7,058	7,058
Mean of outcome	0.71	0.71	0.71	26.8	26.8	26.8	26.8	28.1	28.5

Notes. Standard errors (in parentheses) clustered at the primary sampling unit level. In this Table, we perform further robustness checks of the main results for fathers' employment and weekly hours worked. In columns (1), (2), (4) and (5), we drop individual fixed effects. We then control for ethnic group dummy variables, specifically whether an individual is Black, Asian or Other Race (relative to a White reference group). Ethnic group is missing for 10 individuals in the full sample of men, which is why there are fewer observations in these regressions compared to the baseline. In columns (1) and (4), we run pooled OLS regressions, whilst in column (2) we run a Probit model given the binary nature of the employment dummy variable, and in column (5) we run a Tobit model to deal with the fact that weekly hours are censored from below at 0. In column (7), we use Honore's trimmed LS estimator for estimation of censored fixed effects models, given the incidental parameters problem when attempting to estimate a fixed effects Tobit model. Observations in regressions in columns (3) and (6) are unweighted, whilst observations in all other regressions are weighted by the a_indinus_xw weight as in the baseline. Finally, in columns (8) and (9) we use two different definitions of weekly hours worked. In column (8), we include the self-employed, using a variable jshrs from the UKHLS dataset. In column (9), we add hours worked in second jobs for any individual in the sample working a second job, and again include the self-employed. Hours worked in second jobs are given per month, and so we first divide monthly hours worked in the second job by four and then add to hours worked in the individual's main job.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Balance Tests By Gender of First Child

	Fathers			Mothers		
	Daughter	Son	p-value (H0: Diff = 0)	Daughter	Son	p-value (H0: Diff = 0)
Age	29.834	29.231	0.261	26.636	26.454	0.694
	(0.381)	(0.378)		(0.323)	(0.334)	
Married	0.452	0.338	0.009***	0.379	0.298	0.023**
	(0.032)	(0.029)		(0.026)	(0.025)	
Weekly Hours	38.457	37.153	0.453	31.978	31.865	0.935
	(1.127)	(1.317)		(0.997)	(0.979)	
Employed	0.900	0.874	0.366	0.840	0.858	0.515
	(0.018)	(0.022)		(0.020)	(0.019)	
Religious	0.410	0.368	0.339	0.455	0.359	0.011**
	(0.032)	(0.031)		(0.027)	(0.027)	
Urban	0.810	0.825	0.677	0.819	0.820	0.979
	(0.026)	(0.024)		(0.022)	(0.021)	
Foreign Born	0.172	0.119	0.070*	0.155	0.129	0.354
	(0.023)	(0.018)		(0.020)	(0.022)	
Ethnic Minority	0.108	0.094	0.507	0.080	0.081	0.973
	(0.015)	(0.014)		(0.011)	(0.011)	
Degree	0.376	0.395	0.672	0.438	0.451	0.749
	(0.032)	(0.032)		(0.028)	(0.029)	
GCSE only	0.239	0.302	0.131	0.227	0.219	0.813
	(0.030)	(0.032)		(0.023)	(0.024)	
Monthly Wages	7.707	7.675	0.656	7.382	7.331	0.491
	(0.052)	(0.049)		(0.049)	(0.056)	
Observations	284	292	576	385	359	744

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. We present the weighted sample means of each demographic characteristic for parents with a first born son versus a first born daughter, for fathers and mothers separately. We then present p-values for tests of differences in those weighted sample means across the gender of the first child. We consider the characteristics of mothers and fathers in the first wave of the dataset, which is always before the birth of the first child. We weight observations using the a_indinus_xw sampling weight as in the regression analysis. We consider the natural logarithm of the real monthly wage.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.7: Marriage and Labour Supply

	Men			Women		
	(1)	(2)	(3)	(4)	(5)	(6)
	Weekly Hours	Employed	Weekly Hours Intensive Margin	Weekly Hours	Employed	Weekly Hours Intensive Margin
Married	0.76	0.0039	0.80	-1.34	**-0.055****	-0.24
	(1.15)	(0.020)	(0.95)	(1.24)	**(0.024)**	(1.22)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,136	1,324	861	1,640	1,750	942
Individuals	237	257	178	324	332	183
Mean of outcome	37.9	0.92	42.5	26.3	0.76	37.6

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the results from regressions which ask what effect marriage has on the labour supply of men and women. We restrict the sample to all individuals in the UKHLS (regardless of how many children they have) who start the sample period single, and get married (in a heterosexual marriage) at some point during the sample period. We keep only observations whilst the individual is single or married (we do not include observations from after an individual is separated, divorced or widowed). In columns (1) and (4), we look at the effect of marriage on weekly hours worked. In columns (2) and (5), we look at the effect of marriage on employment. In columns (3) and (6), we look at the effect of marriage on weekly hours worked including only individuals on the intensive margin. Observations are weighted by the a_indinus_xw weight. In all regressions, we control for individual fixed effects, year fixed effects and the individual's age.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Additional Heterogeneity by Socio-Economic Status: By Fathers' Own Characteristics

	Marital Status					Occupation					Wages				
	Employed			Weekly Hours Intensive Margin		Employed			Weekly Hours Intensive Margin		Employed			Weekly Hours Intensive Margin	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Married	Unmarried	Sig Diff?	Married	Unmarried	Non-Professional	Professional	Sig Diff?	Non-Professional	Professional	Low Wage	High Wage	Sig Diff?	Low Wage	High Wage
Child x Boy	**-0.055*****	-0.029	-0.029			-0.018	**-0.046****	-0.048**			-0.024	-0.024	-0.023		
	(0.021)	(0.043)	(0.043)			(0.026)	**(0.023)**	(0.024)			(0.027)	(0.016)	(0.027)		
Child	0.075***	-0.020	0.00021			0.062***	0.039***	0.057***			0.057***	0.042***	0.055***		
	(0.016)	(0.029)	(0.028)			(0.016)	(0.011)	(0.011)			(0.013)	(0.0090)	(0.013)		
Married x Child x Boy			-0.028												
			(0.048)												
Married x Child			-0.0088												
			(0.032)												
Non-Professional x Child x Boy								0.030							
								(0.035)							
Non-Professional x Child								-0.011							
								(0.019)							
High Wage x Child x Boy													-0.0014		
													(0.032)		
High Wage x Child													-0.0099		
													(0.014)		
Child is 0 x Boy				-0.46	**-5.56*****				**-3.19****	-1.64				**-2.31***	**-2.18***
				(0.99)	**(1.74)**				**(1.58)**	(1.09)				**(1.29)**	**(1.29)**
Child is 0				-0.14	1.36				0.38	0.047				-0.37	0.77
				(0.86)	(1.15)				(1.15)	(0.87)				(0.89)	(1.16)
Child older than 0 x Boy				**2.02***	-2.69				**-3.32***	**2.10****				-0.52	1.56
				(1.09)	(1.79)				**(1.81)**	**(1.04)**				(1.20)	(1.48)
Child older than 0				-1.94**	1.09				0.82	-2.01**				-0.64	-2.38**
				(0.81)	(1.36)				(1.40)	(0.79)				(0.93)	(1.05)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	5,554	18,799	24,353	3,458	8,146	8,428	7,747	16,175	5,774	5,745	9,997	4,188	14,185	7,762	3,643
Individuals	1,394	5,684	7,078	905	2,639	2,483	2,023	4,506	1,889	1,619	2,901	1,076	3,977	2,468	990
Mean of outcome	0.85	0.67	0.71	43.4	39.9	0.92	0.96	0.94	38.8	43.2	0.93	0.96	0.94	38.8	45.5

Notes. Standard errors (in parentheses) clustered at the primary sampling unit level. Observations are weighted using the a_indinus_xw weight in all regressions. In this Table, we present additional heterogeneity of the effect by indicators of socio-economic status. We focus on characteristics of the men themselves. In columns (1)-(5), we distinguish by whether the man is initially married or unmarried. In columns (6)-(10), we distinguish men initially in non-professional from professional occupations. Finally, in columns (11)-(15), we distinguish men who are initially on low wages from those initially on high wages. Here, we categorise men as high wage if they earn more than the 75th percentile real monthly wage (among the full sample of observations for employed men). This is £2,723 per month (in 2015 pounds). For each demographic characteristic, we first look at the employment dummy variable, splitting the sample into two sub-samples and then testing for a difference across those sub-samples (columns (1), (2), (3), (6), (7), (8), (11), (12), (13)). We then look at the effect on weekly hours worked on the intensive margin by splitting into the two sub-samples (columns (4), (5), (9), (10), (14), (15)). By initially here, we refer to the first wave of the sample period for men who never have a child, and to the wave before the birth of the first child for those who do. Only men who are initially employed are included in the regressions that categorise by occupation, as these are the only men for whom initial occupation is defined. Only men who are initially employed, and not in self-employment, are included in the regressions that categorise by initial wages, as these are the only men for whom initial wages are defined.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Additional Heterogeneity by Socio-Economic Status: By Partner of Fathers' Characteristics

	Partner Education Level					Partner Occupation					Partner Wages				
	Employed			Weekly Hours Intensive Margin		Employed			Weekly Hours Intensive Margin		Employed			Weekly Hours Intensive Margin	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Degree	No Degree	Sig Diff?	Degree	No Degree	Partner Non-Professional	Partner Profes-sional	Sig Diff?	Partner Non-Professional	Partner Profes-sional	Low Wage	High Wage	Sig Diff?	Low Wage	High Wage
Child x Boy	**-0.060****	-0.036	-0.035			-0.0080	**-0.070*****	-0.069**			**-0.039***	**-0.10*****	-0.038*		
	(0.026)	(0.035)	(0.035)			(0.027)	**(0.027)**	(0.027)			**(0.023)**	**(0.039)**	(0.023)		
Child	0.045**	0.082***	0.074***			0.040**	0.058***	0.060***			0.058***	0.053*	0.057***		
	(0.019)	(0.023)	(0.023)			(0.018)	(0.019)	(0.019)			(0.017)	(0.028)	(0.017)		
Partner Degree x Child x Boy			-0.025												
			(0.044)												
Partner Degree x Child			-0.018												
			(0.029)												
Partner Non-Professional x Child x Boy								**0.061**							
								(0.038)							
Partner Non-Professional x Child								-0.023							
								(0.024)							
Partner High Wage x Child x Boy													-0.066		
													(0.045)		
Partner High Wage													0.00050		
													(0.031)		
Child is 0 x Boy				-0.47	**-3.91****				**-3.11***	-1.18				**-2.41****	-0.49
				(1.11)	**(1.65)**				**(1.72)**	(1.08)				**(1.21)**	(1.62)
Child is 0				-0.49	2.03*				2.36*	-0.81				0.84	-0.76
				(0.93)	(1.12)				(1.35)	(0.87)				(0.96)	(1.28)
Child older than 0 x Boy				1.21	-0.38				0.58	0.66				1.52	-0.34
				(1.28)	(1.50)				(1.75)	(1.26)				(1.43)	(1.44)
Child older than 0				-1.57*	0.31				-0.29	-1.40				-1.24	-0.70
				(0.85)	(1.35)				(1.31)	(0.85)				(0.93)	(1.22)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,638	4,718	8,356	2,517	2,754	2,907	3,887	6,794	2,017	2,664	4,287	1,826	6,113	3,025	1,298
Individuals	862	1,259	2,121	628	777	751	919	1,670	552	672	1,078	407	1,485	813	300
Mean of outcome	0.92	0.81	0.86	43.5	43.1	0.91	0.93	0.92	42.9	44.0	0.92	0.92	0.92	43.1	44.7

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. Observations are weighted using the a_indinus_xw weight in all regressions. In this Table, we present additional heterogeneity of the effect by indicators of socio-economic status. We focus on characteristics of the men's partners. In columns (1)-(5), we distinguish by whether the man's partner initially has a degree or not. In columns (6)-(10), we distinguish men by whether their partner is initially in a non-professional or professional occupation. In columns (11)-(15), we distinguish men whose partners are initially paid low wages from those initially paid high wages. Here, we categorise men's partners as high wage if they earn more than the 75th percentile real monthly wage (among the full sample of observations for men's partners who are employed). This is £2,628 per month (in 2015 pounds). For each demographic characteristic, we first look at the employment dummy variable, splitting the sample into two sub-samples and then testing for a difference across those sub-samples (columns (1), (2), (3), (6), (7), (8), (11), (12) and (13)). We then look at the effect on weekly hours worked on the intensive margin by splitting into the two sub-samples (columns (4), (5), (9), (10), (14) and (15)). By initially here, we refer to the first wave of the sample period for men who never have a child, and to the wave before the birth of the first child for those who do. Note, of the 7,078 men in the full sample, 2,392 initially have a partner, of whom 2,126 have a partner who is interviewed in the UKHLS in that wave. Only men whose partners are interviewed in the UKHLS and are initially employed (and not in self-employment) are included in the regressions that categorise by partner occupation and wage.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Mothers' Leisure Time with Child

	Several Times Per Week		Almost Every Day	
	(1)	(2)	(3)	(4)
	OLS	Probit	OLS	Probit
Boy	0.014	0.043	0.012	0.030
	(0.040)	(0.12)	(0.045)	(0.12)
Year Fixed Effects	Yes	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes	Yes
Observations	590	590	590	590
Mean of outcome	0.69	0.69	0.38	0.38

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the results from regressions which ask whether the gender of the first child affects the time that mothers spend engaged in leisure activities with that child. The UKHLS asks how often you and your child spend time together on leisure activities or outings outside the home, such as going to the park or the zoo, movies, sports or picnics. This question is asked in waves 1, 3 and 5 of adults with children. By definition, we cannot look at changes in behaviour before and after having a child, and so we now run cross-sectional regressions on the sample of mothers (dropping childless adults, for whom this question is not asked). We take the observation in the earliest wave after the birth of the first child for each individual. This variable is non-missing for 590 out of 744 mothers. Answers are given on a six point scale, from which we create two dummy variables. The first, used in columns (1) and (2), is a dummy variable that takes the value 1 if the parent spends leisure time with their child at least several times per week (the second highest category on the six-point scale). The second, used in columns (3) and (4) takes the value 1 if the parent spends leisure time with their child almost every day (the highest category on the six-point scale). Observations in all regressions are weighted by the a_indinus_xw weight. In all regressions, we control for year fixed effects and the age of the mother.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Mothers' Leisure Time and Child Development: UKHLS

	Mother Leisure Time			Difference by Child Gender?		
	(1)	(2)	(3)	(4)	(5)	(6)
	Cognitive	Socio-Emotional	Physical	Cognitive	Socio-Emotional	Physical
Mother Leisure	**0.14****	**0.10***	**0.086***	0.092	0.13	0.15**
	(0.063)	**(0.061)**	**(0.051)**	(0.082)	(0.087)	(0.071)
Boy	-0.29***	-0.17***	-0.077	-0.34***	-0.14*	-0.011
	(0.063)	(0.061)	(0.048)	(0.10)	(0.082)	(0.077)
Mother Leisure x Boy				0.093	-0.051	-0.12
				(0.13)	(0.12)	(0.099)
Wave Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Mother Education Controls	Yes	Yes	Yes	Yes	Yes	Yes
Mother Ethnic Group Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	902	919	919	902	919	919
Mean of outcome	-0.032	-0.0086	-0.039	-0.032	-0.0086	-0.039
Standard Deviation of outcome	0.99	0.82	0.71	0.99	0.82	0.71

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the results from OLS regressions that ask whether mothers' leisure time is correlated with child development. We look at three areas of child development. In columns (1) and (4), we look at cognitive skills; in columns (2) and (5), we look at socio-emotional skills, and in columns (3) and (6), we look at physical skills. The dependent variable in each case is the first common factor from a factor analysis across a range of indicators of the given area of child development. For every indicator underlying these three common factors, parents are asked to judge their child's development, and the questions are asked only of parents with a child aged 3 years old. These regressions are not restricted to first born children only. The common factor for cognitive skills is based on 18 indicators, the common factor for socio-emotional skills is based on 8 indicators, and the common factor for physical skills is based on 4 indicators. The key independent variable is a dummy variable that takes the value 1 if the mother engages in leisure activity with the child at least several times per week. The dependent variables are measure in waves 3 to 6 of the UKHLS, whilst mothers' leisure time is measured in waves 3 and 5, and so we pool observations from waves 3 and 5. In each regression we control for wave fixed effects, the gender of the child, the education level of the mother (dummies for degree, other higher, A-level, GCSE and other qualification, with base group of no qualification) and the ethnic group of the mother (dummies for Black, Asian and Other race, relative to white base group). Observations in all regressions are weighted using the a_indscus_xw weight, as answers are taken from the self-completion questionnaire.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: Parent Engagement and Child Cognitive Development: MCS

	Parent Engagement				Difference by Child Gender?				Difference by Parent Gender?			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Vocab	Readiness	Vocab	Readiness	Vocab	Readiness	Vocab	Readiness	Vocab	Readiness	Vocab	Readiness
Parent Engagement (Low)	**6.34*****	**9.20*****			6.91***	10.7***						
	(0.54)	**(0.62)**			(0.86)	(0.91)						
Parent Engagement (High)			**4.09*****	**5.82*****			4.69***	5.85***				
			(0.42)	**(0.40)**			(0.55)	(0.50)				
Parent Engagement (Low) x Boy					-1.00	-2.64**						
					(1.04)	(1.20)						
Parent Engagement (High) x Boy							-1.21*	-0.060				
							(0.72)	(0.66)				
Mother Read Once a Week									12.4***	16.8***		
									(1.45)	(1.75)		
Father Read Once a Week									9.12***	9.17***		
									(0.85)	(0.97)		
Mother Read Every Day											9.56***	10.4***
											(0.76)	(0.71)
Father Read Every Day											5.18***	6.65***
											(0.84)	(0.78)
Boy	-7.44***	-6.58***	-7.50***	-6.71***	-7.39***	-6.43***	-7.46***	-6.71***	-7.94***	-7.30***	-7.83***	-7.25***
	(0.58)	(0.59)	(0.57)	(0.60)	(0.58)	(0.61)	(0.57)	(0.60)	(0.66)	(0.69)	(0.65)	(0.67)
Child Ethnic Group controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Parent Education	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	14,489	13,788	14,489	13,788	14,489	13,788	14,489	13,788	9,423	9,026	9,423	9,026
Mean of outcome	48.1	56.7	48.1	56.7	48.1	56.7	48.1	56.7	51.2	60.5	51.2	60.5
St. dev of outcome	30.6	30.5	30.6	30.5	30.6	30.5	30.6	30.5	30.5	29.7	30.5	29.7
p-value Wald Test: Father Engagement = Mother Engagement									**0.088**	**0.001**	**0.001**	**0.003**

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains results from OLS regressions which ask whether parental time engagement with a child is correlated with that child's cognitive development. The dependent variable in columns (1), (3), (5), (7), (9) and (11) measures performance in the Naming Vocabulary test, which assesses the spoken vocabulary of the child. The dependent variable in columns (2), (4), (6), (8), (10) and (12) measures performance in the School readiness test, which assesses a range of other cognitive skills covering six educational concepts: colours, letters, numbers, sizes, comparisons and shapes. In both cases, the dependent variable is the percentile at which the child's age-adjusted performance lies relative to a test-specific reference distribution of children. Higher values therefore correspond to a better performance. All children are aged 3 when they sit these tests, which come from the second round of the MCS. The key independent variables are the first common factors from a set of measures of parental time engagement with children. For the Parent Engagement (Low) variable we use a factor analysis to combine measures for whether anyone in the household takes the child to the library, helps the child to learn the alphabet, learn to count, learn songs, reads to the child, and whether the child has eaten with other family members at least once in the past week. For the Parent Engagement (High) variable, we use a factor analysis to combine measures for whether any household member has helped the child to learn the alphabet, count, learn songs or read to the child every day, or taken the child to the library at least once a week. In columns (9)-(12), we have separate measures of whether the mother or father reads once a week to the child, and every day to the child. In columns (5)-(8), we allow for a separate effect of parental time engagement by child gender, and in columns (9)-(12), we allow for a difference in the effect of father and mother time engagement. We include dummies for whether the child is Black, Asian or Other race in all regressions. We control for whether the mother has a degree in columns (1)-(8), and for both whether the mother or the father has a degree in columns (9)-(12). Observations in all regressions are weighted by the bovwt2 weight.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.13: Parent Engagement and Child Socio-Emotional Development: MCS

	Parent Engagement		Difference by Child Gender?		Difference by Parent Gender?	
	(1)	(2)	(3)	(4)	(5)	(6)
	Low	High	Low	High	Low	High
Parent Engagement (Low)	**0.16***** **(0.022)**		0.17*** (0.032)			
Parent Engagement (High)		**0.12***** **(0.014)**		0.14*** (0.020)		
Parent Engagement (Low) x Boy			-0.0079 (0.044)			
Parent Engagement (High) x Boy				-0.028 (0.028)		
Mother Read Once a Week					0.25*** (0.068)	
Father Read Once a Week					0.18*** (0.037)	
Mother Read Every Day						0.16*** (0.030)
Father Read Every Day						0.078** (0.032)
Boy	-0.18*** (0.021)	-0.18*** (0.021)	-0.18*** (0.021)	-0.18*** (0.021)	-0.21*** (0.025)	-0.21*** (0.025)
Child Ethnic Group controls	Yes	Yes	Yes	Yes	Yes	Yes
Parent Education	Yes	Yes	Yes	Yes	Yes	Yes
Observations	9,008	9,008	9,008	9,008	6,031	6,031
Mean of outcome	-0.00013	-0.00013	-0.00013	-0.00013	0.053	0.053
St. dev of outcome	0.92	0.92	0.92	0.92	0.90	0.90
p-value Wald Test: Father Engagement = Mother Engagement					0.35	0.10

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains results from OLS regressions which ask whether parental time engagement with a child is correlated with that child's socio-emotional development.The dependent variable in all regressions is the first common factor from a factor analysis of 35 indicators of the child's socio-emotional development. All of the underlying indicators are dummy variables. Each relates to either a positive or a negative statement about the child's socio-emotional development. In each case, we create a dummy variable that takes the value 1 if the parent believes that a positive statement is certainly true, or a negative statement is not true. Then higher values indicate better development. All children are aged 3 when they sit these tests, which come from the second round of the MCS. The key independent variables are the first common factors from a set of measures of parental time engagement with children. For the Parent Engagement (Low) variable we use a factor analysis to combine measures for whether anyone in the household takes the child to the library, helps the child to learn the alphabet, learn to count, learn songs, reads to the child, and whether the child has eaten with other family members at least once in the past week. For the Parent Engagement (High) variable, we use a factor analysis to combine measures for whether any household member has helped the child to learn the alphabet, count, learn songs or read to the child every day, or taken the child to the library at least once a week. In columns (5) and (6), we have separate measures of whether the mother or father reads once a week to the child, and every day to the child. In columns (3) and (4), we allow for a separate effect of parental time engagement by child gender, and in columns (5) and (6), we allow for a difference in the effect of father and mother time engagement. We include dummies for whether the child is Black, Asian or Other race in all regressions. In columns (1)-(4), we control for whether the mother has a degree. In columns (5)-(6), we control both for whether the mother or the father has a degree. Observations in all regressions are weighted by the bovwt2 weight.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.14: Child Development Gender Gap

	UKHLS			MCS		
	(1)	(2)	(3)	(4)	(5)	(6)
	Cognitive	Socio-Emotional	Physical	Cognitive: Vocab	Cognitive: Readiness	Socio-Emotional
Boy	**-0.33***** **(0.058)**	**-0.20***** **(0.052)**	**-0.11**** **(0.044)**	**-7.90***** **(0.66)**	**-7.20***** **(0.70)**	**-0.18***** **(0.024)**
Mother Degree	0.13** (0.051)	0.11 (0.065)	0.023 (0.048)	14.2*** (1.09)	19.5*** (0.95)	0.44*** (0.038)
Mother Degree x Boy	0.048 (0.084)	0.036 (0.087)	0.088 (0.067)	0.30 (1.59)	-0.11 (1.43)	-0.056 (0.048)
Child Ethnic Group controls	No	No	No	Yes	Yes	Yes
Mother Ethnic Group controls	Yes	Yes	Yes	No	No	No
Wave Fixed Effects	Yes	Yes	Yes	No	No	No
Observations	1,740	1,758	1,759	14,542	13,824	9,009
Mean of outcome	0.0056	0.0047	0.0016	48.1	56.7	-0.000033
St. dev of outcome	0.93	0.80	0.67	30.6	30.5	0.92

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains results from OLS regressions which first asks whether child gender is correlated with child development, and second asks whether there is a gradient in the gender gap by socio-economic status. The regressions in columns (1)-(3) use data from the UKHLS. The dependent variable in columns (1)-(3) is the first common factor from a factor analysis across a range of indicators of the given area of child development. For every indicator underlying these three common factors, parents are asked to judge their child's development, and the questions are asked only of parents with a child aged 3 years old. In column (1), we look at cognitive development using a common factor based on 18 indicators, in column (2), we look at socio-emotional development using a common factor based on 8 indicators, and in column (3), we look at physical development using a common factor based on 4 indicators. These questions are asked in waves 3 to 6 of the UKHLS, and so we pool observations across those four waves. In each regression we control for wave fixed effects and the ethnic group of the mother (as the ethnic group of the child is not measured). Observations in columns (1)-(3) are weighted by the a_indscus_xw weight. The regressions in columns (4)-(6) use data from the MCS. The dependent variable in column (4) measures performance in the Naming Vocabulary test, which assesses the spoken vocabulary of the child. The dependent variable in column (5) measures performance in the School readiness test, which assesses a range of other cognitive skills covering six educational concepts: colours, letters, numbers, sizes, comparisons and shapes. In both cases, the dependent variable is the percentile at which the child's age-adjusted performance lies relative to a test-specific reference distribution of children. Higher values therefore correspond to a better performance. The dependent variable in columns (6) is the first common factor from a factor analysis of 35 indicators of the child's socio-emotional development, where again higher values indicate better development. All children are aged 3, as the data comes from the second round of the MCS. We include dummies for whether the child is Black, Asian or Other race in the regressions in columns (4)-(6), and observations are weighted by the bovwt2 weight. In all columns, the sample is not restricted to first born children only.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.15: Timing of the Gender of First Child Effect on Fathers' Labour Supply

	Observed with Child Aged 0 and Older		
	(1)	(2)	(3)
	Weekly Hours	Weekly Hours Intensive Margin	Employed
Child is 0 x Boy	**-5.11*****	**-2.13****	**-0.074*****
	(1.49)	**(0.99)**	**(0.027)**
Child is 0	1.00	0.17	0.012
	(0.99)	(0.79)	(0.017)
Child older than 0 x Boy	-1.40	0.57	-0.029
	(1.38)	(1.04)	(0.025)
Child older than 0	-1.73	-1.74**	-0.023
	(1.05)	(0.76)	(0.018)
Year Fixed Effects	Yes	Yes	Yes
Individual Fixed Effects	Yes	Yes	Yes
Control for Age	Yes	Yes	Yes
Observations	21,143	11,084	23,520
Individuals	6,461	3,423	6,891
Mean of outcome	26.5	40.8	0.71

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. Observations in all regressions are weighted by the a_indinus_xw weight. In this Table, we look at the timing of the effects on employment and hours worked. In all regressions, we restrict the sample to men who are observed at least once when their child is aged 0 and at least once when their child is aged older than 0. We make this sample restriction due to the unbalanced structure of the panel. When comparing the effect when the child is aged 0 to the effect when the child is aged older than 0, we wish to understand whether differences are driven by a genuine timing of the effect, or by which fathers happen to be observed at different ages of their child. In columns (1) and (2), the dependent variable is weekly hours worked. In column (2), we restrict the sample to men on the intensive margin. In column (3), the dependent variable is the dummy for employment.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.16: Validating the News Source Variable

	Wife's Role	
	(1)	(2)
	OLS	Probit
Non-Progressive News	**0.025****	**0.14*****
	(0.010)	**(0.049)**
Control for Age	Yes	Yes
Controls for Educational Attainment	Yes	Yes
Controls for Ethnic Group	Yes	Yes
Observations	6,389	6,389
Mean of outcome	0.13	0.13

Notes. Standard errors (in parentheses) are clustered at the primary sampling unit level. This Table contains the results from cross-sectional regressions which ask whether the not-progressive news source dummy variable is correlated with non-progressive attitudes about women's role in work versus home, for men. The former is measured in wave 3, and the latter in wave 4, and so the sample is restricted to men observed in both waves. In both columns, the dependent variable is a dummy variable that takes the value 1 if the man agrees or strongly agrees with the statement that 'A husband's job is to earn money, a wife's job is to look after the home and family'. In both regressions we control for the man's age, education level and ethnic group. Observations in all regressions are weighted using the a_indscus_xw weight, as the gender attitude is asked in the self-completion questionnaire.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.17: Balance Tests By Gender of First Child: News Source Sub-Sample

	Fathers		
	Daughter	Son	p-value (H0: Diff = 0)
Age	29.655 (0.636)	29.378 (0.539)	0.742
Married	0.342 (0.049)	0.262 (0.042)	0.208
Weekly Hours	39.748 (1.695)	37.466 (1.802)	0.358
Employed	0.935 (0.026)	0.901 (0.030)	0.398
Religious	0.336 (0.050)	0.377 (0.047)	0.549
Urban	0.813 (0.042)	0.822 (0.035)	0.868
Foreign Born	0.135 (0.036)	0.099 (0.025)	0.383
Ethnic Minority	0.067 (0.019)	0.069 (0.017)	0.936
Degree	0.397 (0.054)	0.459 (0.048)	0.400
GCSE only	0.209 (0.046)	0.247 (0.042)	0.542
Monthly Wages	7.702 (0.089)	7.670 (0.078)	0.789
Observations	104	126	230

Notes. Standard errors (in parentheses) are clustered at primary sampling unit level. We present the weighted sample means of each demographic characteristic for fathers with a first born son versus a first born daughter within the sub-sample of 230 fathers for whom the non-progressive news source variable is defined. We then present p-values for tests of differences in those weighted sample means across the gender of the first child. We consider the characteristics of fathers in the first wave of the dataset, which is always before the birth of the first child. We weight observations using the a_indinus_xw sampling weight as in the regression analysis. We consider the natural logarithm of the real monthly wage.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.18: Balance Tests By Gender of First Child: Fraction of Chores Sub-Sample

	Fathers		
	Daughter	Son	p-value (H0: Diff = 0)
Age	30.292	29.913	0.595
	(0.527)	(0.489)	
Married	0.440	0.341	0.098*
	(0.044)	(0.040)	
Weekly Hours	40.203	39.573	0.770
	(1.361)	(1.666)	
Employed	0.935	0.909	0.426
	(0.021)	(0.026)	
Religious	0.415	0.381	0.577
	(0.045)	(0.041)	
Urban	0.798	0.812	0.769
	(0.037)	(0.032)	
Foreign Born	0.147	0.100	0.232
	(0.033)	(0.023)	
Ethnic Minority	0.054	0.070	0.433
	(0.014)	(0.015)	
Degree	0.457	0.501	0.494
	(0.046)	(0.043)	
GCSE only	0.181	0.190	0.864
	(0.036)	(0.034)	
Monthly Wages	7.792	7.720	0.445
	(0.070)	(0.062)	
Observations	140	158	298

Notes. Standard errors (in parentheses) are clustered at primary sampling unit level. We present the weighted sample means of each demographic characteristic for fathers with a first born son versus a first born daughter within the sub-sample of 298 fathers for whom the fraction of chores variable is defined. We then present p-values for tests of differences in those weighted sample means across the gender of the first child. We consider the characteristics of fathers in the first wave of the dataset, which is always before the birth of the first child. We weight observations using the a_indinus_xw sampling weight as in the regression analysis. We consider the natural logarithm of the real monthly wage.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.19: Data Appendix

Variable Name	Data Sources	Method
Employed	jbhas, jboff, jbstat (UKHLS)	Dummy variable taking value 1 if individual is employed and 0 if they are not. We follow the ONS definition of employment. An individual is employed if they either did paid work last week (jbhas = 1), or if they did not do paid work last week but have a paid job (jboff = 1). An individual is also employed if they are on a government training scheme, or are doing unpaid work for a family business (jbstat = 9 or 10).
Weekly Hours	jbhrs, jbot (UKHLS)	We sum the weekly hours an individual is expected to work in a normal week (jbhrs) and the number of overtime hours they work in a normal week (jbot). Both variables are defined only for those in paid employment. Therefore, our weekly hours variable takes value 0 if an individual is in neither paid nor self-employment, a positive value if they are in paid employment, and missing values for the self-employed.
Weekly Hours: Including Self-Employed	jbhrs, jbot, jshrs (UKHLS)	For individuals in paid employment, we take the sum of jbhrs and jbot as before. For individuals in self-employment, we use the variable jshrs.
Weekly Hours: Including Second Jobs	jbhrs, jbot, jshrs, j2hrs (UKHLS)	We define this variable for all individuals who are in paid or self-employment. We take the 'Weekly Hours: Including Self-Employed' variable, and add to it hours worked in a second job. The variable j2hrs measures hours worked per month in a second job, so we first divide this by 4 before adding to the previous measure.
Real Monthly Wages	paygu_dv (UKHLS), All Items CPI Index (ONS)	We take gross monthly wages (paygu_dv) and convert into real wages (base year 2015) using the All Items CPI Index. This variable is missing for any individual who is not in paid employment. We then take the natural logarithm of the real monthly wage.
Child	lprnt, relationship_dv, alwstat (UKHLS)	We use the lprnt variable in the first wave of the dataset, which asks whether an individual has ever had a biological child before the start of the sample period. We restrict the sample to individuals who have never had a biological child. There is no variable in the UKHLS that cumulatively tracks the total number of children an individual has ever had. We must create this ourselves from the household grid file 'egoalt'. This file contains an observation for every pair of individuals within each household. The variable relationship_dv tells us the relationship of the ego to the alter. We use this to select only natural children of the ego. In each wave we calculate the total number of new natural children of that individual added to the household, using the variable alwstat. We count only children who are 'newborn' or 'new entrant' to the sample according to the alwstat variable. In each wave we add the number of new biological children added in that wave to the total number of biological children in the previous wave, to create a cumulative total number of children that individual has ever had. By only counting children who are 'newborn' or 'new entrant', we ensure that we do not double count children across waves. A new entrant is a new born child who has never appeared in the dataset who is older than 0 years of age. We call this new cumulative measure 'numchild'. Once we have created 'numchild', we define a dummy variable Child that takes value 1 if an individual has ever had a child and 0 if they have not.
Boy	sex (UKHLS)	We merge information on child gender from the child file to the main file of adults. We ensure that we keep only the children who are counted in the cumulative measure of biological children ever had, 'numchild', that we created above (see 'Child'). We must make this adjustment because the child file is constructed on the basis of the 'relationship' variable (rather than the corrected 'relationship_dv' variable, which we used to create 'numchild') and so incorrectly refers to some individuals as biological children when they are not. After dealing with this error, we create a variable that takes the value 1 if the first born child is a Boy and 0 otherwise. This variable is time invariant and takes value 1 even in the waves before the parent has that first child.
Child is 0	Child (Created above)	We create a dummy variable that takes value 1 if Child takes value 1 and the first born child is aged 0.
Child older than 0	Child (Created above)	We create a dummy variable that takes value 1 if Child takes value 1 and the first born child is aged older than 0.
Married	mlstat, mstatch (UKHLS)	We use the variables mlstat and mstatch to create a variable that tells us an individual's current marital status. The mstatch variables tell us whether the marital status reported in the first wave of the dataset in mlstat has changed in each wave. We create a dummy variable that takes value 1 if the individual is married and 0 if they are not. For the heterogeneity analysis by marital status, we consider the marital status of the individual in the 'initial wave' of the dataset. We take the initial wave as the first wave of the dataset for individuals who never have a child, and the wave before the birth of the first child for those who do. We do this because for any characteristic that could change over time, we want the most recent information for individuals just before they have their first child, for those who have a child. For example, an individual may start the sample period without a partner, meet someone during the sample period, marry and then have a child, and so to treat them as unmarried would be misleading. We use this definition of 'initial' for each of the demographic characteristics in the heterogeneity analysis.

Data Appendix: Continued

Variable Name	Data Sources	Method
Single	ppno (UKHLS)	The variable ppno takes value 0 if an individual does not have a partner living in the household with them. We create a dummy variable that takes value 1 if the individual does not have a partner living in the household with them (they are single), and 0 otherwise.
Add Child	numchild (Created above)	We use the variable numchild created as a first step in creating the variable Child. From this, we define a new variable that takes value 1 if the number of children you have ever had in the current wave is greater than the number of children you had ever had in the previous wave, which itself was greater than 0. This variable is an impulse dummy variable for whether you have added at least one child since the previous wave, counting only children after the first born child.
Black, Asian, Other Race	racel_dv	We use the variable racel_dv, which captures an individual's race, to define dummy variables for Black, Asian, and Other Race. For Black, we include individuals who are 'White and Black Caribbean', 'White and Black African', 'Caribbean', 'African', and 'any other Black background'. For Asian, we include individuals who are 'White and Asian', 'Indian', 'Pakistani', 'Bangladeshi', 'Chinese', and 'any other Asian background'. For Other Race, we include individuals who are 'any other mixed background', 'Arab', 'any other ethnic background', and 'Gypsy or Irish traveller'.
Birthweight	bwtlb, bwtoz, bwtk (UKHLS)	We convert the birthweight of newborn children to ounces, as some report in kilograms. These variables are contained in the 'newborn' data files. We then merge this information to their parents for only the first born children. Birthweight is not observable for all new born children.
Time Since	numchild (Created above)	We create a cumulative count of the number of years since an individual last had a child. We start this count from when the individual was aged 15. The 'clock' starts again from value 1 in the wave after a child is born.
Parity	numchild (Created above)	We create a variable that tells us the birth parity of the parent. If a parent has never had a child, their birth parity is 1. If they have had one child, their birth parity is two, and so on.
Partner Variables	hidp, ppno (UKHLS)	We create many of the same variables that we do for each individual for their partner. We merge partner information in a given wave using the household identifier and the person number of the partner.
Degree	hiqual_dv (UKHLS)	We create a dummy variable that takes value 1 if an individual has degree level education, i.e. if hiqual_dv = 1. For the heterogeneity analysis by education level, we consider the education level of the individual in the initial wave of the dataset. We take the initial wave as the first wave of the dataset for individuals who never have a child, and the wave before the birth of the first child for those who do.
Non-Professional	jbsoc00_cc (UKHLS)	We create a dummy variable that takes value 1 if an individual is in a non-professional occupation, and 0 otherwise. To do so, we use the Standard Occupational Classification 2000, produced by the ONS. We treat the following three one-digit codes as 'professional': Managers and Senior Officials, Professional Occupations, Associate Professionals and Technical Occupations. We treat the following six one-digit codes as 'non-professional': Administrative and Secretarial Occupations, Skilled Trades Occupations, Personal Service Occupations, Sales and Customer Services Occupations, Process, Plant and Machine Operatives, and Elementary Occupations. For the heterogeneity analysis by occupation, we consider the occupation of the individual in the initial wave of the dataset. We take the initial wave as the first wave of the dataset for individuals who never have a child, and the wave before the birth of the first child for those who do.
High Wage	Real Monthly Wages (Created above)	We define a dummy variable that takes value 1 if an individual lies in the top quartile of real monthly wages, and 0 otherwise. When we distinguish men by their own wages, the 75th percentile of the real monthly wage distribution is £2,723. When we distinguish men by their partners' wages, the 75th percentile of the partners' real monthly wage distribution is £2,628. For the heterogeneity analysis by wage level, we consider the wage level of the individual in the initial wave of the dataset. We take the initial wave as the first wave of the dataset for individuals who never have a child, and the wave before the birth of the first child for those who do.
Change Occupation	jbsoc00_cc (UKHLS)	We create an impulse dummy variable that takes value 1 if the occupation code in the current wave differs from the occupation code in the previous wave. This tells us whether the individual has change occupation since the last wave.
Change Industry	jbsic07 (UKHLS)	We create an impulse dummy variable that takes value 1 if the industry code in the current wave differs from the industry code in the previous wave. This tells us whether the individual has change industry since the last wave.
Change Either	Change Occupation, Change Industry (Created above)	We create an impulse dummy variable that takes value 1 if either of the variables Change Occupation or Change Industry take value 1. This is a dummy variable that tells us whether an individual changed occupation or industry since the previous wave.

Data Appendix: Continued

Variable Name	Data Sources	Method
Cognitive Development (UKHLS)	cdvla, cdvlb, cdvlc, cdvld, cdvle, cdvlf, cdvlg, cdvlh, cdvli, cdvlj, cdvll, cdvln, cdvlo, cdvlp, cdvlq, cdvlr, cdvls, cdvlt (UKHLS)	The UKHLS contains eighteen subjective measures of child's cognitive development, as answered by the child's parent. The UKHLS asks parents, for a list of skills, to indicate whether your child can now do these things. There are three possible responses: yes, to some extent, and no. The majority of children can do each of the skills, and so we create a dummy variable that takes the value 1 if the parent responds yes, and 0 if the parent responds: to some extent, or no. We then take the first common factor across these eighteen dummy variables. A more positive value indicates a higher level of development. The eighteen skills are: Can your child: understand brief instructions such as 'go and get your shoes', form sentences with at least two words, speak in full sentences, listen attentively to a story for five minutes or longer, pass on simple messages such as 'dinner is ready', use a spoon to eat without assistance and without dripping, blow nose without assistance, use the toilet to do a number two, put on pants and underpants the right way round, brush teeth without assistance, open doors with the door handle, cut paper with scissors, paint/draw recognisable shapes on paper, call familiar people by name, participate in games with other children, get involved in role-playing games, show a special liking for particular playmates or friends, call own feelings by name. These questions are asked from the third to the sixth wave.
Socio-Emotional Development (UKHLS)	cd3pera, cd3perb, cd3perc, cd3perd, cd3pere, cd3perf, chpat, chdelay (UKHLS)	The UKHLS contains eight subjective measures of a child's socio-emotional development, as answered by the child's parent. For the first six (cd3pera-cd3perf), the UKHLS asks parents to what extent they agree with the following statements about the child, where they may respond: agree completely, agree somewhat, disagree somewhat, disagree completely. For each, we create a dummy variable that take the value 1 for the most positive response, and 0 otherwise. For a positive statement (cd3pera, cd3perd, cd3pere, cd3perf), the dummy variable takes the value 1 if the parent agrees completely. For a negative statement (cd3perb, cd3perc), the dummy variable takes the value 1 if the parent disagrees completely. The statements are: child is usually happy and content, child is easily irritated and cries frequently, child is difficult to comfort when crying, child is curious and active, child is communicative and likes to talk, child shows empathy when others are sad. For the variable chpat, the UKHLS asks whether the child is generally an impatient child or a child with a lot of patience. They respond on a seven-point scale. From this, we create a dummy variable that takes the value 1 if the parent gives the child the highest rating: 7 (very patient). For the variable chdelay, the UKHLS asks whether the child is not at all or very impulsive, and again responses are on a seven point scale. We create a dummy variable that takes the value 1 if the parent gives the child the highest rating: 1(not at all impulsive). We then take the first common factor from across these eight dummy variables, where a more positive value indicates a higher level of development. These questions are asked from the third to the sixth wave.
Physical Development (UKHLS)	cddis, cd3perg, cdvlk, cdvlm (UKHLS)	The UKHLS contains four subjective measures of a child's physical development, as answered by the child's parent. The first, cddis, asks whether the child has long-term health conditions that have been diagnosed by a health professional. We create a dummy variable taking the value 1 if not. The second, cd3perg, asks parents whether they agree with the statement: I am worried about my child's health. We create a dummy variable that takes the value 1 if they disagree completely. The third, cdvlk, asks whether the child can walk forwards down the stairs. We create a dummy variable that takes the value 1 if they respond: yes. The fourth, cdvlm, asks whether the child can climb up playground equipment and other high playground structures. We create a dummy variables that takes the value 1 if they respond: yes. We then take the first common factor from across these four dummy variables, where a more positive value indicates a higher level of development. These questions are asked from the third to the sixth wave.
Cognitive Development (MCS)	bdbasp00, bdsrcp00 (MCS)	The MCS contains two objective measures of a child's cognitive performance. The first is the British Ability Scales (BAS) Naming Vocabulary test. The test is intended to reflect the child's expressive language skills, vocabulary knowledge of nouns, ability to attach verbal labels to pictures, general knowledge, general language knowledge, retrieval of names from long-term memory and level of language stimulation. The MCS measures the percentile above which the child's performance lies in a reference distribution (used more generally for the BAS test). This measure is age-adjusted. The second is the Bracken School Readiness score. This is intended to measure a child's development in six basic concepts: colours, letters, numbers/counting, sizes, comparisons and shapes. Again, the MCS measures the percentile above which the child's performance lies in a reference distribution (used more generally for the BSRA test). This measure is again age-adjusted.

Data Appendix: Continued

Variable Name	Data Sources	Method
Socio-Emotional Development (MCS)	bmsews, bmseht, bmseao, bmseuq, bmsedt, bmsena, bmsdpf, bmsdsr, bmsdor, bmsdhu, bmsdgf, bmsdlc, bmsdky, bmsdvh, bmsdst, bmsdte, bmsems, bmseoe, bmseef, bmseia, bmsdro, bmsdhs, bmsdtt, bmsdsp, bmsdmw, bmsdfs, bmsdfb, bmsdud, bmsddc, bmsdnc, bmsdoa, bmsdpb, bmsdcs, bmsdgb, bmsdfe (MCS)	The MCS contains thirty-five subjective measures of a child's development, as answered by the main respondent (typically the child's mother). In each case, the parent is asked whether a given statement is true. They can answer on a four-point scale: not true, somewhat true, certainly true, can't say. There are sixteen positive statements, for which we create dummy variables that the value 1 if the respondent answers: certainly true. There are nineteen negative statements, for which we create dummy variables that take the value 1 if the respondent answers: certainly not true. We then take the first common factor from across these thirty-five dummy variables, where a more positive value indicates higher development. The statements are as follows: the child: likes to work things out for self, shows wide mood swings, does not need much help with tasks, gets over excited, chooses activities on their own, is easily frustrated, gets over being upset quickly, persists in the face of difficult tasks, can move on to a new activity after finishing a task, is impulsive, is considerate of other people's feelings, is restless, overactive and can't stay still for long, often complains of headaches, stomach-aches or sickness, shares readily with other children, often has temper tantrums, is rather solitary (tends to play alone), is generally obedient, has many worries, is helpful if someone is hurt, upset or feeling ill, is constantly fidgeting or squirming, has at least one good friend, often fights with other children or bullies them, is often unhappy, down-hearted or tearful, is generally liked by other children, is easily distracted, is nervous or clingy in new situations, is kind to younger children, is often argumentative with adults, picked on or bullied by other children, often volunteers to help others, can stop and think before acting, can be spiteful to others, gets on better with adults than with other children, has many fears, sees tasks through to the end.
Parent Engagement (MCS)	bmtoli, bmalph, bmnumb, bmsong, bmeatw, bmofre, bpofre, bmofli, bmofab, bmofco, bmofso (MCS)	The MCS contains several measures of parental time engaged with the child. We use these to create two variables. We first create a low-level parental engagement measure. The UKHLS asks whether anyone at home ever does four activities with the child: take the child to the library, help the child learn the alphabet, learn counting or teach the child songs, poems or nursery rhymes. In each case, we create a dummy variable that takes the value 1 if someone in the household has ever done that activity with the child. It also asks whether or not the child has eaten with anyone in the household in the past week. We create a dummy variable that takes the value 1 if they have. Finally, it asks both the main respondent and their partner whether they personally ever read to the child. We create a dummy variable that takes the value 1 if either ever reads to the child. We then take the first common factor across these six dummy variables. Second, we create a high-level parental engagement measure. For each of the five activities above (library, alphabet, counting, songs and reading), the MCS asks how often someone in the household does that activity with the child. For the library measure, the responses are: on special occasions, once a month, once a fortnight or once a week. We create a dummy variable that takes the value 1 if someone in the household takes the child to the library once a week. For the alphabet, counting and songs measures, the responses are: less than once a week, 1-2 days per week, 3 times a week, 4 times a week, 5 times a week, 6 times a week, or 7 times a week. We create a dummy variable in each case that takes the value 1 if someone in the household does the activity with the child every day (7 times a week). For the reading measures, the responses are: every day, several times a week, once or twice a week, once or twice a month, less often, not at all. We create a dummy variable that takes the value 1 if either the main respondent or partner reads to the child every day. We then take the first common factor across these five dummy variables. Whilst most of the underlying indicators of both of our measures refer to 'anyone in the household', it is likely that the person doing the activity with the child is the parent. If not (e.g. an older sibling), that individual is effectively playing the role of a parent, which is what we are interested in when trying to understand the correlation between time investment and child development.
Father Leisure, Mother Leisure (UKHLS)	socialkid (UKHLS)	The UKHLS asks: how often do you and your child spend time together on leisure activities or outings outside the home such as going to the park or zoo, going to the movies, sports or have a picnic? Parents can respond on a six-point scale: never or rarely, once a month or less, several times a month, about once a week, several time a week, almost every day. We create two dummy variables. The first takes the value 1 if the parent responds: several times a week or almost every day, and 0 otherwise. The second takes the value 1 if the parent responds: almost every day, and 0 otherwise. These questions are asked only in the first, third and fifth wave of the UKHLS. We take the observation from the first wave after the parent had his first child. If the child first appears with his parent in the second or third wave of the UKHLS, we take the parent's response in wave 3. If the child first appears with his parent in the fourth or fifth wave of the UKHLS, we take the parent's response in wave 5. We are therefore measuring parents' responses in the same wave or one wave after their child first appears in the dataset, and so when the child is very young (in the first or second year of the child's life typically). This coincides with when the main gender of first child effects are observed. We cannot include parents who have their first child in the sixth wave of the data, or parents who are missing in the waves in which this question is asked.

Data Appendix: Continued

Variable Name	Data Sources	Method
Not-Progressive News, Fraction Chores	paperm2, tvm2, netm2, hubuys, hufrys, humops, huiron, hupots, hudiy (UKHLS)	We create two measures of the progressiveness of a man's gender attitudes. The first is a measure of his news sources. The UKHLS asks which sources provide you personally with the most information about news and current affairs. If the respondent mentions tv, newspapers (national or local), or the internet, they are asked a follow up question about which one tv channel, newspaper or website is the source they use most frequently. We create a dummy variable that takes the value 1 if the respondent says that the newspaper or website version of The Sun, The Daily Mail, The Daily Express or The Daily Star is their most frequently used news source. The dummy variable takes the value 0 if the respondent uses either tv, newspapers or the internet to receive their news and current affairs, and for no medium mentions one of these four organisations as their main source. This variable is measured in the third and sixth waves only. For childless men we take their response in the third wave, and for the fathers with a child who first enters the dataset in waves four to six, we take their response in the third wave. We cannot include fathers who have their first child before the third wave, as their news source could respond to the gender of their first child. The second measure is the fraction of six household chores that a man does mostly himself. The UKHLS asks individuals with partners who mostly does each of six household chores: grocery shopping, cooking, cleaning, washing/ironing, gardening and DIY jobs. The individual may respond: mostly self, mostly spouse/partner, shared, paid help, other. We create a dummy variable that takes the value 1 if the man responds: mostly self. We then calculate the fraction of these six chores that he does mostly himself. These questions are asked in the second, fourth and sixth waves of the data. For childless men, we take their response in the second wave. For fathers, we take their response in the first wave in which the question is asked before they have their first child (again to ensure that their response is not affected by the gender of their first child). For fathers whose first child first appears in the UKHLS in the third or fourth wave, we take their response in the second wave. For fathers whose first child first appears in the UKHLS in the fifth or sixth wave, we take their response in the fourth wave. Therefore, we take fathers' responses as close to (and before) the birth of their first child as possible.
Wife's Role	scopfamf (UKHLS)	The UKHLS directly measures the gender attitude of interest in the fourth wave of the dataset. We cannot use this for the heterogeneity by progressiveness analysis, as we could only include fathers whose first child first appears in the fifth or sixth wave of the dataset to ensure this measure does not respond to the gender of their first child (which leaves too small a sub-sample). We use it as a validation for the news source measure. The UKHLS asks respondents whether they personally agree with the statement: A husband's job is to earn money, a wife's job is to look after the home and family. Responses are on a five point scale: strongly agree, agree, neither agree nor disagree, disagree, strongly disagree. We create a dummy variable that takes the value 1 if the man agrees or strongly agrees.

2.B Calculations to Aggregate Effects

In this Section, we present several calculations to aggregate the effects we have estimated. In Section 2.B.1, we calculate the total gender of first child effect on weekly hours in the first year of the child's life, to compare our results to those in the existing literature in Section 2.3. In Section 2.B.2, we aggregate across the gender of first child effects for fathers and their partners, to calculate the total additional hours of parental time a first born son can expect to receive, compared to a daughter, over the first five years of their lives. Throughout the calculations, we use point estimates from any result from the regressions that is statistically significant at least at the 10% level.

2.B.1 The Total Gender of First Child Effect on Fathers' Hours Worked in the First Year of the Child's Life

The gender of first child effect on fathers' weekly hours in the full sample of men is transitory and lasts for just the first year of the child's life, as documented in Table A.15. We use the results from the event study presented in Table A.1. In the first year of the child's life, fathers work 4.9 fewer hours per week after a first born son rather than a daughter. The base year for these regressions is the wave before the birth of the first child, and so the gender of first child effects upon which we make the following calculations are relative to the gender of first child effect the wave before the birth of the first child. The UKHLS does not measure the number of weeks that an individual works per year. In the UK, statutory paid leave is six weeks per year. We use this to make the assumption that men work forty-six weeks per year. Then:

Total effect in first year of child's life $= -4.90 \times 46 = -225.$

Fathers work 225 fewer hours in the first year of a first born son's life relative to a first born daughter.

2.B.2 The Total Difference in Parental Time Spent with a First Born Son compared to a First Born Daughter

We use the results from the event studies for the fathers and their partners by the fathers' degree status, presented in Table 2.5. The base year for these regressions is the wave before the birth of the first child, and so the gender of first child effects upon which we make the following calculations are relative to the gender of first child effect the wave before the birth of the first child. We make the same assumption as in Section 2.B.1 that individuals work forty-six weeks per year. We estimate gender of first child effects up until the children are aged four years old, though the effects are more vulnerable to compositional effects the older the child is. In the following calculations, we aggregate the effects across the fathers and their partners.

Degree Educated Fathers:

Degree educated fathers work 4.22 fewer hours per week after a first born son rather than a daughter in the first year of the child's life. Their partners work 4.66 fewer hours per week when the child is aged 2, and 6.37 fewer hours per week when the child is aged 3.

Total effect before the child turns four years old $= (-4.22\times 46)+(-4.66\times 46)+(-6.37\times 46) = -194.3 - 214.4 - 292.9 = -701.6$

Assuming that the additional time off work is spent engaged with the child, first born sons of degree educated fathers receive an additional 702 hours of parental time before their fourth birthday than first born daughters.

Non-Degree Educated Fathers:

Non-degree educated fathers work 5.13 fewer hours per week after a first born son rather than a daughter in the first year of the child's life. Their partners work 5.01 more hours per week when the child is aged 2.

Total effect before the child turns three years old $= (-5.13 \times 46) + (5.01 \times 46) = -236.0 + 230.3 = -5.7$

Then the first born sons of non-degree educated fathers receive only an additional 5.7 hours of parental time before their third birthday than first born daughters. The effects for the non-degree educated fathers and their partners almost exactly cancel out.

2.C Validating Indicators of the Progressiveness of Men's Attitudes Towards Women's Role in the Labour Market

We validate our categorisation of news sources statistically in Table A.16. The UKHLS specifically measures a man's attitude towards women's role in the labour market. The UKHLS asks respondents to what extent they agree with the statement that: 'A husband's job is to earn money, a wife's job is to look after the home and family'. We cannot directly use this variable as the indicator of progressiveness for the heterogeneity analysis in Section 2.4.2.2, since this question is only asked in the fourth wave of the dataset. As such we could only include fathers who have their first child in waves five or six (to ensure that his attitude does not respond to the gender of the first child), which leaves too small a sub-sample. Nonetheless, we can use it to validate our news source measure. We create a dummy variable, $Wife's\ Role$, that takes the value 1 if a man agrees or strongly agrees with this traditional attitude. We create a sample of all men who are interviewed in both the third and fourth waves (and so are asked both questions), and run cross-sectional regressions of $Wife's\ Role$ on *Non-Progressive* $News$, controlling for a man's age, educational attainment and ethnic group. We find that men who get their news and current affairs from at least one of these four sources are 2.5 percentage points more likely to agree or strongly agree that a wife's job is to look after the home and family, an effect that is statistically significant at a 5% level. This effect is large, given that only 12.5% of men agree or strongly agree. The effect is half the size of the effect of having a degree, in the opposite direction.

As additional validation of the analysis in Section 2.4.2.2, we perform balance tests on the sub-samples of fathers for whom each of the progressiveness indicators is defined, in Tables A.17 and A.18. The balance tests are very similar to those for the main sample, and in general demonstrate that first wave demographic characteristics are balanced by the gender of the first child.

Chapter 3

Do Female Political Leaders Improve Child Learning?

Abstract

I find evidence that women in the State Legislative Assembly improve learning outcomes in reading and mathematics in rural India for both boys and girls and help to reduce the gender gap in learning in mathematics, during the period from 2008 to 2014. To identify this effect, I use the fraction of seats in a district won by women in close elections against men as an instrument for the fraction of seats in that district won by women. The main results focus on the ability to read at a Grade 1 level and subtract a two-digit number from another. The most plausible mechanism for the effects is an increase in effort by political leaders seeking re-election, where additional effort from female leaders is targeted towards education. The effects are timed in election years and disappear soon after. The effects are concentrated in the worst performing Indian states according to initial gender inequality in learning.

JEL: H75, I28, O53.

Keywords: female leaders, test scores, learning, India.

3.1 Introduction

The identity of elected politicians matters for policy outcomes in India (e.g., Pande (2003); Besley et al. (2004)). The gender of elected politicians matters for school enrolment

(Beaman et al., 2012; Clots-Figueras, 2012), which is now nearly universal in India, with 96.7% of 6-14 year olds enrolled (Pratham, 2015). However, school enrolment is a means to an end: learning. Learning levels remain very low (Hanushek & Woessman, 2008; Muralidharan, 2013; Kremer et al., 2013). In 2014, 32% of Grade 5 children in rural India could not read at a Grade 1 level, whilst 48% could not subtract a two digit number from another.[1] Learning levels are worsening over time. In 2008, 19% of Grade 5 students could not read at a Grade 1 level and 29% could not do subtraction. Further, very little is learnt in each grade in Indian schools. Just one in four students who could not read at a Grade 1 level in Grade 4 have acquired that skill by Grade 5, with only one in six learning to subtract during that year.[2]

In this paper, I therefore ask whether the gender of elected politicians matters for children's learning levels in India. To measure child learning, I use reading and mathematics test scores from the Annual Status of Education Reports (ASER), collected by the non-governmental organisation Pratham over the period from 2008 to 2014. The reading and mathematics scores are ordinal, and so I investigate the probability that a child can perform a particular reading or mathematics task. For the main results, I consider the ability to read at a Grade 1 level and the ability to subtract a two-digit number from another. The ASER dataset covers children across all of rural India. I focus on children who are enrolled in school. I match the most recent State Legislative Assembly election that occurred in a child's district between one and five years before the test was taken, using information on elections for the period from January 2005 to November 2013 from the Election Commission of India. Since this means that I have a mix of lags between elections and test scores in my main results, I also separately match elections to tests taken a given number of years after the election, for lags from zero to six years.

India has a federal government structure, comprising central, state and local levels. All twenty eight states and two[3] of the seven union territories have a State Legislative As-

[1] I calculate this using the ASER data.
[2] Comparing learning levels between Grade 4 and Grade 5 students in 2014.
[3] Delhi and Pondicherry.

sembly, also known as the Vidhan Sabha. Each state is divided into many districts, which are in turn divided into several constituencies. Each constituency elects one Member of the Legislative Assembly (MLA)[4] through first-past-the-post elections. State Legislative Assembly elections occur once every five years and different states have elections in different years. To form the ruling party in the state, a political party must have a majority of constituency seats, either by itself or as part of a coalition.

In practice, responsibility for education policy and expenditure lies with the State Legislative Assembly (Clots-Figueras, 2012), which is why I focus on the State Legislative Assembly in this paper. Female leaders may affect learning in several ways. They can act as role models directly to girls (for which evidence has been found at the village level of government (Beaman et al., 2012)), and can improve parents' perception of the value of female education, which can increase parental inputs into daughters' learning. In addition, previous papers have demonstrated a stronger political preference for education amongst women (e.g. Alesina & La Ferrara (2005)). If this preference is shared by the women who select into positions of political leadership, we might expect different policy choices by a female MLA, who is more strongly motivated to improve learning. MLAs have a variety of policy levers with which to target learning. They can lobby the Department of Education in their state to direct funds to district education offices (Clots-Figueras, 2012). They can improve monitoring and accountability of schools. This lever is particularly important as poor school governance is a major barrier to improvements in school learning (Muralidharan, 2013). Even if female MLAs are only perceived to care more about education, headteachers might expect greater monitoring and accountability, and take steps to improve school performance as a result.

To identify the effect of female MLAs, I use close elections between female and male politicians as a source of plausibly exogenous variation in female political leadership. I use the fraction of seats won by a woman over a man in a close election in a district as an instrument for the fraction of seats won by women in that district, in instrumental

[4]An MLA is an elected member of the State Legislative Assembly.

variables regressions. The key assumption for identification is that women winning in close elections against men is uncorrelated with unobservables that also influence child learning. If an election between a woman and man is won by only a few votes, it is decided by whatever idiosyncratic factors influenced those last few individual voters to turn up and vote the way they did, rather than a systematic, unobservable preference for the candidate who actually won, which might be related to learning outcomes. Since the existence of a close election between a woman and man itself is not random, I control for the existence of close elections between men and women in both the first and second stages. My identification strategy follows that used in Clots-Figueras (2012).

I find that an additional woman elected to the State Legislative Assembly in a district increases the probability that a child in that district can do subtraction by 4 percentage points and can read at a Grade 1 level by 3.4 percentage points. Whilst these effects might seem small in size, they are in line with effect sizes in the existing test score literature (McEwan, 2014). For subtraction, there is additionally a reduction in the gender gap equivalent in size to over half the mean gender gap across the sample period.

The timing of the effects is striking and suggests that the most plausible mechanism is an increase in effort of political leaders seeking re-election, where female leaders target additional effort towards education. The effects are concentrated in the next election year, a lag of five years after the initial election. Further, the effects die out immediately after the next election, which is inconsistent with an alternative explanation that it simply takes time for female leaders to have an effect. The data I have does not allow me to say precisely how female leaders have an effect on learning in election years. However, I find no evidence that they hire new teachers around the time of the next election or at any other time during their reign, or that they increase school monitoring. I also find no evidence that the effects occur through remedial education programs. The effect is regionally heterogeneous, and is concentrated in the Indian states that initially performed worst in terms of gender inequality in learning, which are largely located in Northern India. It is also concentrated amongst primary school children, as we should expect given the recent focus on primary education in India at central and state levels of government.

The results must be taken cautiously, however, as placebo tests suggest that women won in close elections against men in districts where learning outcomes were already higher the year before the election. The size of the placebo test failure is only a small fraction of the size of the effect in the next election year, and so can only explain part of the results. Further, it is possible that women win in close elections in districts where women previously won in close elections, which could drive the placebo test failure. However, I cannot test this with the data I have as I do not observe enough successive elections.

In Section 3.2, I briefly outline the existing literatures on female leaders and school learning. I present the identification strategy and data in Section 3.3. I then present the main results in Section 3.4, and robustness checks in Section 3.5. In Section 3.6, I investigate mechanisms for the effects. In Section 3.7, I discuss the placebo test failure in detail. Section 3.8 concludes.

3.2 Literature Review

This paper brings together two parallel literatures, both of which focus on India. The first looks at the impact of the identity of elected politicians, both on policy choices and on outcomes for the electorate. This literature has often looked specifically at the gender identity of elected politicians. The second, a literature on the economics of education, has increasingly recognised the importance of learning outcomes rather than school enrolment alone. I extend the literature on the gender identity of elected politicians by looking at learning outcomes, following the shift in focus in the education literature.

Female political leaders in India make different policy choices to male leaders, which influences educational, health, legal and public good outcomes for their constituents (e.g. Chattopadhyay & Duflo (2004)). For education, female leaders improve girls' career aspirations, raise years of schooling for both girls and boys, and increase school construction. Effects have been found for female leaders at the village and state levels of government.

Village level analyses have used India's system of reservation of Pradhan (village council leader) positions as a source of exogenous variation in female leadership (Beaman et al., 2012). Clots-Figueras extended the literature to look at the State Legislative Assembly by using close elections between men and women as an alternative source of exogenous variation in female leadership (Clots-Figueras, 2011, 2012; Bhalotra et al., 2013). My paper contributes to the literature on female leaders in the State Legislative Assembly.

Despite high enrolment rates, little is actually learnt in school in many developing countries, including India (Pritchett, 2013). The crucial bottlenecks to child learning in India are governance and pedagogy. A body of evidence using randomised controlled trials has demonstrated that improvements to monitoring and accountability of teachers, as well as pedagogical reforms (such as remedial education programs), can improve learning (Banerjee et al., 2007; Muralidharan & Sundararaman, 2010, 2011; Duflo et al., 2012). The more effective policies increase test scores by in the region of 0.08 to 0.15 standard deviations within a year or two of the program (McEwan, 2014). Simply scaling up 'more of the same' school inputs, for example providing libraries and books, has little or no effect on learning (Kremer et al., 2013; Borkum et al., 2013).

To the best of my knowledge, only a single paper in the female leaders literature has looked at any learning outcome. I performed a search of the female leaders literature using EconLit and Google Scholar.[5] The paper by Beaman et al. (2012) looks at the effect of female leaders at a village level on reading and mathematics scores amongst nine year olds in a single district in West Bengal. In this paper, I therefore attempt to fill the gap in the female leaders literature left by the focus on enrolment rather than learning. Large datasets of learning outcomes have only recently become available, owing to the enormous effort undertaken by Pratham in creating the Annual Status of Education Reports, and no paper has yet taken advantage of these. Whilst this paper is related to Clots-Figueras (2012) in terms of the choice of the identification strategy, I study the effect of female leaders on the educational outcome that we are ultimately interested in, learning, rather

[5]For example, I search EconLit using the keywords 'female leaders' and 'women leaders', which returned 98 and 233 items respectively.

than school enrolment. In addition, I consider more recent State Legislative Assembly elections from 2005 onwards, and look at rural areas across the whole of India, whilst Clots-Figueras (2012) considers elections during the period from 1967 to 1992, and focuses on urban areas in sixteen states. Further, I look at the role of politicians' incentives for re-election in determining children's educational outcomes.

3.3 Identification Strategy and Data

3.3.1 Identification Strategy: Close Elections

I wish to identify the causal effect of the gender identity of MLAs on learning outcomes for children. Women are not elected into the State Legislative Assembly randomly. Districts in which women are elected may differ in other, unobservable, ways that matter for learning outcomes. For example, if learning levels are low, voters care about their children's education, and voters hold beliefs that female politicians are more likely to improve education, then we might expect women to be elected where learning levels are low. These unobservable differences might also matter for gender gaps in learning outcomes. If voters have more progressive attitudes towards gender equality they might hold (and vote upon) a greater desire for women to be in political leadership and put more effort at home into improving the learning levels of girls in the family. Then girls will be performing well relative to boys exactly where women are elected. To tackle these threats to identification I use an instrumental variables approach. I estimate the following equations, where equation (3.2) is the first stage.

$$Y_{idt} = \beta F_{dt} + X'_{idt}\eta + \pi_t + \mu_s + \epsilon_{idt} \tag{3.1}$$

$$F_{dt} = \gamma FC_{dt} + X'_{idt}\theta + \pi_t + \mu_s + u_{idt} \tag{3.2}$$

Here, Y_{idt} is a binary variable indicating whether child i in district d at time t attains the level of learning of interest, F_{dt} indicates the fraction of seats won by a woman in district d at time t, and FC_{dt} indicates the fraction of seats won by a woman in a close election against a man in district d at time t.

I instrument for the fraction of seats in a district won by women with the fraction of seats in that district won by women against men in a 'close election'. If an election contested by a woman and man is decided by only a few votes, it could easily have been won by the politician of the other gender. The outcome is decided by whatever idiosyncratic factors influenced those last few individual voters to turn up and vote the way they did, rather than a wider, unobservable preference for the candidate who actually won, which could be related to learning outcomes. My exclusion restriction is that women winning in close elections against men is not correlated with unobservables that directly affect learning. Whilst the outcome of close elections between men and women may be as good as random, the existence of close elections between men and women in the first place is unlikely to be. To deal with this concern, I control for the existence of close elections between men and women in both the first and second stages in X'_{idt}.

I therefore estimate a local average treatment effect for complier districts. Complier districts are districts that have a female MLA because a woman won against a man in a close election.[6] The average treatment effect in these districts might differ from other Indian districts. An additional complication arises when we try to interpret the effects at an individual level. Not all individuals who live in complier districts can be thought of as compliers. The unit of analysis is the district, and yet elections occur at the constituency level. There are multiple constituencies within each district. Therefore,

[6] Previous authors have suggested that all districts in which there are close elections between men and women are the complier districts for this LATE (e.g. Clots-Figueras (2011)). However, consider districts in which women are elected comfortably in some constituencies *and* in which women win against men in close elections in other constituencies. These districts receive the treatment regardless of the outcome of the close elections, yet they receive a stronger treatment because women won the close elections too. These districts are then a mix of 'always-taker' and 'complier' in the language of local average treatment effects, and so it is not clear whether the treatment effect is representative for them. What I can say is that the LATE is not relevant for the districts in which there are no close elections between men and women. Further, the LATE is relevant for the districts in which there are close elections between men and women, and no women win in either a comfortable election against a man or an election against another woman.

even if an individual lives in a complier district, if the female leader elected in a close election against a man is elected in a different constituency, that individual is not induced into receiving the treatment by the instrument. Indeed, if a man is elected in their constituency, they do not receive the treatment at all. As the econometrician, I can only observe that individuals living in complier districts face some probability greater than zero of being induced into the treatment by the instrument, and so are only compliers with some probability greater than zero.

With an infinite number of elections, I could use an infinitesimally tight definition of 'close election'. With the finite number of elections I observe, I face a trade-off. A tighter close election definition is a more convincing identification strategy, but also means I categorise fewer elections as 'close' and so have less variation in FC. My identification strategy follows Clots-Figueras (2012) and so I follow her definition of close elections to be consistent with the existing literature. I categorise an election as close if the winning margin between first and second place is less than 3.5% of total votes. I check the robustness of the results to 3%, 4%, 4.5% and 5% definitions.

I control for state fixed effects, μ_s, and year fixed effects, π_t, in all regressions. Additionally I control for a child's age, gender, school grade, and the type of school she or he goes to (for which there are four categories: government, private, madarsa and other) in X'_{idt}.

In Table 3.1, I check for balance in the demographic structure of the population according to the census. I compare districts in which more women beat men in close elections with districts in which more men beat women in close elections. I present balance tests for the weakest definition of close elections (5%), where we might expect the greatest chance of imbalance. I find no statistically significant differences, and the absolute size of differences is small too.

In Table 3.2, I check for balance in politicians' characteristics and features of the elections. I compare constituency seats in which women win against men in close elections with those in which men win against women in close elections. There are two statistically significant differences. Women who win are significantly younger than men who win. However, this is a broader feature of Indian politics: female candidates are on average younger than male. My concern is that female winners of close elections are more likely to be part of the BJP party than male. I therefore control for the fraction of district seats won by the BJP in all regressions in X'_{idt} to ensure that results are not driven by the party identity (rather than gender identity) of winning candidates.

I also allow for a heterogeneous effect of female leaders across female and male children. I can then understand whether female leaders have an effect on the gender gap in learning. To do so, I estimate the following equations, again using an instrumental variables approach. Equations (3.4) and (3.5) are the first stages.

$$Y_{idt} = \beta F_{dt} + \delta(F_{dt} \times Female_{idt}) + X'_{idt}\eta + \pi_t + \mu_s + \epsilon_{idt} \tag{3.3}$$

$$F_{dt} = \gamma_1 FC_{dt} + \rho_1(FC_{dt} \times Female_{idt}) + X'_{idt}\theta_1 + \pi_t + \mu_s + u_{idt} \tag{3.4}$$

$$(F_{dt} \times Female_{idt}) = \gamma_2 FC_{dt} + \rho_2(FC_{dt} \times Female_{idt}) + X'_{idt}\theta_2 + \pi_t + \mu_s + v_{idt} \tag{3.5}$$

That is, I interact the key right hand side variable, F, with an indicator for whether the child is a girl, $Female$. I add a second instrument, the interaction of the original instrument, FC, with the indicator for whether the child is a girl, $Female$.

3.3.2 Data

3.3.2.1 Outcome Variables: Test Scores

I take data on test scores from the Annual Status of Education Reports (Pratham), for the period from 2008 to 2014. A standardised test in mathematics and reading is administered to a random sample of children aged 5-16 in rural areas across all Indian states each year.[7] The learning measures for mathematics and reading are both ordinal,

[7] In each district in each year 600 households are sampled from 30 villages. Indian districts vary considerably in size. With a fixed number sampled per district, the probability a household in a large

and have four levels. From lowest to highest learning level, mathematics is graded as the ability to perform the following tasks: 1) recognise numbers from 1 to 9 (single digit recognition), 2) recognise numbers from 10 to 99 (double digit recognition), 3) do subtraction (a two digit number minus another two digit number, requiring carry over), and 4) do division (a three digit number divided by a single digit number). Reading is graded as the ability to: 1) recognise letters, 2) recognise words, 3) read a Grade 1 (first year of school) level text, equivalent to a short paragraph, 4) read a Grade 2 (second year of school) level text, equivalent to a short story. My choice of outcome variable uses this ordinal information as it is intended to be used. I focus on whether or not a child can perform a specific task. This contrasts with many other authors in the test score literature who convert ordinal test scores into z-scores as if learning were measured on a continuous interval (e.g. French & Kingdon (2010)). In the main results, I use the second highest level of ability for both mathematics and reading: subtraction and the ability to read at a Grade 1 level. I later look at the highest learning levels too. I focus on in-school children in all regressions. In the final dataset for the baseline regressions I can match over 2.5 million in-school children with non-missing values for age, school grade, school type, and test performance to elections.[8]

3.3.2.2 Treatment: Female Leaders

I take data on elections from the Election Commission of India. They have published detailed information for each State Legislative Assembly election for all twenty-eight states and two union territories (Delhi and Pondicherry) that had a State Legislative Assembly over the period from January 2005 to November 2013. For each election, I know the name, gender, constituency, district, and total votes cast for each candidate. The district is the most disaggregated geographic unit in the ASER data, and so for each district-election year, I must aggregate the elections data from the constituency to the

district is sampled is much lower than for a household in a small district. To reflect this sampling design, I weight observations inversely proportional to the probability the household is sampled in all regressions.

[8]Specifically, I match 2,535,288 and 2,544,897 children who sat maths and reading tests respectively.

district level. I calculate the fraction of constituency seats that are won by a woman in each district. For the main results, I merge each round of ASER data to the most recent election that occurred in the district between one and five years before the ASER test year. Given my election information starts from 2005, different states have elections in different years (see Table A.1), and State Legislative Assembly elections occur typically once every five years, it is only from 2010 onwards that I observe an election for every state covered by the ASER dataset. Table 3.3 outlines the number of constituency seats won by a woman for the elections matched to the ASER 2008-14 test scores included in the main results. This table therefore outlines the source of variation in the key right hand side variable F in the main results. I observe 5,777 constituency elections across 880 district-election years, across 560 districts. Of these, 7.7% are won by a woman.

3.3.2.3 Instrument: Close Elections Won by a Woman against a Man

I also use the elections data from the Election Commission of India to create the instrument. For each district-election year, I calculate the fraction of constituency seats in each district that are won by a woman against a man, in second place, in a close election. Table 3.3 outlines the number of constituency seats won by a woman against a man in close elections, for different definitions of close election. Of the 5,777 constituency seats observed in the full sample, 1.45% are won by a woman in a close election against a man at a 3.5% close election definition.

3.3.2.4 Descriptive Statistics: Learning

In Figure 3.1, I run local polynomial regressions for each of the outcomes: whether or not the child can do subtraction (Panel A), and whether or not the child can read at a Grade 1 level (Panel B), on age. There exists a small but persistent gender gap in maths across India during the period 2008-14. In the full sample, by the end of lower primary school, age 11, 69% of boys compared to 66% of girls can do subtraction. This 3 to 4% gap had opened up by age 10 and persists until age 16 (the oldest children in the sample).

National average gaps in reading are very small.

The national pattern masks considerable variation across states, as shown in Figure 3.2, where again I run local polynomial regressions of learning outcomes on age. Several Northern Indian states have wide gender gaps in both maths and reading. The gaps for maths are particularly pronounced. In Uttar Pradesh, for example, by the end of lower primary school at age 11, 57% of boys compared to 47% of girls can do subtraction. In several Southern Indian states there exist gender gaps in the opposite direction for both maths and reading. In Kerala for example, 95% of girls compared to 92% of boys can read at a Grade 1 level by the end of lower primary, age 11. Boys tend to do relatively better in maths and girls relatively better in reading across the country, as has been documented across countries.

3.4 Results

Female MLAs have a positive effect on children's learning in both maths and reading. The IV point estimates suggest that an additional seat won by a woman in the median district (the median district contains five constituency seats)[9] leads to an increase in the probability that a child can do subtraction by 4 percentage points and can read at a Grade 1 level by 3.4 percentage points (Table 3.4). Since 53% of children can do subtraction and 62% read at a Grade 1 level in the sample, these effects are equivalent to 7.5 and 5.5 percent increases from the means respectively. These effects are significant at the 1% level.

Whilst the size of the effects might seem small, the literature has demonstrated that even the most effective direct interventions in schools in randomised controlled trials have only small effects on test scores. Most existing papers create continuous measures of learning, and then discuss effect sizes in terms of standard deviations of that continuous measure. The typical effect size for an effective intervention is between 0.08 and 0.15

[9]The median district in the full sample of elections 2005-13 contains five constituency seats. Throughout the paper I therefore focus on the effect of an increase in F by 0.2.

standard deviations (McEwan, 2014), with perhaps the largest effect a 0.47 standard deviation improvement in maths test scores from a computer assisted learning program (Banerjee et al., 2007) (although even this effect reduced to 0.1 standard deviations after two years). The outcome variable I use is binary, and so it does not make sense to calculate effect sizes in terms of standard deviations and compare to previous papers with continuous measures. Nonetheless, finding only a small effect on test scores is in line with previous papers.

The coefficient on the instrument in the first stage, the fraction of district seats won by women in close elections against men, is 0.86 and is highly statistically significant. Holding constant the number of seats decided in close elections in the district, a 10% increase in the fraction of seats won by women in a close election against a man increases the number of seats in the district won by women by 8.6%.

OLS estimates of β are smaller in absolute size and not statistically significant. The difference between the IV and OLS estimates arises from two sources. First, the IV estimates remove bias in OLS estimates from the non-random election of women to the State Legislative Assembly as explained in Section 3.3.1. Women being elected may be correlated with unobservables that also influence learning outcomes for children. This is the motivation behind using the instrumental variables technique in the first place. Second, I am estimating a local average treatment effect in the IV estimates, as also explained in Section 3.3.1. There is no reason to believe the effect of a woman in power is homogeneous across India. One female leader's preferences may differ considerably from another's, and her ability to enact change in her district depends on many factors (not least the preferences of other elected politicians in her state). The local average treatment effect that I have estimated for the sub-population of compliers might be a different effect to the average effect in the overall population, which is captured by the OLS estimates.

The effect on learning occurs for both girls and boys. I run the baseline regression on the sub-samples of girls and boys separately (in Table A.2). For both maths and reading the effects are statistically significant at the 5% significance level. In Table 3.5, I allow for a heterogeneous effect by gender. I find that for maths there is a statistically significantly greater effect for girls than boys, at the 5% significance level. Female MLAs not only improve learning in maths for both girls and boys, but also reduce the maths gender gap. The IV point estimates suggest an additional woman in the State Legislative Assembly in the median district increases the probability that a boy can do subtraction by 3.3 percentage points, and a girl by 4.8 percentage points. This gender difference in the effect is equivalent to a catch up of 57% of the mean gender gap in subtraction. There is no evidence of a gender difference in the effect for reading. The coefficient on the interaction term is not statistically significant, and the point estimate is approximately one third of that in the maths regressions in absolute size.

The effects for both maths and reading are driven by primary school children and not secondary, as demonstrated in Table A.3. Primary school consists of Grades 1 to 8. I group what is known as secondary (Grades 9 and 10) and higher secondary (Grades 11 and 12) together in my definition of 'secondary' here. For maths and reading, the point estimate of β is approximately three and six times larger in magnitude for the primary school compared to secondary school children. The effects are statistically significant at the 1% level for primary school children, whilst the effects are not statistically significant for either learning outcome for secondary school children. The gender difference in the effect for subtraction is also statistically significant for the primary school children and not secondary. We should perhaps expect the effects to be concentrated amongst primary school children, because the most prominent government education initiatives over the past fifteen years in India have focussed on primary education (education for 6-14 year olds). This includes Sarva Shiksha Abhiyan, a flagship program introduced in 2000 to universalise primary education and raise school quality, and the Right to Education Act 2009 for free and compulsory education for children aged 6 to 14.[10]

[10]For details, see: http://mhrd.gov.in/school-education.

3.5 Robustness

3.5.1 Changes in the Definition of Close Elections

The main results are robust to changes in how I define close elections. Panels A and B of Figure 3.3 demonstrate that for maths the estimate of β is significant at least at the 5% level for all five close election definitions, whilst for reading this is true for four out of five close election definitions (p-value of 0.056 for the 4.5% definition). The gender difference in the effect for maths is also robust to changes in how I define close elections. Panel C of Figure 3.3 demonstrates the gender difference is significant at least at the 5% level for four out of five close election definitions.

Across all four panels of Figure 3.3, as the close election definition becomes tighter, the IV estimates of β and δ move away from the OLS estimate. This is intuitive. The tighter the close election definition, the less the estimates are influenced by the biases contaminating the OLS estimates as a consequence of the non-random election of women to the State Legislative Assembly. The confidence intervals for the effect become wider as the definition becomes tighter, as I have fewer 'close' elections and so less variation in FC (from which I am identifying the effect).

3.5.2 Inclusion of a Vote Margin Running Variable

If I had test scores at the constituency level, I could directly use a regression discontinuity design in a constituency-level analysis. The running variable would be the vote margin between the female and male candidate, and I would identify the effect of a female leader from the change in test scores as the vote margin crosses the threshold of zero. I cannot do this as I only have district-level identifiers in the test score dataset. However, I can control for the direct effect of the running variable in the district-level analysis, like I would do in a regression discontinuity design (which is therefore referred to as being 'in the spirit of regression discontinuity design' (Clots-Figueras, 2012)). In Table 3.6, I try specifications in which I control for a first, second and third order polynomial in the smallest vote margin between a female and male candidate in those districts in which there was at least one election between a woman and man (in first and second place in either order). The main results are robust. The absolute size of the estimate of β in fact increases for both subtraction and Grade 1 reading, as does the gender difference in the effect, δ, for subtraction (the estimates of β and δ increase slightly as I include higher order polynomials). Only the 363 districts with at least one election between a woman and man are included in these regressions.

3.5.3 Inclusion of Additional Control Variables

The main results are also robust to the inclusion of additional household level control variables and the interaction of state and year fixed effects (Table A.4). The former include whether the household has an electricity connection, a TV, a house made of pucca building material, and whether the child's mother had ever gone to school. The absolute size of β reduces slightly in both cases for both subtraction and Grade 1 reading, but remains statistically significant at the 1% level.

3.6 Mechanisms

In this Section, I explore possible mechanisms for the effect of female leaders on child learning. In Section 3.6.1, I present evidence consistent with an explanation that the effect is driven by the incumbent leaders' desire for re-election. Around the time of elections, politicians increase effort in the hope of getting re-elected, and for women that effort is targeted towards education. The effect of female leaders on learning is concentrated in the next election year, and dies out immediately after. In Section 3.6.2, I ask how female leaders achieve those gains in learning in election years. Finally, in Section 3.6.3, I demonstrate that the effects are concentrated in the states with the initially worst gender inequality in learning, which are predominantly located in Northern India.

3.6.1 Re-election

3.6.1.1 The Effect on Child Learning is Concentrated in the Next Election Year

If politicians care about re-election, and female leaders have a preference for education, we might expect the effect of female leaders on child learning to be concentrated around the time of the next election. Politicians increase effort around the time of elections, and for female leaders that effort is focussed on education more than male leaders. For the main results in Section 3.4, I matched learning outcomes to the most recent election that occurred between one and five calendar years before the calendar year in which a test was taken. The effects were then identified from a mix of lags between elections and tests (from one to five years inclusive). I now match elections to tests taken a fixed number of years after the election. For example, for a lag of one year, I take states with elections in 2008 and match those elections to the 2009 test scores in that state. I take states with elections in 2009 and match those elections to the 2010 test scores in that state, and so on. By contrast, for a lag of two years, I instead take the states with elections in 2008 and match those elections to the 2010 test scores in that state, and so on. I therefore run a separate regression for each lag, and I look at lags from minus two to six years.[11] The five year lag coincides with the next election year for all twenty-eight states included in this timing analysis.

To ensure that only the timing of the effects and not regional heterogeneity in the effects drives differences in the estimates across lags, I use a single election from each state wherever possible. For lags from minus one to two years, I use elections from 2008-12. Every state had exactly one round of elections across these five years. For lags from three to five years, I use elections from 2005-9.[12] Every state had at least one set of

[11] Given that test scores are only observed from each state in a single year in this setup, by controlling for state fixed effects I have effectively controlled for year fixed effects too, and so I drop those year fixed effects from the regression.

[12] I can also consider a lag of two years using elections from 2005-09, which produces a very similar insignificant result as when using the 2008-12 elections, but with larger confidence intervals.

elections 2005-9. I consider only the most recent elections for the two states (Haryana and Jharkhand) that had two sets of elections in these five years. For the lags of minus two and six years, I cannot match every state to test scores, since I only have test score data from 2008 to 2014.[13] For the lag of minus two years, I use elections from 2009 to November 2013, and can match twenty-four states.[14] For the lag of six years, I use elections from 2005 to 2008, and can match a different subset of twenty-four states.[15] For this reason, I have joined the lags minus one to five years with a dashed line in Figure 3.4, to indicate that the effects estimated for these lags are based on the same group of states.[16]

Figure 3.4 demonstrates a similar pattern for both maths and reading. There is no statistically significant effect of female MLAs at any lag from zero to four years. However, in the next election year there is a statistically significant effect at the 5% level for maths and reading. At the five year lag the estimates of β jump to 0.72 for maths and 0.68 for reading. These coefficients are larger than the main results by a factor of 3.6 and 4.0 respectively. An additional woman in the State Legislative Assembly increases the probability that a child can do subtraction by 14 percentage points, and can read at a Grade 1 level by 14 percentage points, five years after the year she is elected (although confidence intervals are also wider here than the main results). The gender difference in the effect for maths also exists at a lag of five years (see Column (1) of Table A.5), with the point estimate for δ taking the value 0.12, significant at the 5% level. Test scores are also higher the year before the election, which I discuss in detail in Section 3.7.

This evidence suggests that a considerable proportion of the effect arises after five

[13] I can also use test scores from 2007 for the lags of minus one and minus two. I did not use test scores from 2007 in the main analysis as very few test scores in that year could be matched to past elections, given I only observe elections as far back as 2005. To include this year in the main results would have distorted the regional balance of the sample towards those states that had elections in 2005 and 2006. I do not observe test scores before 2007.

[14] Chattisgarh, Jammu & Kashmir, Madhya Pradesh, Mizoram and Rajasthan are missing from the estimates for the lag of minus two.

[15] Andhra Pradesh, Arunachal Pradesh, Orissa, Maharashtra and Sikkim are missing from the estimates for the lag of six years.

[16] Since Jammu & Kashmir is missing from the 2010 ASER data, and therefore would be missing from specific lags, I drop Jammu & Kashmir from regressions at all lags from minus one to five years, to ensure regional comparability across those lags.

years. I test this claim directly. I run the baseline regression again, but now match test scores to the most recent election that occurred between zero and four calendar years before the year the test was taken (rather than one to five as in the main results). I find that removing the five year lag in this way nearly halves the point estimates for β (to 0.12 for maths and 0.09 for reading), and they are now only significant at the 5% and 10% significance levels respectively (see columns (2) and (3) of Table A.5).

An alternative explanation to the re-election mechanism is that it simply takes time for female leaders to have an effect on learning. If that is the case, we would expect the effect to persist in the next year after the election. However, for the lag of six years, the point estimate for β is 0.01 and 0.20 for maths and reading respectively, neither of which are significant at even the 10% level.[17] The effect of female leaders at a five year lag disappears the following year, which is difficult to reconcile with the hypothesis that learning gains at the five year lag accrued gradually from five years of effort from the female leader.[18]

Is it plausible that test scores could increase and decrease so quickly? The existing literature on learning suggests so. Levitt et al. (2012) show that there exists considerable slack in test score performance. They find that providing previously unannounced and small financial incentives at the time of low stakes tests can cause large improvements in test score performance, particularly in maths, as children become motivated to put in effort.[19] It is conceivable that female leaders find (or motivate schools to find) ways of removing slack in test score performance in these low stakes tests in election years. However, with that in mind, and given that the test score improvements disappear the year after the election, it is not clear that we can interpret the effects of female leaders

[17]Further, when I run the five year lag regressions again on the same subset of states that are included for the six year lag, I still get a positive and significant effect at the 5% level for both maths and reading. In fact, the point estimates of β are larger than when the excluded states were included. This suggests that the lack of effect at a six year lag is not driven by the exclusion of particular states from that analysis.

[18]The Kleibergen-Paap F statistic is 7.6 and 7.5 for maths and reading respectively for the six year lag regressions, and so I am concerned that my instrument is weak in this sub-sample.

[19]They provide $20 incentives to Chicago school children to improve on a recent test score, and find an increase in average test score performance by one tenth of one standard deviation, a large effect in the learning literature.

estimated here as evidence of a deeper, lasting impact on child learning.

3.6.1.2 Is the Effect Concentrated Where Women Stand for Re-election?

If the effect is driven by the additional effort of politicians seeking re-election, then we should expect to see the effect concentrated in the districts where female leaders stand for re-election. I can match female leaders across elections, since I have candidate names.[20] One difficulty here is that I can only use test scores matched to the first round of elections from each state, as it is only the first round of elections for which I observe whether the woman ran in the *next* election. In addition, for some states I do not observe more than one round of elections.[21] This reduces my sample size substantially.

I define a dummy variable, *Stood*, that takes the value 1 if any female leader in the district ran for re-election in the next round of elections in any constituency in that same district. I control for this variable directly, and interact it with the main right hand side variable, F. I create a second instrument as the interaction between *Stood* and the original instrument, FC. I therefore test whether the effect of female leaders is greater in districts where they stand for re-election. In the district-election years for which I observe a subsequent election, there are 223 female leaders. Of those, 60.1% (134) stood for re-election. These 134 women stood in 106 different districts. The fact that such a high fraction of female leaders run for re-election lends credibility to the re-election mechanism. Of those who stood for re-election, 48.5% (65) won re-election, across 59 different districts.

Unfortunately, the regression results are not conclusive. In Table A.6, I first run the baseline regression for the sub-sample of observations which I can match to the first round

[20] There are no candidate ID numbers in the data, and so to match candidates across elections I have to use candidate names. Small spelling differences across years means that I must do this process by hand. For example, 'Promila Rani Brahma' in Kokrajhar, Assam in 2006 is spelt 'Pramila Rani Brahma' in the same district in 2011; and 'Poongothai Aladiaruna' in Tirunelveli, Tamil Nadu in 2006 is recorded as 'Dr. Poongothai Aladi Aruna' in the same district in 2011.

[21] I use elections data from 2014 and February 2015, which is the most recent published by the ECI (but which I have not used for the main analysis, as I only have test scores up until 2014). However, for the elections in December 2013, district identifiers have not been published, and so there are four states for which I only have one round of elections data. These are: Chattisgarh, Madhya Pradesh, Mizoram and Rajasthan.

of elections in districts in which I observe a second round of elections, to see if the effect still exists in that sub-sample. The point estimates are larger for both maths and reading (columns (1) and (4)) than in the baseline regressions, but less precisely estimated. The effect is still statistically significant at the 10% level for reading (p-value 0.073), but not for maths (p-value 0.175). In spite of the lower precision, there is some suggestion that the effects do exist in this sub-sample, and so I proceed to allow for a difference in the effect in the districts where women stand for re-election in columns (2) and (5).

Whilst the point estimates of the coefficient on $F \times Stood$ are positive for both maths and reading, consistent with the re-election mechanism, neither are statistically significant and the standard errors are large. The reason is the low explanatory power of the instrument in the sub-sample of districts where women run for re-election. The SW F-statistic for the first stage regression of $F \times Stood$ is just 4.0. Further, when I run the baseline regression in the sub-sample of districts where I observe that female leaders stand for re-election, the KP F-statistic is just 0.86. There are simply not enough close elections between women and men in the sub-sample of districts where I observe that female leaders run for re-election, and this is true even when I use a less tight definition of close elections (the 5% definition).

In columns (3) and (6), I allow for a difference in the effect in the districts where a female leader won re-election. If winning the election is an indicator that the politician put in a particularly high level of effort, we would expect the effects to be even stronger where the women won re-election. The point estimates for the coefficients on $F \times Won$ are positive and larger than on $F \times Stood$, as predicted. However, again the standard errors are very large owing to the limited explanatory power of the instrument in the districts where women won re-election, and so I cannot say anything conclusive.

3.6.1.3 Ruling Parties

If the effect is driven by the additional effort of politicians seeking re-election, then we might expect the effect to be greatest where the female leaders are part of the ruling party in their state. The Department of Education, controlled by the ruling party in the state, might offer more funding and support to schools in the districts where the ruling party's own representatives are in power and seeking re-election. I use a dataset recently compiled by Asher and Novosad (Asher & Novosad, 2015),[22] which outlines which parties form the ruling coalition (or outright majority party in that event) in the state. Information is available for all but four state elections covered by my sample (Goa and Manipur 2012, Meghalaya and Nagaland 2013). I create a dummy variable, *Ruling*, that takes value one if at least one female MLA in the district is a member of any of the ruling parties in that state. I interact this with F, and add both *Ruling* and $F \times Ruling$ to the baseline regression specification. As a second instrument, I interact *Ruling* with the original instrument, FC.

Table A.7 demonstrates that the coefficient on $F \times Ruling$ is indeed positive and large for both maths and reading. However, the coefficient is not statistically significant and is not precisely estimated. The problem here, as was true for the results in Section 3.6.1.2, is that the instrument does not have much explanatory power in the sub-sample of districts where female leaders are part of the ruling party. Women do not often win in close elections against men in districts where they are part of the ruling party, in my sample. The KP F-statistic in a regression on the sub-sample of districts in which female leaders are part of the ruling party is just 2.7. Therefore, unfortunately the evidence is not conclusive.

3.6.2 How Do Female Leaders Increase Learning in Election Years?

In Section 3.6.1, I showed that female leaders improve test scores in the next election year. In this Section, I ask how. There is little evidence that female leaders increase school inputs or monitoring, or that the effect occurs through remedial education programs.

[22] I am very grateful to Sam Asher for sending me this dataset.

3.6.2.1 School Inputs and Monitoring

Do school inputs increase after women are elected to the State Legislative Assembly? The National University of Educational Planning and Administration has collected school report cards, measuring a wide range of school inputs from the 2005-6 school year until the 2013-4 school year. This dataset is called the District Information System for Education (DISE). It is intended to be a census of all schools in the country. A similar dataset is available for secondary schools, but I focus only on primary schools, where the learning effects are driven. I consider only schools in the same districts as covered by the ASER dataset that I use in my main results.[23] Within these districts I consider only schools in rural areas, again since the ASER dataset focuses on children in rural areas only.

I match schools to elections using the same timing as described in Section 3.6.1.1. I am particularly interested in whether there is an effect in the next election year. For comparability across lags, I use elections from 2005 to 2009 (when each state had one set of elections)[24] for all lags from one to five years. I can also use elections 2008-12 for a lag of one year.[25] I adopt the same identification strategy as in the main results, but I replace child control variables in X with school control variables. Specifically, I control for whether the school is all-boys or all-girls, a set of dummies indicating the category of grades taught at the school,[26] and a set of dummies indicating the school management type. There are small and unavoidable differences in the states included at each lag such that the estimates across lags are not perfectly comparable.[27] All regressions are

[23] Not all districts are covered by both ASER and DISE. There are 16 districts I cannot match from the DISE dataset that are included in the main analysis.

[24] As in Section 3.6.1.1, I consider 2009 elections in Haryana and Jharkhand, not their 2005 elections.

[25] For school report cards 2010-1 onwards, I observe the start month of the school. I can therefore match elections to school years more exactly than I could to test scores. For a one year lag, I match elections to school years that started one to twelve months after the election. For a two year lag, I match elections to school years that started thirteen to twenty four months after the elections and so on. For school years before 2010-1, I take the modal start month of schools in that state in 2010-1 as the school start month. School start months are state specific (typically over 90% of schools in a given state start in the same month) and do not change much over time.

[26] Different schools teach different categories of grades. For example, some schools teach only lower primary (Grades 1 to 5) students, some both lower and upper primary (Grades 1 to 8), some upper primary only (Grades 6 to 8), whilst some have secondary students too. I create a dummy for each school type, taking lower primary only as the base category.

[27] Karnataka state had its 2008 elections in the same month as the start of the school year. The

unweighted, as the DISE dataset is intended to be a complete census.

Whilst the DISE report cards measure many variables, most of those variables are exactly the sort of school inputs that the existing literature on learning has demonstrated have little effect, since these are unfortunately also the most tangible and so easy to measure (Pritchett, 2013). It is much harder to measure subtle pedagogical changes in the classroom or improvements in teacher incentives that are more important for learning outcomes. For that reason, I consider only one of these standard school inputs, one that has been found to have small effects in the existing literature: the pupil to teacher ratio (Muralidharan & Sundararaman, 2013).

I then consider three variables with which I intend to capture the extent of monitoring and accountability of teachers and schools, which better reflects where learning effects have been found in the existing literature. These are: the number of academic inspections, the number of visits by personnel from the Block Resource Centre (BRC), and the number of visits by personnel from the Cluster Resource Centre (CRC). The BRC and CRC are local government institutions created under Sarva Shiksha Abhiyan (the flagship government elementary education scheme), and have several functions. They have a mandate to visit schools to address pedagogical issues, monitor the progress of school quality, and meet with community members to understand how best to improve schools (EdCIL Limited, 2010). Whilst I use these BRC and CRC visits as a proxy for monitoring and accountability, they could also partly capture pedagogical support offered to teachers and schools. The BRC and CRC organise teacher training and other support for teachers. Pritchett (2013) notes that accountability is really a system property, and so difficult to compartmentalise into individual variables like, say, the number of visits by the BRC. If the BRC does not deliver information on school performance to parents or the wider community effectively, or if there is no punishment for poor performance, a large number

first school year after the elections then starts after twelve months have passed, and so there are no observations for Karnataka for any outcomes at a one year lag. Tamil Nadu is missing several districts for the school report cards in 2008-9, which affects the three year lag only for all outcomes. Sixty fewer districts are included in the five year lag as I only observe school report cards up to 2013-4. Finally, for the pupil-to-teacher ratio at a two year lag I do not observe information from Madhya Pradesh, as the number of teachers was not recorded in the 2010-1 school year.

of visits by itself does not have much bite as an accountability tool. In spite of these weaknesses, these are the best available measures of the type of school input most likely to matter for learning outcomes.

I find no statistically significant effect for any outcome variable for any lag, including in the next election year. However, the estimates are not sufficiently precise that I can rule out small to moderate sized effects in either direction for most lags. For the regressions using 2005-09 elections presented in Figure A.1, the 95% confidence interval for the effect of an increase in F by 0.2 (an additional woman in the State Legislative Assembly in the median district) is between a 0.2 standard deviation decrease and 0.2 standard deviation increase in each of the four outcome variables for most lags. The nulls are more precise for the 2008-12 elections, for which I can only consider a one year lag (Table A.8). For those elections the 95% confidence interval rules out effects outside the interval $(-0.1, 0.09)$ standard deviations for academic inspections, $(-0.05, 0.11)$ standard deviations for BRC visits, $(-0.09, 0.14)$ standard deviations for CRC visits and $(-0.12, 0.06)$ standard deviations for the pupil-to-teacher ratio. There is little evidence here that female MLAs improve pupil-to-teacher ratios or the monitoring and accountability of teachers and schools as measured by inspections and BRC/CRC visits.

3.6.2.2 Remedial Education

Remedial education programs have been implemented in many states since the late 1990s, specifically targeting children who have fallen behind their classmates. One example is the Balsakhi Program, introduced by Pratham and evaluated in Banerjee et al. (2007), where young women from the local community help the poorest performing children in government schools. Perhaps female leaders encourage the use of these remedial education programs in their districts. To test this, I ask whether there are effects of female leaders at the higher end of the learning distribution too, which we would not expect to see if the effects were driven by remedial education programs.[28]

[28]It is possible that remedial education programs could improve learning outcomes for more able children too, since removing less able children from the classroom reduces class size and allows teachers

I find that the impact of female MLAs does also occur at the higher end of the learning distribution. In the main results, I consider the ability to do subtraction and read a Grade 1 level text. These are the second highest learning levels in the test. In Table A.9, I consider the highest learning outcomes tested: the ability to do division, and the ability to read a Grade 2 level text. An additional woman in the State Legislative Assembly in the median district increases the probability that a child can do division by 2.8 percentage points, and can read at a Grade 2 level by 2.2 percentage points. These effects are statistically significant at the 5% and 10% level respectively. A large proportion of children in the sample cannot do division or read at a Grade 2 level (68% and 53% respectively in the full sample, and even 29% of secondary students cannot do division). The effect of female leaders is therefore not limited to the minority of children who are 'falling behind'.

The size of the effects on division and Grade 2 reading is comparable to the main results when considered in context of the smaller fraction of the sample who have attained that level of learning. For division and Grade 2 reading the effect of an additional female MLA is equivalent to an 8.6 and 4.8 percent increase from the mean respectively. The effects occur for both girls and boys, and there are no statistically significant gender differences in the size of the effects. The effects are also concentrated amongst primary school students as demonstrated in Table A.10. The pattern of timing of these effects closely reflects that for subtraction and Grade 1 reading. The only lag (from zero to five) for which there is a statistically significant effect is at a lag of five, for which the effect is significant at the 10% level for both division and Grade 2 reading. The size of the estimate of β jumps at that lag by a similar multiple as for subtraction and Grade 1 reading. The estimate of β increases to 0.38 for division and 0.36 for Grade 2 reading, an increase by a factor of 2.7 and 3.2.

to teach at a higher level for the remaining, more able children. However, Banerjee et al. (2007) finds that remedial education programs in India directly help the less-able children who were in the program, and have no effects for the more able who were left with smaller classes.

3.6.3 The Effect is Concentrated in States with the Largest Initial Gender Gaps in Learning

There exists considerable variation in both the level of learning and the degree of gender inequality in learning across states. I therefore allow for heterogeneity across states. I look separately at the states that were initially doing best and worst in terms of gender inequality in learning. I divide the best fifteen and worst fourteen states in terms of the initial gender gap in 2007 (the year before the sample period for my main results begins),[29] separately for the fraction who can subtract and read at a Grade 1 level. The two darkest shades in Figure 3.5 highlight the fourteen worst performing states in each case.

The effect of female MLAs on learning for both boys and girls for both maths and reading is driven by the states where gender inequality was initially worst. The catch up of learning in maths is driven by these states too, such that the improvements in girls' learning relative to boys is exactly where it is needed most (Table A.11). These states are predominantly in Northern India, and include the 'BIMARU' states of Bihar, Madhya Pradesh, Rajasthan and Uttar Pradesh for both maths and reading. An additional woman in the State Legislative Assembly in these states, in a district with five constituency seats, increases the probability that a child can do subtraction by 4.8 percentage points and can read at a Grade 1 level by 4.6 percentage points, a slightly larger effect than the main results. These increases are respectively 9.4 and 9 percent of the mean. The magnitude of the gender difference in the effect in these states is comparable to the main results. There is no statistically significant effect in the fifteen states who do best in terms of gender inequality in learning, and the estimates of β are considerably smaller too.

[29]Mizoram is not included in the ASER dataset for 2007, and so I take its learning gender gap in 2008.

3.7 Placebo Tests

Female MLAs should not affect test scores before they are elected. In Table 3.7, I present placebo tests of this claim. I use the same approach as described in Section 3.6.1.1, but consider test scores before women are elected as the outcome variables. In columns (1) to (4), I use test scores from the year before women are elected.[30] For those regressions I use elections from 2008-12, which are the only elections for which I can observe test scores the year before (I have observations from 525 districts). I find that this placebo test fails for both maths and reading. The effects on test scores one year before the election are statistically significant at the 1% level for both subjects. However, these effects are much smaller than the effects in the subsequent election year (the five year lag), and so can only explain a small fraction of those effects, as demonstrated in Figure 3.4. For maths and reading, the point estimates of the effects one year before the election are, respectively, less than one quarter and less than one fifth of the size of the effects after five years.

One explanation for the failed placebo test is that women might be elected in close elections against men in districts where women had won in close elections against men in previous elections. If that is true, and female leaders affect learning outcomes around the time of the next election, then we would expect a failure of the placebo test.[31] Consistent with this explanation, I find no evidence that test scores were higher two years before the election in districts where women win in close elections in columns (5) and (6) of Table 3.7. I perform a placebo test using test scores two years before women are elected as the outcome variable, which I can do for those states that had elections 2009-13, which is all but five states. The estimates of β fall to 0.05 and 0.04 for maths and reading

[30] In columns (1) and (2), I use the baseline regression specification, whilst in columns (3) and (4), I include the vote margin running variable control as explained in Section 3.5.2. The results are very similar.

[31] The main effect occurs at a five year lag, which would imply this placebo test failure should occur at a zero lag rather than the year before. However, previous papers have shown election cycles in test scores at both the year before and the year of elections (Fagernas & Pelkonen, 2014). Further, since I do not observe the exact month in which test scores are taken, and election months vary across states and election years, it is possible that effects both one year before an election and the year of an election capture an election cycle effect from the previous incumbent.

respectively, and are not statistically significant (p-values 0.58 and 0.56).

If this is the explanation for the failed placebo, then I have omitted a variable from the right hand side of my regression equations: test scores in the year before elections took place. If female leaders raise test scores, test scores are positively correlated over time, and women win in close elections against men in the same districts across successive elections, my estimate of β is biased upwards, and can be thought of as an upper bound for the true effect. To investigate this further, I first ask whether it is indeed the case that in districts in which a woman wins in a close election against a man, a woman wins in a close election against a man again in the next round of elections. As explained in Section 3.6.1.2, I can only do this for the subset of district-election years where I observe a subsequent election. I only observe one election in some states, and even in the states for which I observe two, I can only do this for the first round of elections.[32] Of the thirty-nine districts in which at least one woman was elected in a close election against a man in the first round of elections, eight had at least one woman win against a man in a close election in the next round of elections too.[33]

Secondly, I try to directly control for test scores the year before elections took place. I set up the data in a similar way to the baseline regressions. I match test scores from 2009-14 to the most recent election that occurred in the district between one and five years before the test was taken, using elections from 2008-12. I then additionally control for the district average test score of someone of the same age and gender as the child in the year before the election. So for an eight-year-old girl from Vadodara district (Gujarat) whose test score I observe in 2014, I control in both the first and second stage for the average test score across all eight-year-old girls in Vadodara in 2011, the year before the 2012 Gujarat elections. I ask whether the fraction of district seats held by women in the most recent election has explanatory power over current test scores over and above the pre-election average test scores in the district. As demonstrated in the final two columns

[32] As explained in Section 3.6.1.2 I use elections data from 2014 and February 2015, which is the most recent published by the ECI, but I cannot use elections in December 2013 for which district identifiers are not published.

[33] I use a 3.5% close elections definition here.

of Table 3.7, I find no such effect. Pre-election test scores have significant explanatory power, at the 1% significance level, for both maths and reading. However, the temporal structure of the datasets again prevents me from testing this as I would like. I can only observe both a five year and minus one year lag for sixteen states in the sample. I cannot match the thirteen states with elections in 2010, 2011, and 2012 both to subsequent test scores five years later and one year before. For those thirteen states, test scores at a five year lag were therefore not included in the regressions in columns (7) and (8) (although the shorter lags for these states were included), and yet I previously found most of the effect exactly at the five year lag.

To summarise, the evidence presented in this Section gives reason to interpret the main results cautiously. It is possible that the placebo test failure arises because women win in close elections against men in districts where women previously won in close elections against men. However, I do not observe enough successive elections to test this claim as I would like.

3.8 Conclusion

I find evidence that women in the State Legislative Assembly improve learning outcomes for children in their district in rural India in both maths and reading, and help to reduce the gender gap in maths, during the period from 2008 to 2014. The most plausible mechanism for the effect is an increase in effort by political leaders seeking re-election, where additional effort from female leaders is targeted towards education. The effect is concentrated around the next election year, and dies out immediately afterwards. It is not clear that these effects can be interpreted as a deeper improvement in child learning, since they appear to be transitory. The effect occurs for primary school children and in the Indian states that initially performed worst in terms of gender inequality, which are largely concentrated in Northern India. I stress that the results must be taken cautiously, however, as placebo tests suggest that women won in close elections against men in districts where learning outcomes were already higher the year before the election. It is possible that women win in close elections in districts where women previously won in close elections, which could drive the placebo test failure. However, this is something I cannot test with the data I have, as I do not observe enough successive elections.

3.9 Tables and Figures

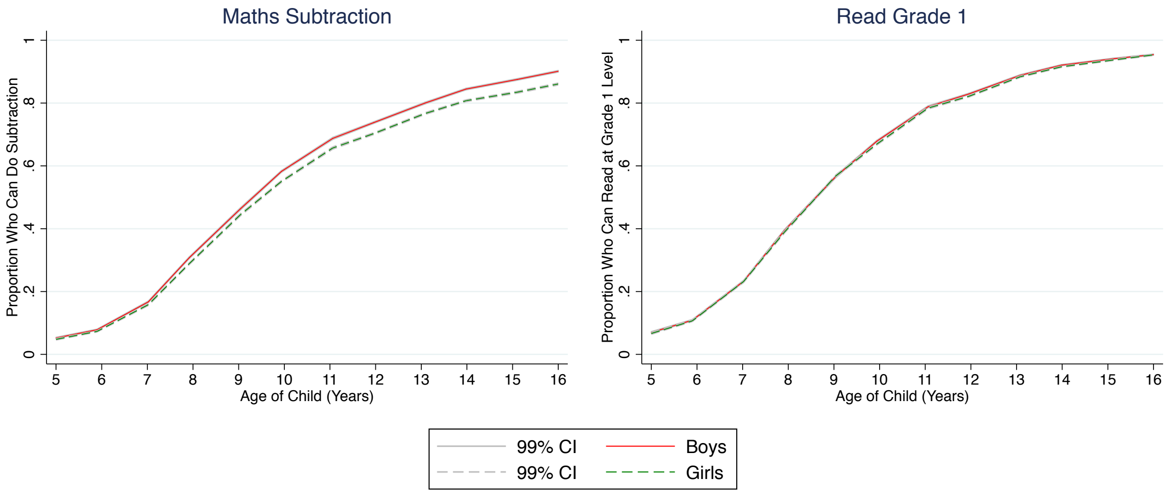


Figure 3.1: *Notes.* **National Gender Gaps in Learning.** I run local polynomial regressions of each of the two outcome variables: in Panel A whether or not the child can do subtraction, and Panel B whether or not the child can read at a Grade 1 level, on age. I plot 99% confidence intervals. I have 3,147,284 and 3,158,912 in-school children for subtraction and Grade 1 reading respectively. Due to the large number of observations, the confidence intervals are very tight. I use a bandwidth of 0.05 (which is larger than the mean integrated squared error minimising bandwidth, but produces a smoother fit to the data).

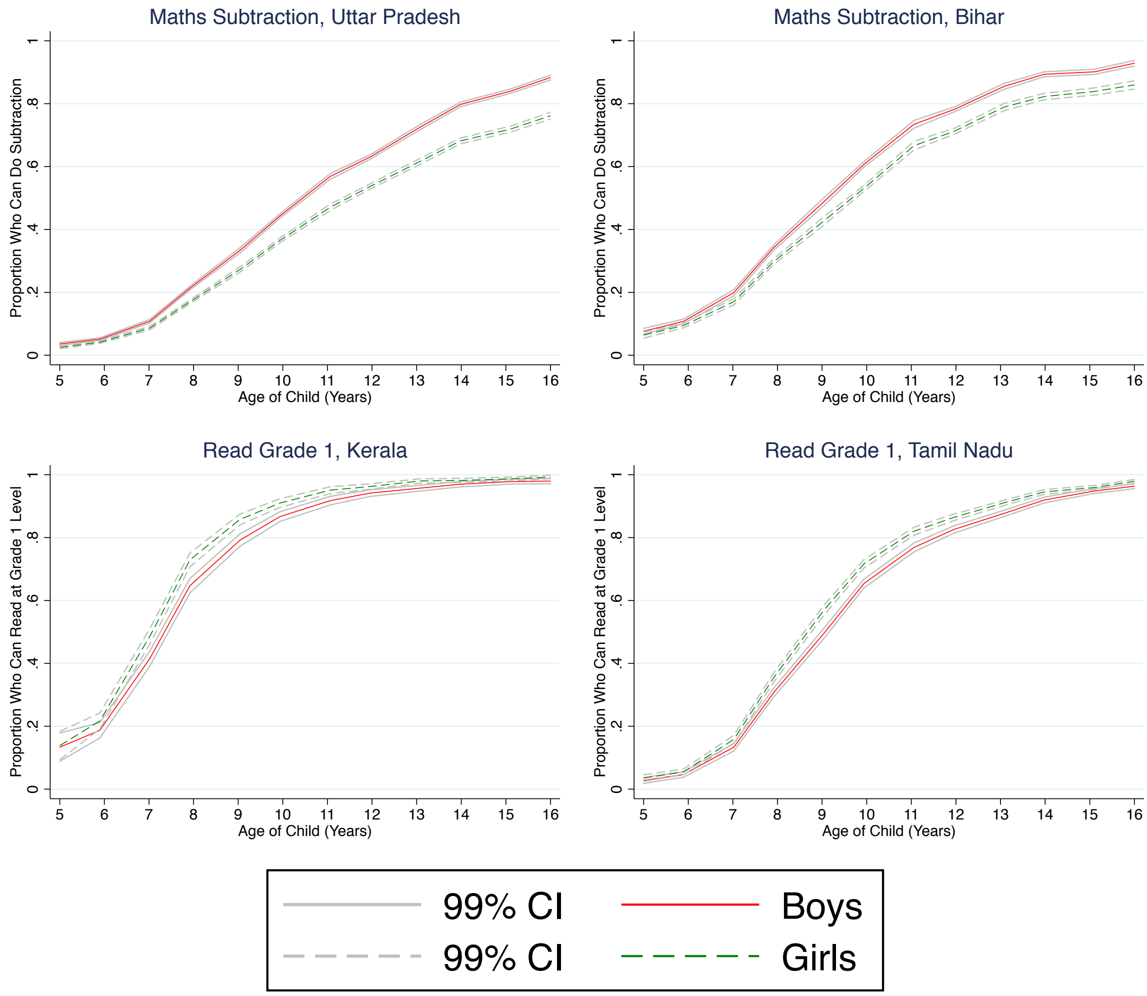


Figure 3.2: *Notes.* **State Specific Gender Gaps in Learning.** I run local polynomial regressions of whether or not the child can do subtraction (Panels A and B), or whether or not the child can read at a Grade 1 level (Panels C and D), on age. I plot 99% confidence intervals. I have 482,033, 271,097, 60,825, and 139,265 in school-children who sat the tests for Uttar Pradesh, Bihar, Kerala and Tamil Nadu respectively. I use a bandwidth of 0.05 (which is larger than the mean integrated squared error minimising bandwidth, but produces a smoother fit to the data).

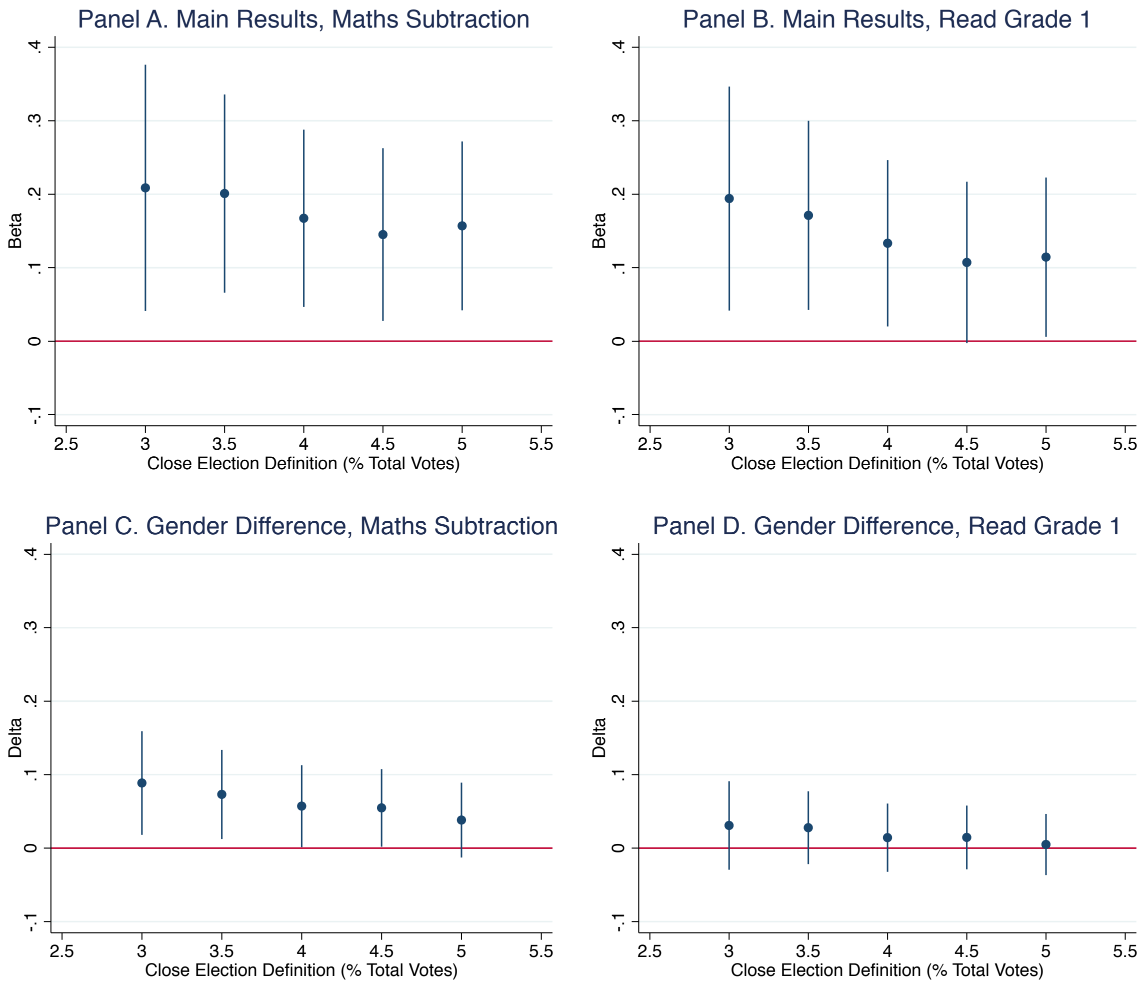


Figure 3.3: *Notes.* **Robustness across close election definitions.** Panels A and B plot estimates of β, the effect for the full sample, across close election definitions. Panels C and D plot estimates of δ, the gender difference in the effect, across close election definitions. In all cases I plot 95% confidence intervals.

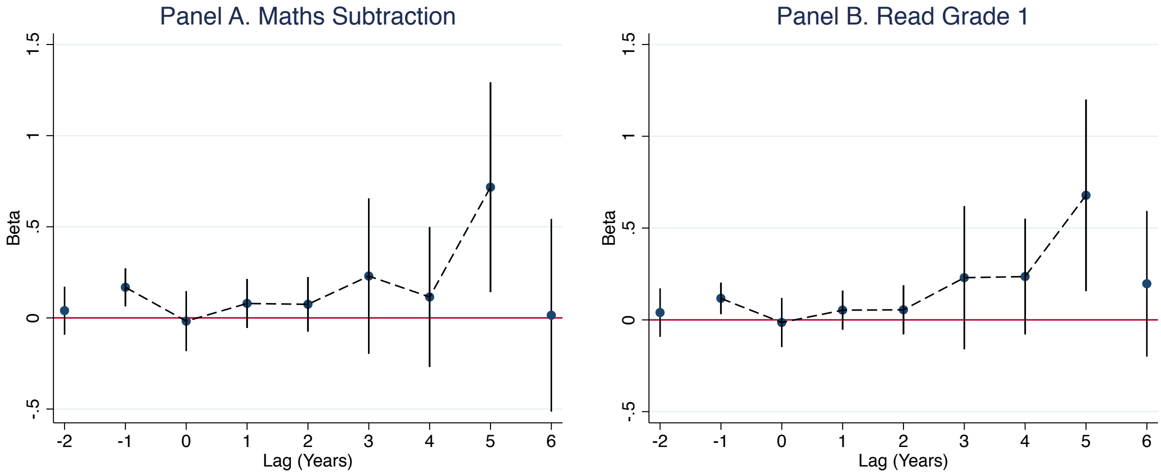
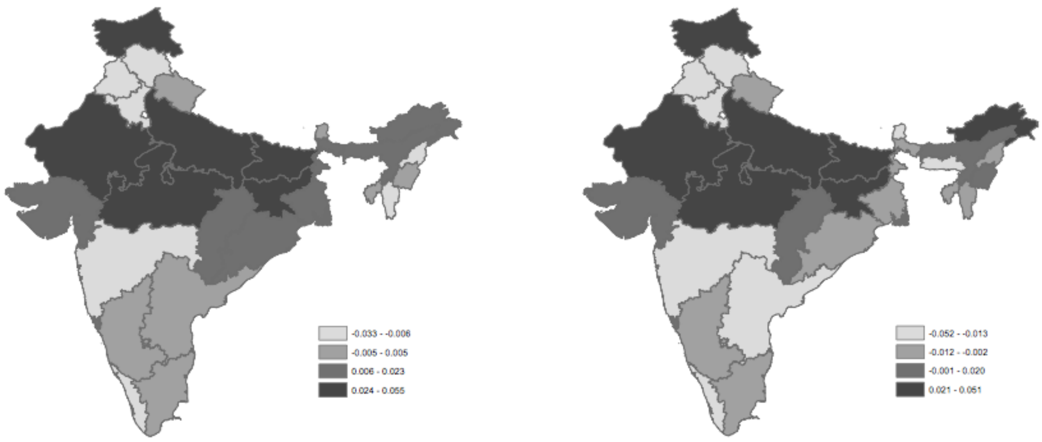


Figure 3.4: *Notes.* **Timing.** I plot 95% confidence intervals for the estimate of β, using a 3.5% close election definition. The estimates for the lags from minus one to five years are comparable in their regional coverage, as they are obtained using data from all (29) states other than Jammu & Kashmir. To indicate that these estimates are comparable, I join them with the dashed line. The estimates for the lags minus one to two years use elections 2008-12, and the estimates for the three to five year lags use elections 2005-9. The instrument has less explanatory power in the first stage for the 2005-09 elections, which is why confidence intervals are larger for the lags from three to five years. For the lag minus two, I use elections from 2009 to November 2013, and can only include 24 states (these estimates do not include information from Chattisgarh, Jammu & Kashmir, Madhya Pradesh, Mizoram, or Rajasthan). For the six year lag, I use elections from 2005-08, and can only include 24 states (these estimates do not include information from Andhra Pradesh, Arunachal Pradesh, Maharashtra, Orissa or Sikkim).



(a) Maths Subtraction (b) Read Grade 1

Figure 3.5: *Notes.* **Regional Differences in Gender Inequality in Learning.** I plot the level of gender inequality in a learning level (subtraction or Grade 1 reading), calculated as the fraction of boys who can perform that learning level minus the fraction of girls who can perform that learning level (using test scores from 2007, the year before the sample period for my main results begins). Higher numbers indicate greater inequality against girls. I separate states into four categories of eight, seven, seven and seven, from lowest to highest gender inequality against girls. The two darkest shades represent the fourteen worst states in terms of gender inequality against girls in that learning level, which is how I split the sample in the regional regression analysis. That regression analysis highlights that these are also the states where the effects occur.

Table 3.1: Balance Test: Census

	More Men Win Against Women than Vice-Versa in Close Elections	More Women Win Against Men than Vice-Versa in Close Elections	p-value (H0: Diff = 0)
Total Population (Millions)	2.584	2.777	0.492
	(0.203)	(0.192)	
Rural Population (Proportion)	0.719	0.706	0.686
	(0.022)	(0.021)	
Female Population (Proportion)	0.484	0.486	0.351
	(0.001)	(0.001)	
SC/ST Population (Proportion)	0.269	0.271	0.941
	(0.019)	(0.018)	
0-6 Years Population (Proportion)	0.140	0.135	0.306
	(0.003)	(0.003)	
Female Fraction of 0-6 Population	0.478	0.479	0.435
	(0.001)	(0.001)	
Female Literacy Rate	0.558	0.559	0.941
	(0.013)	(0.013)	
Male literacy rate	0.690	0.696	0.671
	(0.010)	(0.009)	
Working Population (Proportion)	0.386	0.399	0.162
	(0.007)	(0.008)	
Observations	88	98	186

Standard errors of the estimates of the means are in parentheses. I present p-values for the t-test for the difference in means. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. I use a 5% close election definition. I use data from the 2011 Census.

Table 3.2: Balance Test: Elections

	More Men Win Against Women than Vice-Versa in Close Elections	More Women Win Against Men than Vice-Versa in Close Elections	p-value (H0: Diff = 0)
Total Votes (100,000s)	1.209 (0.034)	1.279 (0.033)	0.144
Vote Difference (Proportion of Total)	0.024 (0.001)	0.025 (0.001)	0.484
Fraction of Electorate Who Voted	0.696 (0.011)	0.697 (0.011)	0.944
Female Fraction of Candidates	0.179 (0.008)	0.180 (0.008)	0.926
Age of Winner	50.403 (0.852)	45.171 (0.841)	0.000***
Winner is SC/ST	0.338 (0.039)	0.392 (0.038)	0.317
Winner in BJP	0.182 (0.034)	0.291 (0.034)	0.023**
Winner in INC	0.221 (0.035)	0.272 (0.034)	0.294
Observations	154	158	312

Standard errors of the estimates of the means are in parentheses. I present p-values for the t-test for the difference in means. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. I use a 5% close election definition.

Table 3.3: Women in State Government

	Constituency Seats	% Total Seats
Woman won close election against man (3%)	72	1.25
Woman won close election against man (3.5%)	84	1.45
Woman won close election against man (4%)	100	1.73
Woman won close election against man (4.5%)	113	1.96
Woman won close election against man (5%)	123	2.13
Woman won	445	7.70
Total	5777	100

Table 3.4: Effect of Female Leaders on Learning

	Maths Subtraction			Read Grade 1		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	IV First Stage	OLS	IV	IV First Stage
F	0.015 (0.020)	**0.20***** **(0.069)**		0.0018 (0.017)	**0.17***** **(0.066)**	
FC (Instrument)			**0.86***** **(0.11)**			**0.86***** **(0.11)**
Child And Election Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,535,288	2,535,288	2,535,288	2,544,897	2,544,897	2,544,897
KP F-statistic		64.2	64.2		64.2	64.2
Mean of outcome	0.53	0.53		0.62	0.62	

Notes. Standard errors (in parentheses) clustered at district level. This Table presents the baseline regression results. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: Effect of Female Leaders on the Gender Gap in Learning

	Maths Subtraction		Read Grade 1	
	(1)	(2)	(3)	(4)
	OLS	IV	OLS	IV
F	0.019 (0.021)	**0.17**** **(0.071)**	0.0048 (0.017)	**0.16**** **(0.066)**
F x Female	-0.0066 (0.014)	**0.073**** **(0.031)**	-0.0065 (0.012)	0.028 (0.025)
Child And Election Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,535,288	2,535,288	2,544,897	2,544,897
SW F-statistic for F		171.2		171.3
SW F-statistic for F x Female		173.4		173.5
Mean of outcome	0.53	0.53	0.62	0.62

Notes. Standard errors (in parentheses) clustered at district level. This Table presents regression results allowing for a separate effect of female political leaders by child gender. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Robustness: Inclusion of Controls for Vote Margin

	Maths Subtraction						Read Grade 1					
	Linear		Second Order		Third Order		Linear		Second Order		Third Order	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	IV	IV Gender Difference	IV	IV Gender Difference	IV	IV Gender Difference	IV	IV Gender Difference	IV	IV Gender Difference	IV	IV Gender Difference
F	**0.23*****	**0.19****	**0.23*****	**0.19****	**0.23*****	**0.19****	**0.19****	**0.18****	**0.19****	**0.18****	**0.20****	**0.19****
	(0.082)	**(0.083)**	**(0.082)**	**(0.083)**	**(0.085)**	**(0.086)**	**(0.078)**	**(0.079)**	**(0.078)**	**(0.079)**	**(0.080)**	**(0.081)**
F x Female		**0.089****		**0.089****		**0.089****		0.022		0.022		0.022
		(0.042)		**(0.042)**		**(0.042)**		(0.033)		(0.033)		(0.033)
Min. Vote Margin	-0.13***	-0.13***	-0.13**	-0.13**	-0.13**	-0.13**	-0.12***	-0.12***	-0.12***	-0.12***	-0.14**	-0.14**
	(0.049)	(0.049)	(0.049)	(0.049)	(0.066)	(0.066)	(0.042)	(0.042)	(0.042)	(0.042)	(0.057)	(0.057)
Min. Vote Margin Sq			0.061	0.061	0.071	0.071			0.021	0.021	0.041	0.041
			(0.15)	(0.15)	(0.15)	(0.15)			(0.11)	(0.11)	(0.12)	(0.12)
Min. Vote Margin Cub					0.11	0.10					0.20	0.20
					(0.46)	(0.47)					(0.35)	(0.35)
Child And Election Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,397,163	1,397,163	1,397,163	1,397,163	1,397,163	1,397,163	1,402,217	1,402,217	1,402,217	1,402,217	1,402,217	1,402,217
KP F-statistic	52.5		53.1		50.2		52.5		53.1		50.1	
SW F-statistic for F		94.0		93.9		93.5		94.0		93.9		93.5
SW F-statistic for F x Female		93.6		93.6		93.7		93.6		93.7		93.8
Mean of outcome	0.53	0.53	0.53	0.53	0.53	0.53	0.63	0.63	0.63	0.63	0.63	0.63

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I demonstrate the robustness of the results to the inclusion of vote margin running variables. In columns (1), (2), (7) and (8), I control linearly for the smallest vote margin between a female and male candidate in the district. In columns (3), (4), (9) and (10), I control for a second order polynomial in that vote margin, and in columns (5), (6), (11) and (12), a third order polynomial. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7: Placebo Tests

	Minus One Year Lag				Minus Two Year Lag		Control for Lag Test Score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Maths Subtraction IV	Read Grade 1 IV	Maths Subtraction IV	Read Grade 1 IV	Maths Subtraction IV	Read Grade 1 IV	Maths Subtraction IV	Read Grade 1 IV
F	**0.17*****	**0.12*****	**0.19*****	**0.13*****	0.047	0.040	0.040	0.017
	(0.053)	**(0.044)**	**(0.060)**	**(0.048)**	(0.085)	(0.067)	(0.049)	(0.041)
Pre-election Maths							0.47***	
							(0.018)	
Pre-election Reading								0.60***
								(0.015)
Child And Election Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State and Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Vote Margin Running Variables	No	No	Yes	Yes	No	No	No	No
Observations	474,011	475,751	281,578	282,765	372,232	374,087	1,410,756	1,415,535
KP F-statistic	117.1	117.3	92.3	92.2	127.0	127.3	99.9	99.8
Mean of outcome	0.58	0.64	0.58	0.64	0.58	0.64	0.52	0.63

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I present results from placebo tests. In columns (1), (2), (3) and (4), I look at the effect of female political leaders on test scores one year before the election. In columns (3) and (4), I add the vote margin running variable. In columns (5) and (6), I look at the effect of female political leaders on test scores two years before the elections. In columns (7) and (8), I run the baseline regression, but additionally control for test scores one year before the election. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix

3.A Additional Tables and Figures

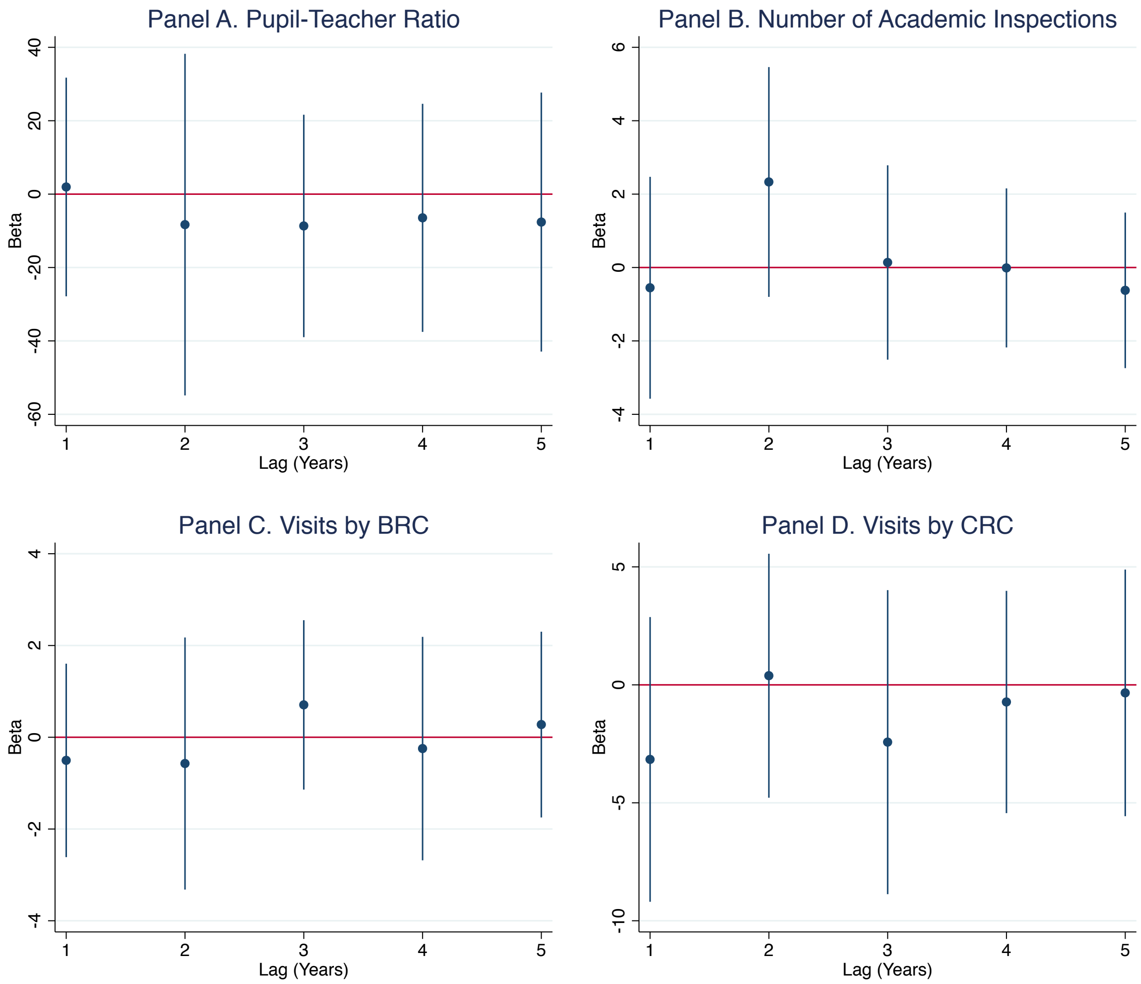


Figure A.1: *Notes.* **School Inputs.** I plot 95% confidence intervals for the estimate of β using school inputs as the outcome variables. I use a 3.5% close election definition. For all four panels and all lags I have used elections from 2005-09.

Table A.1: State Legislative Assembly Election Dates

Election Year	States
2005	Bihar, Haryana, Jharkhand
2006	Assam, Kerala, Tamil Nadu, Pondicherry, West Bengal
2007	Goa, Gujarat, Himachal Pradesh, Manipur, Punjab, Uttaranchal, Uttar Pradesh
2008	Chattisgarh, Jammu and Kashmir, Karnataka, Madhya Pradesh, Meghalaya, Mizoram, Nagaland, Rajasthan, Tripura
2009	Andhra Pradesh, Arunachal Pradesh, Haryana, Jharkhand, Maharashtra, Orissa, Sikkim
2010	Bihar
2011	Assam, Kerala, Tamil Nadu, Pondicherry, West Bengal
2012	Goa, Gujarat, Himachal Pradesh, Manipur, Punjab, Uttaranchal, Uttar Pradesh
2013	Karnataka, Meghalaya, Nagaland, Tripura

Table A.2: Separate Sub-Samples Girls and Boys

	Girls		Boys	
	(1)	(2)	(3)	(4)
	Maths Subtraction IV	Read Grade 1 IV	Maths Subtraction IV	Read Grade 1 IV
F	**0.22*****	**0.19*****	**0.18*****	**0.16****
	(0.071)	**(0.068)**	**(0.069)**	**(0.065)**
Child And Election Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,185,903	1,190,403	1,349,385	1,354,494
KP F-statistic	67.9	68.0	60.6	60.6
Mean of outcome	0.52	0.62	0.55	0.62

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I present the baseline regression results splitting the sample into girls and boys. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Primary versus Secondary Students

	Primary				Secondary			
	Maths Subtraction		Read Grade 1		Maths Subtraction		Read Grade 1	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	IV	IV Gender Difference	IV	IV Gender Difference	IV	IV Gender Difference	IV	IV Gender Difference
F	**0.21*****	**0.18****	**0.19*****	**0.16****	0.078	0.041	0.030	0.023
	(0.074)	**(0.076)**	**(0.072)**	**(0.066)**	(0.055)	(0.061)	(0.023)	(0.026)
F x Female		**0.072****		0.028		0.076		0.015
		(0.032)		(0.025)		(0.047)		(0.017)
Child And Election Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,150,867	2,150,867	2,159,742	2,544,897	384,421	384,421	385,155	385,155
KP F-statistic	66.0		66.0		49.6		49.7	
SW F-statistic for F		172.1		171.3		152.6		152.7
SW F-statistic for F x Female		175.1		173.5		150.3		150.5
Mean of outcome	0.47	0.47	0.56	0.62	0.88	0.88	0.96	0.96

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I split the sample into primary and secondary age students and run the baseline regressions and allow for a difference in the effect by gender. In columns (1) to (4), I look at primary aged children, and in columns (5) to (8), I look at secondary aged children. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.4: Robustness: Inclusion of Additional Control Variables

	Household Controls		State-Year Fixed Effects	
	(1)	(2)	(3)	(4)
	Maths Subtraction IV	Read Grade 1 IV	Maths Subtraction IV	Read Grade 1 IV
F	**0.18***** **(0.059)**	**0.15***** **(0.055)**	**0.17***** **(0.065)**	**0.17***** **(0.060)**
Child And Election Controls	Yes	Yes	Yes	Yes
Household Controls	Yes	Yes	No	No
Year Fixed Effects	Yes	Yes	No	No
State Fixed Effects	Yes	Yes	No	No
State-Year Fixed Effects	No	No	Yes	Yes
Observations	2,387,717	2,396,273	2,535,288	2,544,897
KP F-statistic	64.6	64.6	62.5	62.5
Mean of outcome	0.53	0.62	0.53	0.62

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I present two additional tests of robustness. In columns (1) and (2), I include additional controls for whether the household has an electricity connection, a TV, a house made of pucca building material, and whether the child's mother had ever gone to school. In columns (3) and (4), I control for state-year fixed effects (I therefore implicitly also control for state and year fixed effects). Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Additional Timing Results

	Five Year Lag	Zero to Four Year Lags	
	(1)	(2)	(3)
	Maths Subtraction IV Gender Difference	Maths Subtraction IV	Read Grade 1 IV
F	**0.66****	**0.12****	**0.094***
	(0.29)	**(0.057)**	**(0.054)**
F x Female	**0.12****		
	(0.057)		
Child And Election Controls	Yes	Yes	Yes
State and Year Fixed Effects	Yes	Yes	Yes
Observations	367,670	2,693,407	2,703,442
KP F-statistic		86.8	87.0
SW F-statistic for F	132.1		
SW F-statistic for F x Female	123.8		
Mean of outcome	0.49	0.54	0.63

Notes. Standard errors (in parentheses) clustered at district level. In column (1), I look at the effect of female political leaders on the gender gap in learning five years after the election. In columns (2) and (3), I drop the five year lag and re-run the baseline regression. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men. Column (1) uses a five year lag.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Re-election

	Maths Subtraction			Read Grade 1		
	(1) IV	(2) IV	(3) IV	(4) IV	(5) IV	(6) IV
F	0.27	0.29	0.18	**0.32***	0.44	0.29
	(0.20)	(0.27)	(0.20)	**(0.18)**	(0.32)	(0.19)
F x Stood		0.29			0.21	
		(0.52)			(0.46)	
F x Won			0.58			0.47
			(0.55)			(0.50)
Child And Election Controls	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,651,474	1,651,474	1,651,474	1,658,621	1,658,621	1,658,621
KP F-statistic	12.9			13.0		
SW F-statistic for F		14.4	16.9		14.5	17.0
SW F-statistic for F x Stood		4.0			4.0	
SW F-statistic for F x Won			3.8			3.8
Mean of outcome	0.55	0.55	0.55	0.63	0.63	0.63

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I test whether the effect of female political leaders is greater in the districts where those leaders stand for or win re-election. In columns (2) and (5), I interact the main right hand side variable, F, with a dummy variable Stood, which takes value 1 if any female leader in the district stood for re-election in the next round of elections in the same district. As a second instrument, I interact Stood with the original instrument, FC. In columns (3) and (6), I interact the main right hand side variable, F, with a dummy variable Won, which takes value 1 if any female leader in the district won re-election in the next round of elections in the same district. As a second instrument, I interact Won with the original instrument, FC. I can only include observations from the years between the first and second round of elections in districts in which I observe a second round of elections, and so in columns (1) and (4), I run the baseline regression on this restricted sub-sample to see if the main effects still exist in that sub-sample. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Effect of Female Leaders in the Ruling Party

	(1)	(2)
	Maths Subtraction IV	Read Grade 1 IV
F	**0.21***** **(0.081)**	**0.20**** **(0.089)**
F x Ruling	0.22 (0.23)	0.15 (0.22)
Child And Election Controls	Yes	Yes
Year Fixed Effects	Yes	Yes
State Fixed Effects	Yes	Yes
Observations	2,511,696	2,521,281
SW F-statistic for F	164.4	163.9
SW F-statistic for F x Ruling	34.5	34.5
Mean of outcome	0.53	0.62

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I allow for a difference in the effect of female political leaders in the districts where at least one female leader is part of the ruling party in her state. I define a variable Ruling, which takes the value 1 if any female leader in the district is part of any of the ruling parties (there may be more than one ruling party in the event of a coalition), and interact this with the main right hand side variable, F. The second instrument is an interaction of Ruling and the original instrument, FC. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men. I control for the variable Ruling in both regressions.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: School Inputs 2008-12 Elections, Lag One Year

	(1)	(2)	(3)	(4)
	Pupil-to-Teacher ratio IV	Number of Academic Inspections IV	Visits by BRC IV	Visits by CRC IV
F	-5.49 (8.12)	-0.11 (0.77)	0.46 (0.59)	0.66 (1.78)
School And Election Controls	Yes	Yes	Yes	Yes
State And Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,001,651	1,015,147	1,015,231	1,011,551
KP F-statistic	80.3	77.9	77.7	77.5
Standard deviation of outcome	34.6	3.15	2.83	6.04

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I look at the effect of female political leaders on school inputs one year after the election, using elections from 2008 to 2012. Observations in all regressions are unweighted. I use a 3.5% close election definition. In the school controls I include dummy variables for whether the school is all-boys or all-girls (base category co-ed), for the category of grades taught at the school, and for the school management type. The election controls include the fraction of district seats won by the BJP party as well as the fraction of seats decided in a close election between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.9: Effect of Female Leaders on Higher Learning Levels

	Maths Division				Read Grade 2			
	Baseline		Gender Difference		Baseline		Gender Difference	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
F	0.025	**0.14****	0.031	**0.12***	0.0062	**0.11***	0.015	**0.11***
	(0.020)	**(0.062)**	(0.021)	**(0.067)**	(0.017)	**(0.061)**	(0.017)	**(0.061)**
F x Female			-0.013	0.046			-0.019	0.0021
			(0.013)	(0.031)			(0.013)	(0.028)
Child And Election Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	2,535,288	2,535,288	2,535,288	2,535,288	2,544,897	2,544,897	2,544,897	2,544,897
KP F-statistic		64.2				64.2		
SW F-statistic for F				171.2				171.3
SW F-statistic for F x Female				173.4				173.5
Mean of outcome	0.32	0.32	0.32	0.32	0.47	0.47	0.47	0.47

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I look at the effect of female political leaders on higher learning levels. In columns (1) to (4), I look at the effect on the ability to do division, and in columns (5) to (8), read to a grade 2 level. In columns (1), (2), (5) and (6), I look at the effect on the level of learning, and in columns (3), (4), (7) and (8), I look at the effect on the gender gap in learning. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.10: Primary versus Secondary Students: Higher Learning Levels

	Primary		Secondary	
	(1)	(2)	(3)	(4)
	Maths Division IV	Read Grade 2 IV	Maths Division IV	Read Grade 2 IV
F	**0.14****	**0.12***	0.085	0.042
	(0.064)	**(0.065)**	(0.096)	(0.052)
Child And Election Controls	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,150,867	2,159,742	384,421	385,155
KP F-statistic	66.0	66.0	49.6	49.7
Mean of outcome	0.25	0.39	0.71	0.90

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I split the sample into primary and secondary age students and run the baseline regressions, looking at the higher learning levels: ability to do division and ability to read at a grade 2 level. In columns (1) and (2), I look at primary aged children, and in columns (3) and (4), I look at secondary aged children. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design. I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.11: Regional Heterogeneity

	Worst 14 States			Best 15 States	
	(1)	(2)	(3)	(4)	(5)
	Maths Subtraction IV	Maths Subtraction IV Gender Difference	Read Grade 1 IV	Maths Subtraction IV	Read Grade 1 IV
F	**0.24*****	**0.21****	**0.23*****	0.093	0.083
	(0.080)	**(0.082)**	**(0.078)**	(0.13)	(0.10)
F x Female		**0.077****			
		(0.034)			
Child And Election Controls	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
State Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	1,629,447	1,629,447	1,584,056	905,841	960,841
KP F-statistic	39.3		32.3	32.2	48.8
SW F-statistic for F		134.4			
SW F-statistic for F x Female		136.2			
Mean of outcome	0.49	0.49	0.58	0.61	0.69

Notes. Standard errors (in parentheses) clustered at district level. In this Table, I split the sample by initial gender inequality in learning. Observations in all regressions are weighted inversely proportional to the probability the household is sampled by ASER, to reflect the sampling design I use a 3.5% close election definition. In the child and election controls I include a child's age, gender, school grade, and type of school (private, madarsa or other, with government school as the base group), the fraction of district seats won by the BJP party as well as the fraction of seats decided in close elections between women and men. The worst 14 and best 15 states are calculated in terms of gender inequality in learning in 2007 across children of all ages. This is done separately for maths subtraction, and Grade 1 reading, and so the states in each category differ slightly across subjects.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.12: Data Appendix

Variable	Data Source	Method
Maths Subtraction, Read Grade 1	ASER	I create a dummy variable taking the value 1 if a child can do subtraction, or read at Grade 1 level, as explained in Section 3.3.2.1. The ASER dataset does not cover Delhi, or Jammu & Kashmir in 2010. Access to the raw ASER dataset can be applied for through the ASER centre. See http://www.asercentre.org for details.
Fraction of Seats Won by a Woman, F	Election Commission of India	For each district-election year, I calculate the fraction of constituency seats that are won by a woman, as explained in Section 3.3.2.2. I have elections data from January 2005 until November 2013. Bihar had elections in both February and October 2005. Throughout the analysis I consider only the October elections for Bihar in 2005. I do not observe elections in December 2013 to match to the 2014 ASER test scores, because the elections data for December 2013 does not include district identifiers. The elections data can be downloaded from: http://eci.nic.in/eci_main1/ElectionStatistics.aspx.
Fraction of Seats Won by a Woman Against a Man in Close Election, FC	Election Commission of India	For each district-election year, I calculate the fraction of constituency seats that are won by a woman against a man, in second place, in a close election. I categorise an election as close if the winning margin between first and second place is less than 3.5% of the total votes in the baseline regressions. The elections data can be downloaded from: http://eci.nic.in/eci_main1/ElectionStatistics.aspx.
Household Level Controls	ASER	I create a dummy variable for whether the child's household has an electricity connection, a TV, a house made of pucca building material, and whether the child's mother had ever gone to school. Details on accessing the ASER data can be found in the first row of this Table.
Ruling	Asher & Novosad (2015)	I create a dummy variable that takes the value 1 if at least one female MLA in the district is a member of any of the ruling parties in that state, as explained in Section 3.6.1.3. This dataset was provided to me by Sam Asher.
School Inputs	DISE	I use a measure of the number of academic inspections, visits by the BRC, visits by the CRC, and the pupil-to-teacher ratio, as described in Section 3.6.2.1. This dataset can be downloaded from: http://schoolreportcards.in/SRC-New/.

Conclusion

Discussion and Avenues for Future Research

This thesis has investigated how parental labour market experiences, gender, and the wider political environment in which a child grows up can each affect childhood outcomes.

In Chapter 1, we find that a one percentage point increase in the unemployment rate causes a 20 percent increase in child neglect in the United States, using a Bartik instrument. We use a new county-level dataset, which contains one observation for every reported incident of child maltreatment, covering the period from 2004 to 2012. One important mechanism for the main result is that parents lack the social and private safety nets required to provide for their children's basic needs after being made unemployed.

Child maltreatment is an area of research in which few studies have been conducted by economists. The scope for new research in this area is huge. In my opinion, two lines of enquiry could be particularly fruitful. First, more work is needed to establish a theoretical framework with which we can think about child maltreatment. A few papers have incorporated child maltreatment into models of human capital formation (e.g., Seiglie (2004); Akabayashi (2006)). Future researchers might also borrow non-cooperative bargaining models from the literature on intimate partner violence, for example Tauchen et al. (1991); Farmer & Tiefenthaler (1997); Bloch & Rao (2002); Eswaran & Malhotra (2011). In that literature, abuse is generally modelled as either instrumental (abuse is a means of altering behaviour or extracting rent) or expressive (the perpetrator gains utility from abuse), and the level of abuse is constrained by the victim's outside option in the event of divorce. To adapt these models to the case of child abuse and neglect, future researchers need to answer several questions. For example, is a child capable of bargaining on their own behalf? Or is the bargaining game between parents, who might place different weights on their child's outcomes in their own utility function? If a child can bargain for themselves, do they really have an outside option, and if so what is it?

Second, future research could ask how we can encourage people to report child maltreatment. For both child maltreatment and domestic violence more generally, neighbours, family and friends can be reluctant to report their suspicions, often out of fear that those suspicions are wrong. For this reason, the Department for Education in the UK launched a nationwide campaign to encourage reporting of child abuse and neglect last year, called 'Together we can tackle child abuse'.[1] Increasing reporting might help to decrease the actual incidence of maltreatment according to Becker's model of crime (Becker, 1968), as perpetrators face a higher threat of detection. The NCANDS dataset that we use in Chapter 1 could be a great place to start, as it contains information on the identity of the reporter and records the date of the report.

Turning to Chapter 2, we find that fathers are 4.5 percentage points less likely to be employed and work 2.2 fewer hours per week on the intensive margin, on average, after having a first born son rather than a daughter in the UK. We find evidence that degree educated fathers and their partners both take time off work to increase the total parental time received by sons, which may be because boys' development is less advanced than girls' development at a young age. We also find evidence that non-degree educated fathers take time off work to reduce the burden of childcare on their partners, possibly to help their partners to return to work sooner after childbirth.

One avenue for future research is to evaluate the impact of Shared Parental Leave, which was introduced close to the end of our sample period in April 2015. Shared Parental Leave enables parents to split fifty-two weeks of leave between them however they choose, including in multiple blocks separated by periods of work, and replaces a previous system

[1] See: https://tacklechildabuse.campaign.gov.uk.

where fathers were only eligible for two weeks of statutory paternity leave. This scheme therefore reduces the costs of substituting mother for father parental inputs early in a child's life. It will be interesting to see at what rate fathers take up this option, how this changes the division of childcare between mothers and fathers early in a child's life, and in turn whether this changes the gender of first child effect. In addition, as more rounds of the UKHLS become available, future research can ask whether the difference in parental inputs received between boys and girls of degree educated fathers has an impact on later life outcomes for those children.

Finally, in Chapter 3, I find that an additional woman elected to the State Legislative Assembly in rural India increases the probability that a child in her district can do subtraction by 4 percentage points and can read at a Grade 1 level by 3.4 percentage points. The most plausible explanation for this result, which occurs in the next election year, is that politicians increase effort around the time of an election in the hope of getting re-elected, and for women that effort is particularly concentrated on education. Future research might then ask where additional effort by male politicians is targeted during election years.

The main priority for future research in this literature, however, is to continue building a body of evidence which examines the impact of pedagogical and governance interventions on child learning in India. The existing literature demonstrates that just providing more school inputs like books and libraries, or improving infrastructure, has little effect on its own.

References

ADDA, J., DUSTMANN, C. & STEVENS, K., 2017. The Career Costs of Children. *Journal of Political Economy*, **125**(2).

AIZER, A., 2010. The Gender Wage Gap and Domestic Violence. *American Economic Review*, **100**(4), 1847.

AKABAYASHI, H., 2006. An Equilibrium Model of Child Maltreatment. *Journal of Economic Dynamics & Control*, **30**, 993–1025.

ALESINA, A. & LA FERRARA, E., 2005. Preferences for Redistribution in the Land of Opportunities. *Journal of Public Economics*, **89**(5-6), 897–931.

ALLEGRETTO, S., DUBE, A., REICH, M. & ZIPPERER, B., 2017. Credible Research Designs for Minimum Wage Studies. *Industrial and Labor Relations Review*, **70**(3).

ALMOND, D. & CURRIE, J., 2011. Killing Me Softly: The Fetal Origins Hypothesis. *Journal of Economic Perpsectives*, **25**(3), 153–72.

ALMOND, D., CURRIE, J. & DUQUE, V., 2017. Childhood Circumstances and Adult Outcomes: Act II. *NBER Working Papers*, (23017).

ALMOND, D., EDLUND, L. & MILLIGAN, K., 2009. Son Preference and the Persistence of Culture: Evidence from Asian Immigrants to Canada. *NBER Working Papers*.

Anderberg, D., Rainer, H., Wadsworth, J. & Wilson, T., 2016. Unemployment and Domestic Violence: Theory and Evidence. *The Economic Journal*, **126**(597), 1947–1979. ISSN 1468-0297.

Angrist, J. & Evans, W., 1998. Children and Their Parents' Labour Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*, **88**(3), 450–77.

Asher, S. & Novosad, P., 2015. Politics and Local Economic Growth: Evidence from India.

Autor, D., Figlio, D., Karbownik, K., Roth, J. & Wasserman, M., 2015. Family Disadvantage and the Gender Gap in Behavioural and Educational Outcomes. *Northwestern University Working Paper Series*.

Autor, D. & Wasserman, M., 2013. Wayward Sons: The Emerging Gender Gap in Education and Labour Markets. *Technical Report, Third Way*.

Baker, M. & Milligan, K., 2013. Boy-Girl Differences in Parental Time Investments: Evidence from Three Countries. *NBER Working Papers*, (18893).

Banerjee, A., Cole, S., Duflo, E. & Linden, L., 2007. Remedying Education: Evidence from Two Randomised Experiments in India. *Quarterly Journal of Economics*, **122**(3), 1235–1264.

Barro, R., 2010. The Folly of Subsidising Unemployment. *Wall Street Journal*.

Bartik, T., 1991. Boon or Boondoggle? The Debate over State and Local Economic Development Policies. *WE Upjohn Institute for Employment Research*.

Beaman, L., Duflo, E., Pande, R. & Topalova, P., 2012. Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India. *Science*, **335**(6068), 582–586.

Becker, G., 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy*, **76**(2), 169–217.

Becker, G. S., 1993. Nobel Lecture: The Economic Way of Looking at Behavior. *Journal of Political Economy*, **101**(3), 385–409.

Ben-Shalom, Y., Moffit, R. & Scholz, J., 2011. An Assessment of the Effectiveness of Anti-Poverty Programs in the United States. *Working Papers, the John Hopkins University*, (579).

Berger, L. M. & Waldfogel, J., 2011. Economic Determinants and Consequences of Child Maltreatment. *OECD Social, Employment, and Migration Working Papers*, (111).

Bertrand, M. & Pan, J., 2013. The Trouble with Boys: Social Influences and the Gender Gap in Disruptive Behaviour. *American Economic Journal: Applied Economics*, **5**(1), 32–64.

Besley, T., Pande, R., Rahman, L. & Rao, V., 2004. The Politics of Public Good Provision: Evidence from Indian Local Governments. *Journal of the European Economic Association*, **2**(2-3), 416–426.

Bhalotra, S., Cassan, G., Clots-Figueras, I. & Iyer, L., 2013. Religion, Politician Identity and Development Outcomes: Evidence from India. *NBER Working Papers*, (19173).

Bitler, M. & Hoynes, H., 2016. The More Things Change, the More They Stay the Same? The Safety Net and Poverty in the Great Recession. *Journal of Labor Economics*, **34**(1).

Bitler, M., Hoynes, H. & Kuka, E., 2017. Child Poverty, the Great Recession, and the Social Safety Net in the United States. *Journal of Policy Analysis and Management*, **36**(2), 358–389.

Bitler, M. & Zavodny, M., 2004. Child Maltreatment, Abortion Availability, and Economic Conditions. *Review of Economics of the Household*, **2**(2), 119–141.

BLANCHARD, O. & KATZ, L., 1992. Regional Evolutions. *Brookings Papers on Economic Activity*, **1**, 1–75.

BLAU, F. & KAHN, L., 2000. Gender Differences in Pay. *Journal of Economic Perpsectives*, **14**(4), 75–99.

BLOCH, F. & RAO, V., 2002. Terror as a Bargaining Instrument: A Case Study of Dowry Violence in Rural India. *American Economic Review*, **92**(4), 1029–1043.

BLOCK, R. & KREBS, N., 2005. Failure to Thrive as a Manifestation of Child Neglect. *Pediatrics*, **116**(5), 1234–7.

BOARDMAN, J., FINCH, B., ELLISON, C., WILLIAMS, D. & JACKSON, J., 2001. Neighbourhood Disadvantage, Stress, and Drug Use Among Adults. *Journal of Health and Social Behaviour*, **42**, 151–65.

BORKUM, E., HE, F. & LINDEN, L., 2013. The Effects of School Libraries on Language Skills: Evidence from a Randomised Controlled Trial in India. *IZA Discussion Paper Series*, (7267).

BROOKS-GUNN, J., SCHNEIDER, W. & WALDFOGEL, J., 2013. The Great Recession and the Risk for Child Maltreatment. *Child Abuse & Neglect*, **37**(10), 721–729.

BRUINS, M., 2016. TANF, Child Care and Child Well-Being in Sole-Parent Families.

BUCKLEY, C. & SECRET, M., 2011. Caseworkers Dispirited Over Charges in Girl's Death. *New York Times*.

CAMERON, C. & MILLER, D., 2015. A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, **50**(2), 317–373.

CANCIAN, M., YANG, M.-Y. & SLACK, K. S., 2013. The Effect of Additional Child Support Income on the Risk of Child Maltreatment. *Social Service Review*, **87**(3), 417–437.

CARD, D. & LEVINE, P., 2000. Extended Benefits and the Duration of UI Spells: Evidence from the New Jersey Extended Benefit Program. *Journal of Public Economics*, **78**, 107–138.

CASE, A. & DEATON, A., 2015. Rising Morbidity and Mortality in Midlife Among White Non-Hispanic Americans in the 21st Century. *Proceedings of the National Academy of Sciences*, **112**(49), 15078–15083.

CHARLES, K. & STEPHENS, M., 2004. Job Displacement, Disability, and Divorce. *Journal of Labor Economics*, **22**(2), 489–522.

CHATTOPADHYAY, R. & DUFLO, E., 2004. Women as Policy Makers: Evidence from a Randomised Policy Experiment in India. *Econometrica*, **72**(5), 1409–1443.

CHILD WELFARE INFORMATION GATEWAY, 2014. Definitions of Child Abuse and Neglect. U.S. Department of Health and Human Services.

CHILDREN'S BUREAU, 2003. Child Protective Services: A Guide for Caseworkers. Child Abuse and Neglect User Manual Series.

CHOI, H., JOESCH, J. & LUNDBERG, S., 2005. Work and Family: Marriage, Children, Child Gender and the Work Hours and Earnings of West German Men. *IZA Discussion Paper Series*, (1761).

CLOTS-FIGUERAS, I., 2011. Women in Politics: Evidence from the Indian States. *Journal of Public Economics*, **95**(7-8), 664–690.

CLOTS-FIGUERAS, I., 2012. Are Female Leaders Good for Education? Evidence from India. *American Economic Journal: Applied Economics*, **4**(1), 212–244.

CULLEN, J. & GRUBER, J., 2000. Does Unemployment Insurance Crowd Out Spousal Labour Supply? *Journal of Labor Economics*, **18**(3), 546–572.

CUNHA, F., ELO, I. & CULHANE, J., 2013. Eliciting Maternal Expectations about the Technology of Cognitive Skill Formation. *NBER Working Papers*, (19144).

CURRIE, J. & TEKIN, E., 2012. Understanding the Cycle Childhood Maltreatment and Future Crime. *Journal of Human Resources*, **47**(2), 509–549.

CURRIE, J. & WIDOM, C., 2010. Long-Term Consequences of Child Abuse and Neglect on Adult Economic Well-being. *Child Maltreatment*, **15**(2), 111–120.

DAHL, G. & MORETTI, E., 2008. The Demand for Sons. *The Review of Economic Studies*, **75**(4), 1085–1120.

DE ZEGHER, F., DEVLIEGER, H. & EECKELS, R., 1999. Fetal Growth: Boys Before Girls. *Hormone Research*, **51**(5), 258–9.

DIPRETE, T. & JENNINGS, J., 2012. Social and Behavioural Skills and the Gender Gap in Early Educational Achievement. *Social Science Research*, **41**(1), 1–15.

DOIRON, D. & MENDOLIA, S., 2012. The Impact of Job Loss on Family Dissolution. *Journal of Population Economics*, **25**(1), 367–398.

DUBE, A., LESTER, T. W. & REICH, M., 2010. Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties. *Review of Economics and Statistics*, **92**(4), 945–964.

DUBE, S., FELITTI, V., DONG, M., CHAPMAN, D., GILES, W. & ANDA, R., 2003. Childhood abuse, Neglect, and Household Dysfunction and the Risk of Illicit Drug Use: the Adverse Childhood Experiences Study. *Pediatrics*, **111**(3), 564–572.

DUBUC, S. & COLEMAN, D., 2007. An Increase in the Sex Ratio of Births to India-Born Mothers in England and Wales: Evidence for Sex-Selective Abortion. *Population and Development Review*, **33**(2), 383–400.

DUFLO, E., HANNA, R. & RYAN, S., 2012. Incentives Work: Getting Teachers to Come to School. *American Economic Review*, **102**(4), 1241–1278.

Dwyer-Lindgren, L., Flaxman, A., Hansen, G., Murray, C. & Mokdad, A., 2015. Drinking Patterns in US Counties From 2002 to 2012. *American Journal of Public Health: Research and Practise*, **105**(6).

Eckholm, E., 2009. States Slashing Social Programs for Vulnerable. *New York Times*.

EdCIL Limited, 2010. Role of Block and Cluster Resource Centres in Providing Academic Support to Elementary Schools.

Elder Jr, G. H., 1974. *Children of the Great Depression: Social Change in Life Experience*. Boulder, CO: Westview Press.

Eliason, M., 2012. Lost Jobs, Broken Marriages. *Journal of Population Economics*, **25**(4), 1365–1397.

Eliason, M. & Storrie, D., 2009. Job Loss is Bad for Your Health - Swedish Evidence on Cause-Specific Hospitalisation Following Involuntary Job Loss. *Social Science and Medicine*, **68**, 1396–1406.

Else-Quest, N., Hyde, J., Goldsmith, H. & Van Hulle, C., 2006. Gender Differences in Temperament: A Meta-Analysis. *Psychological Bulletin*, **132**(1), 33–72.

Eswaran, M. & Malhotra, N., 2011. Domestic Violence and Women's Autonomy in Developing Countries: Theory and Evidence. *Canadian Journal of Economics*, **44**(4), 1222–1263.

Fagernas, S. & Pelkonen, P., 2014. Politics Before Pupils? Electoral Cycles and School Resources in India. *IZA Discussion Paper Series*, (8366).

Fan, X., Fang, H. & Markussen, S., 2017. Parental Employment, Absence and Children's Educational Gender Gap.

Farmer, A. & Tiefenthaler, J., 1997. An Economic Analysis of Domestic Violence. *Review of Social Economy*, **55**(3), 337–358.

Fein, D. J. & Lee, W. S., 2003. The Impacts of Welfare Reform on Child Maltreatment in Delaware. *Children and Youth Services Review*, **25**(1), 83–111.

Feldman, J., 2010. Do Doctors in America Turn Away the Uninsured? *The Huffington Post*.

French, R. & Kingdon, G., 2010. The Relative Effectiveness of Private and Government Schools in Rural India: Evidence from ASER Data. *DoQSS Working Papers*, **10**(03).

Frioux, S., Wood, J., Fakeye, O., Luan, X., Localio, R. & Rubin, D., 2014. Longitudinal Association of County-level Economic Indicators and Child Maltreatment Incidents. *Maternal and Child Health Journal*, **18**(9), 2202–2208.

Goldin, C., 2014. A Grand Gender Convergence: Its Last Chapter. *American Economic Review*, **104**(4), 1091–1119.

Goldin, C., Katz, L. & Kuziemko, I., 2006. The Homecoming of American College Women: The Reversal of the College Gender Gap. *NBER Working Papers*, (12139).

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

Hagedorn, M., Karahan, F., Manovskii, I. & Mitman, K., 2013. Unemployment Benefits and Unemployment in the Great Recession: The Role of Macro Effects. Tech. rep., National Bureau of Economic Research.

Hagele, D., 2005. The Impact of Maltreatment on the Developing Child. *NC Med J*, **66**(5), 356–9.

Halpern, D., 1997. Sex Differences in Intelligence: Implications for Education. *American Psychologist*, **52**, 1091–1101.

Hanushek, E. & Woessman, L., 2008. The Role of Cognitive Skills in Economic Development. *Journal of Economic Literature*, **46**(3), 607–668.

HECKMAN, J. & MOSSO, S., 2014. The Economics of Human Development and Social Mobility. *Annual Review of Economics*, **6**(689-733).

HECKMAN, J., PINTO, R. & SAVELYEV, P., 2013. Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review*, **103**(6), 2052–86.

HONORE, B., 1992. Trimmed Lad and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects. *Econometrica*, **60**(3), 533–565.

HUANG, M. I., O'RIORDAN, M. A., FITZENRIDER, E., MCDAVID, L., COHEN, A. R. & ROBINSON, S., 2011. Increased Incidence of Nonaccidental Head Trauma in Infants Associated with the Economic Recession: Clinical Article. *Journal of Neurosurgery: Pediatrics*, **8**(2), 171–176.

ICHINO, A., LINDSTROM, E. & VIVIANO, E., 2014. Hidden Consequences of a First-Born Boy for Mothers. *Economics Letters*, **123**, 274–278.

ISAACS, J. & HEALY, O., 2014. Public Supports When Parents Lose Work. *Low-Income Working Families Paper*, (28).

KNOW VIOLENCE IN CHILDHOOD, 2017. Ending Violence in Childhood. New Delhi: Global Report 2017.

KOHLER, H., BEHRMAN, J. & SKYTTHE, A., 2005. Partner + Children = Happiness? The Effects of Partnerships and Fertility on Well-Being. *Population and Development Review*, **31**(3), 407–445.

KREMER, M., BRANNEN, C. & GLENNERSTER, R., 2013. The Challenge of Education and Learning in the Developing World. *Science*, **340**(6130), 297–300.

LEE, D., BROOKS-GUNN, J., MCLANAHAN, S. S., NOTTERMAN, D. & GARFINKEL, I., 2013. The Great Recession, Genetic Sensitivity, and Maternal Harsh Parenting. *Proceedings of the National Academy of Sciences*, **110**(34), 13780–13784.

LEE LUCA, D., OWENS, E. & SHARMA, G., 2015. Can Alcohol Prohibition Reduce Violence Against Women? *American Economic Review: Papers and Proceedings*, **105**(5), 625–9.

LEININGER, L., 2009. Partial-Year Insurance Coverage and the Health Care Utilization of Children. *Medical Care Research and Review*, **66**(1).

LEVESON, B., 2012. An Inquiry into the Culture, Practices and Ethics of the Press. National Archives.

LEVITT, S., LIST, J., NECKERMANN, S. & SADOFF, S., 2012. The Behaviouralist Goes to School: Leveraging Behavioural Economics to Improve Educational Performance. *NBER Working Papers*, (18165).

LIN, M., 2008. Does Unemployment Increase Crime? *Journal of Human Resources*, **43**(2), 413–436.

LINDO, J., SCHALLER, J. & HANSEN, B., 2013. Caution! Men Not at Work: Gender-specific Labor Market Conditions and Child Maltreatment. *NBER Working Papers*, (18994).

LUNDBERG, S., 2005. The Division of Labour by New Parents: Does Child Gender Matter? *IZA Discussion Paper Series*, (1787).

LUNDBERG, S., POLLAK, R. & STEARNS, J., 2016. Family Inequality: Diverging Patterns in Marriage, Cohabitation and Childbearing. *NBER Working Papers*, (22078).

LUNDBERG, S. & ROSE, E., 2002. The Effects of Sons and Daughters on Men's Labour Supply and Wages. *Review of Economics and Statistics*, **84**(2), 251–268.

LUNDBERG, S. & ROSE, E., 2003. Child Gender and the Transition to Marriage. *Demography*, **40**(2), 333–349.

LUNDBERG, S., WULFF-PABILONIA, S. & WARD-BATTS, J., 2007. Time Allocation of Parents and Investment in Sons and Daughters.

LUNDBORG, P., PLUG, E. & RASMUSSEN, A., 2014. Fertility Effects on Female Labour Supply: IV Evidence from IVF Treatments. *IZA Discussion Paper Series*, (8609).

LUTCHMAYA, S., BARON-COHEN, S. & RAGGATT, P., 2001. Foetal Testosterone and Vocabulary Size in 18 and 24 Month-old Infants. *Infant Behavior and Development*, **24**(4), 418–424.

LUTCHMAYA, S., BARON-COHEN, S. & RAGGATT, P., 2002. Foetal Testosterone and Eye Contact in 12 Month-old Human Infants. *Infant Behavior and Development*, **25**(327-335).

LUTTMER, E., 2005. Neighbours as Negatives: Relative Earnings and Well-Being. *Quarterly Journal of Economics*, **120**(3), 963–1002.

MAMMEN, K., 2011. Fathers' Time Investments in Children: Do Sons Get More? *Journal of Population Economics*, **24**.

MARKOWITZ, S. & GROSSMAN, M., 1998. Alcohol Regulation and Domestic Violence towards Children. *Contemporary Economic Policy*, **16**(3), 309–320.

MATHEWS, F., JOHNSON, P. & NEIL, A., 2008. You Are What Your Mother Eats: Evidence for Maternal Preconception Diet Influencing Foetal Sex in Humans. *Proceedings of the Royal Society B: Biological Sciences*, **275**(1643).

MAZUMDER, B. & SEESKIN, Z., 2015. Breakfast Skipping, Extreme Commutes, and the Sex Composition at Birth. *Biodemography and Social Biology*, **61**(2), 187–208.

MCEWAN, P., 2014. Improving Learning in Primary Schools of Developing Countries: A Meta-Analysis of Randomised Experiments. *Review of Educational Research*.

MULLIGAN, C. B., 2012. *The Redistribution Recession: How Labor Market Distortions Contracted the Economy*. Oxford University Press.

MURALIDHARAN, K., 2013. Priorities for Primary Education Policy in India's 12th Five Year Plan. *India Policy Forum 2012-13*, **9**, 1–46.

Muralidharan, K. & Sundararaman, V., 2010. The Impact of Diagnostic Feedback to Teachers on Student Learning: Experimental Evidence from India. *The Economic Journal*, **120**(546), F187–F203.

Muralidharan, K. & Sundararaman, V., 2011. Teacher Performance Pay: Experimental Evidence from India. *Journal of Political Economy*, **119**(1), 39–77.

Muralidharan, K. & Sundararaman, V., 2013. Contract Teachers: Experimental Evidence from India. *NBER Working Paper*, (19440).

Nicoletti, C. & Tonei, V., 2015. The Response of Parental Time Investments to the Child's Abilities and Health.

Olivetti, C. & Petrongolo, B., 2016. The Evolution of Gender Gaps in Industrialised Countries. *NBER Working Papers*, (21887).

Pande, R., 2003. Can Mandated Political Representation Increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India. *American Economic Review*, **93**(4), 1132–1151.

Paxson, C. & Waldfogel, J., 1999. Parental Resources and Child Abuse and Neglect. *American Economic Review*, **89**(2), 239–244.

Paxson, C. & Waldfogel, J., 2002. Work, Welfare, and Child Maltreatment. *Journal of Labor Economics*, **20**(3), 435–474.

Paxson, C. & Waldfogel, J., 2003. Welfare Reforms, Family Resources, and Child Maltreatment. *Journal of Policy Analysis and Management*, **22**(1), 85–113.

Pedersen, J., 1980. Ultrasound Evidence of Sexual Difference in Fetal Size in First Trimester. *British Medical Journal*, **281**(6250).

Pennsylvania Department of Human Services, 2016. Annual Child Protective Services Report 2016.

PHILLIPS, M., 2011. Parenting, Time Use, and Disparities in Academic Outcomes. Ch. 10 in Whither Opportunity, Duncan, G. and Murnane, R.

POWDTHAVEE, N. & OSWALD, A., 2010. Daughters and Left-Wing Voting. *Review of Economics and Statistics*, **92**(2), 213–227.

POWDTHAVEE, N., OSWALD, A. & WU, S., 2009. The Effects of Daughters on Health Choices and Risk Behaviour. *Discussion Papers in Economics, University of York*, **10**(3).

PRATHAM, 2015. Annual Status of Education Report 2014.

PRESTON, J., 2013. Grand Jury Cites Deaths of Children in Inquiry. *New York Times*.

PRICE, J., 2008. Parent-Child Quality Time: Does Birth Order Matter? *Journal of Human Resources*, **43**(1), 240–265.

PRITCHETT, L., 2013. *The Rebirth of Education*. Center for Global Development.

RAISSIAN, K., 2015. Does Unemployment affect Child Abuse Rates? Evidence from New York State. *Child Abuse & Neglect*, **48**, 1–12.

READY, D., LOGERFO, L., BURKAM, D. & LEE, V., 2005. Explaining Girls' Advantage in Kindergarten Literacy Learning: Do Classroom Behaviors Make a Difference. *Elementary School Journal*, **106**(1), 21–38.

ROTHSTEIN, J., 2011. Unemployment Insurance and Job Search in the Great Recession. *NBER Working Papers*, (17534).

RUHM, C. J., 2000. Are Recessions Good for Your Health? *Quarterly Journal of Economics*.

RUHM, C. J., 2003. Good Times Make You Sick. *Journal of Health Economics*, **22**(4), 637–658.

Ruhm, C. J., 2005. Healthy Living in Hard Times. *Journal of Health Economics*, **24**(2), 341–363.

Ruhm, C. J. & Black, W. E., 2002. Does Drinking Really Decrease in Bad Times? *Journal of Health Economics*, **21**(4), 659–678.

Rutter, M., Caspi, A., Fergusson, D., Horwood, L., Goodman, R., Maughan, B., Moffit, T., Meltzer, H. & Carroll, J., 2004. Sex Differences in Developmental Reading Disability: New Findings from 4 Epidemiological Studies. *Journal of the American Medical Association*, **291**.

Santos, F., 2013. Thousands of Ignored Child Abuse Allegations Plague Arizona Welfare Agency. *New York Times*.

Schneider, W., Waldfogel, J. & Brooks-Gunn, J., 2017. The Great Recession and Risk for Child Abuse and Neglect. *Children and Youth Services Review*, **72**, 71–81.

Sedlak, A., Mettenburg, J., Basena, M., Peta, I., McPherson, K. & Greene, A., 2010. Fourth National Incidence Study of Child Abuse and Neglect (NIS-4). *Washington, DC: US Department of Health and Human Services*, **9**.

Seiglie, C., 2004. Understanding Child Outcomes: An Application to Child Abuse and Neglect. *Review of Economics of the Household*, **2**(2), 143–160.

Selden, T. & Hudson, J., 2006. Access to Care and Utilization Among Children: Estimating the Effects of Public and Private Coverage. *Medical Care*, **44**(5), 19–26.

Shook, K. & Testa, M., 1997. Cost-savings Evaluation of the Norman Program: Final Report to the Department of Children and Family Services. *Chicago: Illinois Department of Children and Family Services*.

Silles, M., 2016. The Impact of Children on Women's Labour Supply and Earnings in the UK: Evidence Using Twin Births. *Oxford Economic Papers*, **68**(1), 197–216.

Springer, K., Sheridan, J., Kuo, D. & Carnes, M., 2007. Long-term Physical and Mental Health consequences of Childhood Physical Abuse: Results from a Large Population-based Sample of Men and Women. *Child Abuse & Neglect*, **31**(5), 517–530.

Starfield, B. & Shi, L., 2004. The Medical Home, Access to Care, and Insurance: a Review of Evidence. *Pediatrics*, **113**(5), 1493–8.

Stephens-Davidowitz, S., 2013. Unreported Victims of an Economic Downturn. Harvard University.

Szatmari, P., Offord, D. & Boyle, M., 1989. Ontario Child Health Study: Prevalence of Attention Deficit Disorder with Hyperactivity. *Journal of Child Psychology and Psychiatry*, **30**(2), 219–230.

Tarullo, A. & Gunnar, M., 2006. Child Maltreatment and the Developing HPA Axis. *Hormones and Behavior*, **50**, 632–639.

Tauchen, H., Witte, A. & Long, S., 1991. Domestic Violence: A Nonrandom Affair. *International Economic Review*, **32**(2), 491–511.

Tefft, N., 2011. Insights on Unemployment, Unemployment Insurance, and Mental Health. *Journal of Health Economics*, **30**(2), 258–264.

U.S. Census Bureau, 2013. Individual State Descriptions: 2012. 2012 Census of Governments.

US Department of Health and Human Services, 2012. Child Maltreatment 2012. Tech. rep., US Department of Health and Human Services, Administration for Children and Families.

US Department of Health and Human Services, 2015. Child Maltreatment 2015. Tech. rep., US Department of Health and Human Services, Administration for Children and Families.

Voldner, N., Froslie, K., Godang, K., Bollerslev, J. & Henriksen, T., 2009. Determinants of Birth Weight in Boys and Girls. *Human Ontogenetics*, **3**(1), 7–12.

Weinberg, B. A., 2001. An Incentive Model of the Effect of Parental Income on Children. *Journal of Political Economy*, **109**(2), 266–280.

Wells, J., 2000. Natural Selection and Sex Differences in Morbidity and Mortality in Early Life. *Journal of Theoretical Biology*, **202**, 65–76.

Wiedrich, K., Rice, S., Lebaron, S. & Weisman, H., 2017. On Track or Left Behind? Findings from the 2017 Prosperity Now Scorecard. Prosperity Now Scorecard.

Williams, E., 2017. States Can Adopt or Expand Earned Income Tax Credits to Build a Stronger Future Economy. Center on Budget and Policy Priorities.

Woolley, F., 2011. A Rant on Inverse Hyperbolic Sine Transformations. Worthwhile Canadian Initiative Blog.

Wozniak, A., 2010. Are College Graduates More Responsive to Distant Labour Market Opportunities. *Journal of Human Resources*, **45**(4), 944–970.

Wulff-Pabilonia, S. & Ward-Batts, J., 2007. The Effect of Child Gender on Parents' Labour Supply: An Examination of Natives, Immigrants and their Children. *American Economic Review*, **97**(2), 402–406.

Yeung, W., Sandberg, J., Davis-Kean, P. & Hofferth, S., 2001. Children's Time with Fathers in Intact Families. *Journal of Marriage and Family*, **63**, 136–154.