Reaching For Equality

Essays in Education and Gender Economics

Ericsson, Sanna

2020

Document Version:
Publisher's PDF, also known as Version of record

Link to publication

Citation for published version (APA):
Ericsson, S. (2020). *Reaching For Equality: Essays in Education and Gender Economics* (1 ed.). [Doctoral Thesis (compilation), Lund University School of Economics and Management, LUSEM]. Lund University (Media-Tryck).

Total number of authors:
1

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Reaching For Equality

Essays in Education and Gender Economics

Sanna Ericsson

Lund Economic Studies

Number 221

LUND UNIVERSITY

Reaching For Equality

Reaching For Equality

Essays in Education and Gender Economics

Sanna Ericsson

Lund University

DOCTORAL DISSERTATION
By due permission of the School of Economics and Management,
Lund University, Sweden.
To be defended at Holger Crafoords Ekonomicentrum EC3:210
on Friday May 15, 2020, at 10.15 am.

Faculty Opponent
Andreas Kotsadam,
The Frisch Centre and University of Oslo

Organization
LUND UNIVERSITY
Department of Economics
P.O. Box 7082
S-220 07 Lund, Sweden

Document name
DOCTORAL DISSERTATION

Date of issue
May 15, 2020

CODEN:
ISSN 0460-0029

Author
Sanna Ericsson

Sponsoring organization

Title and subtitle
Reaching For Equality: Essays in Education and Gender Economics

Abstract
This thesis consists of three self-contained papers that all relate to the understanding of equality. The first chapter investigates the effects of preschool attendance on children in Kenya and Tanzania. We use a within-household estimator and data from nationally representative surveys of school-age children's literacy and numeracy skills, which include retrospective information on preschool attendance. In both countries, school entry rules are not strictly enforced, and children who attend preschool often start primary school late. At ages 7-9, these children have thus attended fewer school grades than their same-aged peers without pre-primary education. However, they catch up over time: at ages 13-16, children who went to preschool have attended about the same number of school grades and score about 0.10 standard deviations higher on standardized tests in both countries. They are also 3 (5) percentage points more likely to achieve basic literacy and numeracy in Kenya (Tanzania).

The second chapter investigates the interaction between cultural norms and neighbourhood characteristics. I estimate the effect of cultural gender norms on the gender gap in math, and explore whether this effect is mitigated by municipality gender equality. I use high-quality Swedish administrative data on the results of national standardised math tests. To separate the effect of cultural gender norms from formal institutions, I estimate the effect of mothers' source-country gender norms on the gender gap in math for second-generation immigrants. By contrasting the outcomes of opposite-sex siblings, I control for everything that correlates with the source country but that is unrelated to gender. I show that the sibling gender gap in math increases with mothers' adherence to traditional gender norms, such that girls with more gender-traditional mothers perform worse relative to their brothers. To investigate whether the cultural gender norm effect can be mitigated by municipality gender equality, I exploit a refugee placement policy to obtain random variation in municipality characteristics. I show that municipality gender equality can almost completely mitigate the negative cultural norm effect. Taken together, my results imply that while cultural gender norms play an important role for the gender gap in math, they are not immune to the effects of neighbourhood exposure.

The third chapter estimates the effect of female economic empowerment on domestic violence. I use individual level data from high-quality Swedish administrative registers on hospital visits relating to assault. I proxy female economic empowerment with a measure of women's potential earnings, caused by local changes in female-specific labour demand. I show that the causal effect of increasing women's potential earnings on domestic violence is positive and substantial. In addition, I show that increasing women's potential earnings increase the husbands' risk of destructive behaviour, such as stress, anxiety, substance abuse and assault. Taken together, these results indicate that improving women's financial independence triggers a male backlash response, even in a gender-equal country like Sweden.

Key words
Preschool, Education, Sub-Saharan Africa, Cultural Gender Norms, Math Gender Gap, Epidemiological Approach, Refugee Placement Policy, Sibling Fixed Effects, Domestic Violence, Household Bargaining, Male Backlash, Local Labour Demand

Classification system and/or index terms (if any)
JEL Classification: I21, J24, I24, J15, J16, Z13, D13, I12, J12

Supplementary bibliographical information

Language
English

ISSN and key title
0460-0029 Lund Economic Studies no. 221

ISBN
978-91-7895-498-8 (print)
978-91-7895-499-5 (pdf)

Recipient's notes

Number of pages
173

Price

Security classification

Distribution by: Department of Economics, P.O. Box 7082, S-220 07 LUND, Sweden

Signature

Date 2020-03-31

Reaching For Equality

Essays in Education and Gender Economics

Sanna Ericsson

LUND UNIVERSITY

LUND ECONOMIC STUDIES NUMBER 221

Lund University School of Economics and Management, Department of Economics

ISBN: 978-91-7895-498-8 (print)
ISBN: 978-91-7895-499-5 (pdf)
ISSN: 0460-0029 Lund Economic Studies no. 221

Printed in Sweden by Media-Tryck, Lund University, Lund 2020

NORDIC SWAN ECOLABEL
Printed matter
3041 0903

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

To me, nine years ago, worrying if I made the right study choice:
Don't worry, it will all work out great.

Contents

Abstract

This thesis consists of three self-contained papers that all relate to the understanding of equality. The first chapter investigates the effects of preschool attendance on children in Kenya and Tanzania. We use a within-household estimator and data from nationally representative surveys of school-age children's literacy and numeracy skills, which include retrospective information on preschool attendance. In both countries, school entry rules are not strictly enforced, and children who attend preschool often start primary school late. At ages 7-9, these children have thus attended fewer school grades than their same-aged peers without pre-primary education. However, they catch up over time: at ages 13-16, children who went to preschool have attended about the same number of school grades and score about 0.10 standard deviations higher on standardized tests in both countries. They are also 3 (5) percentage points more likely to achieve basic literacy and numeracy in Kenya (Tanzania).

The second chapter investigates the interaction between cultural norms and neighbourhood characteristics. I estimate the effect of cultural gender norms on the gender gap in math, and explore whether this effect is mitigated by municipality gender equality. I use high-quality Swedish administrative data on the results of national standardised math tests. To separate the effect of cultural gender norms from formal institutions, I estimate the effect of mothers' source-country gender norms on the gender gap in math for second-generation immigrants. By contrasting the outcomes of opposite-sex siblings, I control for everything that correlates with the source country but that is unrelated to gender. I show that the sibling gender gap in math increases with mothers' adherence to traditional gender norms, such that girls with more gender-traditional mothers perform worse relative to their brothers. To investigate whether the cultural gender norm effect can be mitigated by municipality gender equality, I exploit a refugee placement policy to obtain random variation in municipality characteristics. I show that municipality gender equality can almost completely mitigate the negative cultural norm effect. Taken together, my results imply that while cultural gender norms play an important role for the gender gap in math, they are not immune to the effects of neighbourhood exposure.

The third chapter estimates the effect of female economic empowerment on domestic violence. I use individual level data from high-quality Swedish administrative registers on hospital visits relating to assault. I proxy female economic empowerment with a measure of women's potential earnings, caused by local changes in female-specific labour demand. I show that the causal effect of increasing women's potential earnings on domestic violence is positive and substantial. In addition, I show that increasing women's potential earnings increase the husbands' risk of destructive behaviour, such as stress, anxiety, substance abuse and assault. Taken together, these results indicate that improving women's financial independence triggers a male backlash response, even in a gender-equal country like Sweden.

Keywords: Preschool, Education, Sub-Saharan Africa, Cultural Gender Norms, Math Gender Gap, Epidemiological Approach, Refugee Placement Policy, Sibling Fixed Effects, Domestic Violence, Household Bargaining, Male Backlash, Local Labour Demand

JEL Classifications: I21, J24, I24, J15, J16, Z13, D13, I12, J12

Acknowledgements

Me ending up where I am right now is the result of a variety of happy coincidences (as well as, of course, a lot of hard work, discipline and an innate love of learning). I did not grow up thinking that I wanted to be an economist (although, to be fair, I do not think a lot of people do), or that I wanted to get a PhD. I wanted to study at university, chose something almost at random, switched a year later, and spent the following three years in and out of agony over whether or not I had made the right choice, and whether or not there are any jobs that require economists (except for banks, where I did not want to work). Now, nine years after I made that first choice, I can confidently say that it all worked out great. This is the best place I could ever have ended up in, and I am so happy that all the coincidences (and all the hard work) led me here. Doing a PhD turned out to be a lot of fun, and a very rewarding experience. Many people have contributed to this journey, and for that I want to thank them sincerely.[1]

First and foremost, I thank my main supervisor, Petter Lundborg. Thank you for seeing straight through the flaws in my research designs, and for always accompanying your comments with suggestions on how to improve, and advice on what to add to my study. It is not a coincidence that many PhD students choose you as their supervisor, because you really know how to encourage and help us move forward.

I was lucky enough to have two co-supervisors. I thank my first co-supervisor, Alessandro Martinello, for being the perfect combination of pushy and inspiring. I always left our meetings feeling excited about research, but with a secret mental list of econometric stuff to Google as soon as I got back to my office. It is partly thanks to you that I started the PhD, as you sparked my interest in applied micro-econometrics when you supervised my master's thesis.

I also thank my second co-supervisor, Therese Nilsson, who started working

[1] I tried to think of many ways to write the acknowledgements without being too cheesy. In the end, I decided to give up, so readers should be prepared for cheesiness.

with me during only the final one and a half years, but who was immediately there to help me with anything I needed. Thank you for reading my drafts and helping me develop my third paper, and for organising seminars and research visits for me during this final year.

I thank Jan Bietenbeck for co-authoring my first paper with me, because I learned so much from you.

Thank you Polina, for creating the best office atmosphere possible, for maintaining the 'good enough'-mentality, and for all the chocolate you have given me over the years. And for all the sparkling wine we have been drinking in our office. Thank you, Hampus and Josefin, for always having your door open and being ready for fun talks during the day. And for all the sparkling wine we have been drinking in your office.

A big thank you to my cohort: Polina, Pol, Yana and Sara, for working hard alongside me and making PhD-life way more fun. I am so happy that we ended up being not just colleagues, but friends. I also want to thank all the other colleagues and friends I have met here at EC and Alpha: Danial, Matthew, Chelsea, Sandra, Devon, Demid, Dominice, Zahra, Hong, Marco, Dominika, Olga, Linn, Adrian, Marcus, Kristoffer, Ovidjius, Charlotta, Prakriti, Emelie, Phillip, Albert, Yousef, Caglar, Vinh, John, Hjördis, Erik, Thomas, Sara Mo, Petra, Alex, Roel, Claes, Olga, Simon, Andreas, Lina-Maria, Margareta, Gunes, Kaveh and Tommy. This is a great workplace, and I am glad I will get to stick around with you all for another year.

Thank you to the great admin staff at EC: Jenny, Nathalie, Rikke, Marie, Anna, Azra, Mariana and Peter, for always answering my questions (sometimes over and over again as I never seem to be able to remember how to fill in the expense reports correctly) and for our late-night dancing at the department Christmas parties.

I thank Heather Royer for hosting me during my research visit in Santa Barbara. I thank Jason Lindo, Núria Rodríguez-Planas, and participants at seminars, conferences and workshops for commenting on my papers and helping me improve them. I thank Hedelius Stiftelse, Stiftelsen för främjande av ekonomisk forskning vid Lunds univesitet, and Stiftelsen Siamon for grants that made it possible for me to travel and attend courses, conferences and the research visit.

But life is more than just EC and Alpha, and a lot of people outside of academia deserve to be thanked for helping me through these years. I thank my mother, for always being ready to analyse, and endlessly discuss, any questions that I have had. For forcing me to dance salsa and for being my closest ally during all these years. I thank my father, for teaching me that doing things 'good enough' is actually the optimal way of doing things. For making me believe in myself and trust that everything

will work out in the end. It is thanks to you that I have been able to get a PhD and still be a happy person.

I thank my oldest friends: Åsa, Linda, Pernie and Elin, for taking me in when we were nine years old and staying with me through 20 years and counting. For realising, even when we were little, that being nerdy is actually kind of cool. I thank Malin and Sofia, for making economics studies fun and for encouraging me to accept the PhD offer. And, of course, for all the Tours of Lund (i.e. some of the best parties I have ever had). I thank My (and David!), for offering housing in Lund when I had none, for queuing tirelessly for the carnival, for being with me almost the entire time after finishing high school, and during eight great years in Skåne.

I thank the Vildanden Playboys: Martin, Ane, Nina, Malin, Jonas, Josef, and Magnus, for amazing times in Lund. Many of you are away doing cool things in cool places now, and (among other things) I miss our frequent evenings at Blekingska's amazingly horrible pub. Special thanks to Jonas, who gives me the daily dose of communism one needs to balance the harsh world of neo-classical economics. I thank Andrea, Karin and Sofia, for great times at Wermland's nation, for all the dinners, all the wine, and the awesome Paris-themed pub we organised.

Finally, I thank Martin, for everything. For all the walks around Alpha, for reading, listening and commenting, for celebrating with me when everything went well, and supporting me when everything went wrong and I was struck by the impostor syndrome. You are simply the most amazing person I have ever met, and you make everything fun.

Lund, March, 2020
Sanna

Introduction

Introduction

1 Equality, education and gender

This thesis consists of three self-contained papers that all relate to the understanding of equality. The first chapter investigates the effects of preschool attendance on children in Kenya and Tanzania. Preschool impacts equality as it can compensate for a lack of early childhood investments by parents. Universal preschool programmes may therefore reduce the impact of any socio-economic gap in parental investments, and contribute to reducing the socio-economic gradient in educational outcomes.[1]

The second and third chapters deal with gender equality, or rather the lack thereof. The second chapter explores the effect of cultural gender norms on the gender gap in math among second-generation immigrants in Sweden, and examines how this effect interacts with exposure to neighbourhood gender equality. The third chapter shows that increasing female economic empowerment increases the risk of domestic violence, hence, it shows that even in a gender-equal country like Sweden there is a risk of a violent backlash response to greater female financial independence.

A second common aspect among the chapters is the methodology. All chapters strive to provide some potential implications for policy and policy makers. In order to advise policy making, it is crucial for a study to be certain of the direction of the cause-and-effect relationship. For example, for a study on an intervention to have meaningful policy implications, it must be able to identify the impact of the intervention and not some spurious underlying relationship.

The ideal set-up to recover causal effects is a randomised controlled trial, where a simple comparison of means between the treated and untreated groups will reflect the causal effect of the intervention. As a perfectly controlled experiment is seldom feasible when we want to study real-world behaviour and outcomes, research

[1] See e.g. Guryan et al. (2008) on socio-economic differences in time spent with ones' children.

in economics has grown to incorporate various methods striving to emulate experimental designs. In recent decades, empirical research has improved its ability to provide meaningful policy recommendations, thanks to better data, cheap computing power and a more stringent approach to research design (Angrist and Pischke, 2010). This strand of economics is commonly referred to as applied microeconometrics, and all three research chapters of this thesis are examples thereof.

1.1 Educational equality: preschool as a remedy for skill deficits in Sub-Saharan Africa

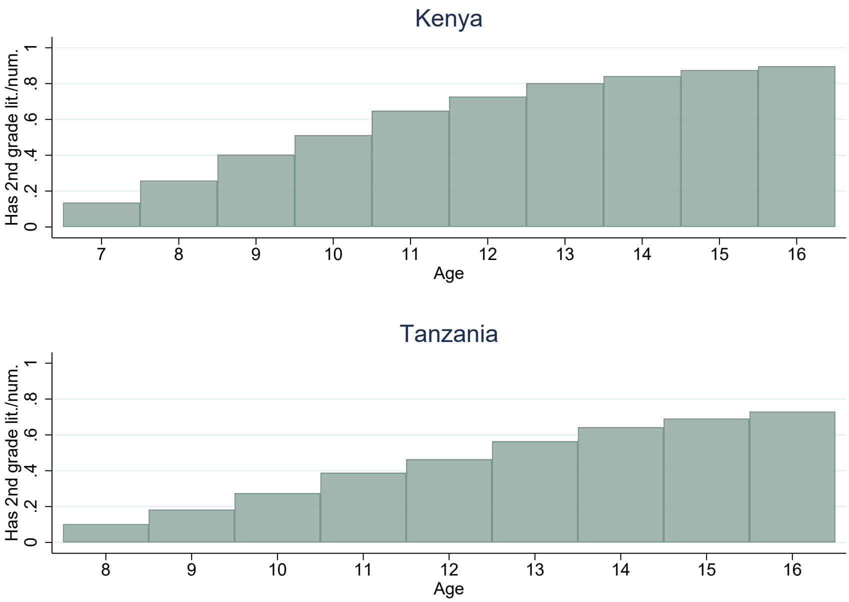
School enrolment in Sub-Saharan Africa has increased substantially over the past two decades, with the gross enrolment rate reaching almost 98% in 2018 (World Bank, 2018b). However, many students fall behind the curriculum early on, and grade repetition and early drop-out are widespread (UNESCO, 2012). Students also seem to learn very little in school. For example, only one in five third-grade students in Eastern Africa manages second-grade literacy and numeracy skills, and less than one third of sixth-grade students in Southern and Eastern Africa can solve a simple subtraction problem (Bietenbeck et al., 2018; Uwezo, 2015).

Figure 1 shows the fraction of children who have the basic competencies they are expected to have attained after two years of schooling, according to the national curricula of Kenya and Tanzania. Children are supposed to be enrolled in second grade when they are seven years old (in Kenya) or eight years old (in Tanzania). As the figure shows, very few of the younger children have the expected knowledge, and a non-negligible fraction of the older children are also missing it.

One possible reason why the students perform poorly is that they enter school unprepared. The theoretical literature on skill formation suggests that skills formed in different stages of life are complimentary, and vary in effectiveness depending on the child's stage of development. Cunha and Heckman (2007) present a model of skill formation over multiple stages of childhood, suggesting it may be optimal to invest relatively more in early stages of childhood rather than in later stages. Skills obtained at an earlier stage persist into the future and augment the skills obtained at a later stage, which increases the efficiency of a child's skill formation (Cunha et al., 2006; Cunha and Heckman, 2007; Cunha et al., 2010). Thus, a lack of education before children start school reduces their efficiency in learning once they start primary school. For this reason, preschool programmes are often seen as a promising way to enhance learning (e.g. World Bank, 2018).

Furthermore, if children from different social backgrounds have different access to schooling opportunities or qualitative parental inputs this could cause intergen-

Figure 1: Fraction of students who manage second grade requirements



Notes: The figure plots the fraction of children who manage second grade literacy and numeracy tests of the Uwezo surveys of 2011 - 2014 for Tanzania, and 2013 - 2014 for Kenya. Data is obtained from Uwezo (2015).

erational transmission of inequality. Due to the complementarity of skills, earlier interventions are most effective if one wishes to remedy the skills gap between children from different socio-economic backgrounds, and a large literature documents positive effects of preschool programmes targeted at disadvantaged children (see e.g. Currie and Thomas, 1995; Currie, 2001; Garces et al., 2002; Deming, 2009).

Less is known, however, about the effects of universal preschool programmes, and especially in developing countries. Dietrichson et al. (2018) reviews the literature on universal preschool programmes, and only five out of the 26 studies reviewed were conducted in a developing country context. Previous literature on developing countries has mainly focused on Asia and Latin America, and is reviewed in Nores and Barnett (2010) and Rao et al. (2014). Studies in Bolivia, Uruguay, Argentina, Eqypt and Indonesia show that preschool attendance has positive effects on children's educational attainment and cognitive skills (Behrman et al., 2004; Berlinski et al., 2008, 2009; Krafft, 2015; Brinkman et al., 2017). In contrast, Bouguen et al. (2018) find negative short-term impacts on test scores of a preschool programme in Cambodia.

In Sub-Saharan Africa, pre-primary education has expanded rapidly, with the gross enrolment ratio growing from 15% in 2000 to 34% in 2018 (World Bank, 2018a). However, the question remains whether these pre-primary programmes are actually effective in improving students' educational outcomes, as the empirical evidence is still scarce.[2] The first chapter of this thesis contributes to the empirical literature on effects of preschool attendance in Sub-Saharan Africa, and evaluates the short- and long-term effects of preschool attendance for children in Kenya and Tanzania.

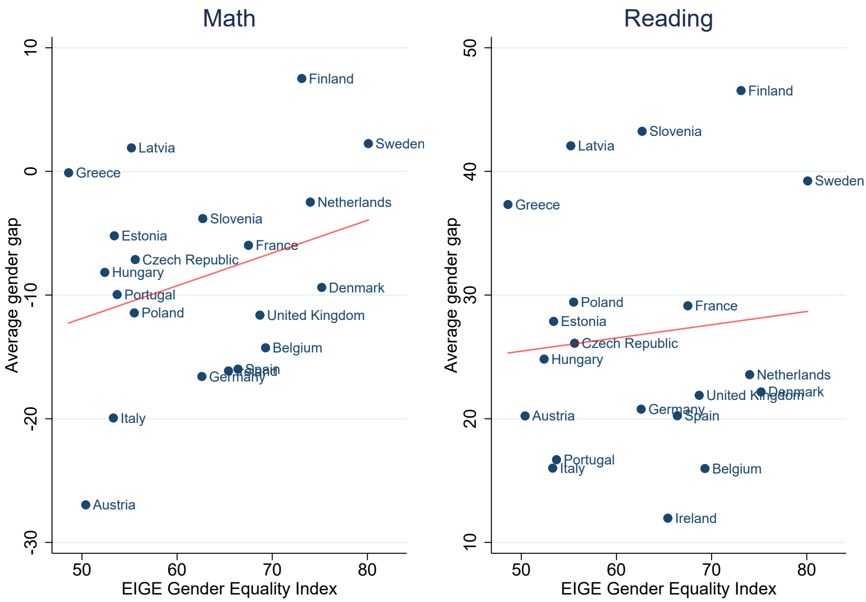
1.2 Gender norms and gender differences in education

Boys and girls differ in their educational achievement. A gender gap in education most often implies one that favours girls, as girls outperform boys along most educational dimensions (DiPrete and Buchmann, 2013). However, one exception is math. Girls systematically perform worse than boys on math tests, particularly at the top of the performance distribution (Bedard and Cho, 2010; Pope and Sydnor, 2010).

Although many factors may be contributing to these educational gender differences, one salient factor is gender norms. Geographical variation in the math gender gap indicates that this gap is not driven solely by innate gender differences in ability (Bedard and Cho, 2010). In addition, the gender gap in math does not exist at the point of school entry, but rather emerges over time when children are socialised into school (Fryer Jr and Levitt, 2010). Thus, social factors appear to be a significant determinant of the gender gaps in educational achievement. Indeed, the gender gap in math has been shown to correlate strongly with societal gender equality. Guiso et al. (2008) and Pope and Sydnor (2010) show that girls in more gender-equal countries, and more gender-equal U.S states, perform better relative to boys. As an example, Figure 2 demonstrates the relationship between the average gender gap in math and reading in the 2018 PISA evaluations, and the country level of gender equality (measured as country score on the EIGE Gender Equality Index) (PISA, 2018; EIGE, 2013).

[2]One exception is Martinez et al. (2012), who report on an experimental evaluation of a model preschool programme in the Gaza province of Mozambique. The authors find that two years after the start of the programme, children were more likely to be enrolled at primary school and had higher cognitive and socio-emotional skills.

Figure 2: Gender gap in test scores and gender equality



Notes: The figure plots the correlation between country-level gender equality and average country-wide gender gap in math and reading, respectively. The gender equality measure is the combined Gender Equality Index of 2013, obtained from the European Institute for Gender Equality (EIGE, 2013). The gender gaps in test scores are defined as girls' test scores - boys' test scores, for which the data are obtained from the PISA assessments of 2018 (PISA, 2018).

One channel through which norms could affect educational outcomes is the formation of identities. Norms shape our expectations regarding the social group with whom we identify, which in turn affect our beliefs of what we are capable of and our preferences for what we spend time on. The merging of economics and identity started with Akerlof and Kranton (2000), who develop a theoretical framework in which there exist several social categories, and each social category comes with a set of prescribed behaviours and ideal physical attributes. Individuals identify with some of the social categories, and derive utility from complying with the behaviour prescribed by the chosen categories. One salient category is gender, where everyone is assigned to either being a 'man' or a 'woman', and where there are prescribed attributes and behaviours that are considered 'manly' or 'womanly'. In this way, gender identity changes the pay-off of different actions, hence, the choice of identity may impact our economic outcomes, including our educational performance. Indeed, in the framework of Akerlof and Kranton (2000), as well as in empirical studies testing its implications, identity concerns have a significant impact on educational outcomes (Akerlof and Kranton, 2002; Schüller, 2015).

But from where do we perceive the norms that, through identity formation, shape our behaviour and impact our economic outcomes? The original answer of Akerlof and Kranton (2000) was to say that norms and category divisions arise from human interaction, and to point to the vast body of research outside of economics. However, in recent years, research in economics has begun to tackle this question (Kranton, 2016).

Several studies find that culture and historical traditions have a significant impact on both the attitudes and the economic outcomes of individuals today (see e.g. Fernandez and Fogli, 2009; Fernández, 2011; Alesina et al., 2013; Nollenberger et al., 2016; Finseraas and Kotsadam, 2017; Rodríguez-Planas and Nollenberger, 2018; Dahl et al., 2020), which demonstrate that cultural values are important determinants of the norms that shape our behaviour. However, another strand of the literature documents significant behavioural impacts of neighbourhood exposure and peer effects (see e.g. Chetty et al., 2016; Chetty and Hendren, 2018; Dahl et al., 2014; Olivetti et al., 2018), which indicate that our on-going exposure to institutions, peers, culture, and other environmental factors may be another important determinant of our behaviour.[3]

Taken together, a large body of literature shows that both family culture and neighbourhood characteristics affect our economic, including educational, outcomes. It is likely that these two channels do not operate independently of each other, however, there is limited empirical literature combining the two channels. The second chapter of this thesis remedies this gap, and investigates the interaction between cultural norms and neighbourhood exposure on the gender gap in math.

[3]The distinction between cultural values and neighbourhood exposure is similar to the framework developed by Bisin and Verdier (2011), who contrast vertical and horizontal transmission of norms. Vertically transmission of norms occurs within the family, from parents to children, and happens if parents believe that their children will benefit from certain cultural traits. Horizontal transmission denotes the socialisation of norms that takes place within a community context, where norms are transmitted by peers and surroundings. However, the culture/neighbourhood distinction noted above is broader compared to that of Bisin and Verdier (2011), as the cultural (i.e. 'vertical') channel includes also parents' peers and networks (sharing the same cultural beliefs), and the neighbourhood (i.e. 'horizontal') channel includes not only the effect of norms, but also the effect of more formal institutions.

1.3 Gender equality and domestic violence

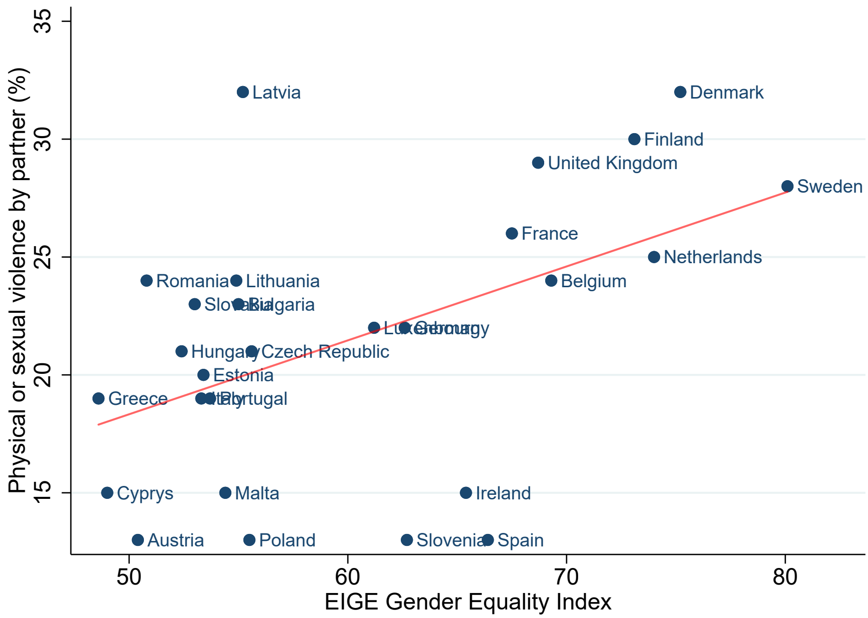
Domestic violence is a major issue for public health, productivity and gender equality. Globally, one in three women will experience violence from a partner at some point during their lifetime (García-Moreno et al., 2013), and the cost of total intimate partner violence is estimated to be 5.2% of world GDP (Hoeffler, 2017). In his 2020 speech to the Human Rights Council, the UN Secretary-General António Guterres argued that "violence against women and girls is the world's most pervasive human rights abuse", and called for action to improve gender equality worldwide (Guterres, 2020).

Gender equality and female empowerment is often cited as the main way to combat domestic violence. However, the theoretical predictions diverge. Models of household bargaining predict a negative relationship between female empowerment and domestic violence. The model predicts that a wife will stay in a marriage as long as her utility from the marriage surpasses the utility she would have in case of a marriage dissolution, i.e. the outside option of the marriage. The utility of the outside option becomes her threat point: the point at which is it no longer beneficial for her to stay in the marriage. Thus, as the outside option of the wife improves, this raises her threat point, which implies that she will be willing to endure less violence before it is beneficial for her to leave the marriage. Therefore, an improved economic position for the wife reduces violence, both indirectly, through more women leaving abusive spouses, and directly, through the deterrent effect of the threat of leaving (Farmer and Tiefenthaler, 1997; Aizer, 2010).

In contrast, the theory of 'male backlash', prominent in the sociological literature, predicts a positive relationship between women's relative economic position and domestic violence. An increase of women's financial independence, relative to their husbands, redefines the power structures of the relationship and violates traditional gender norms. This violation causes stress and anxiety, which could result in increases of violence as husbands try to reclaim authority over their wives (Hornung et al., 1981; Macmillan and Gartner, 1999).

Figure 3 shows a replication of the results from the Violence Against Women Survey, conducted by the Fundamental Rights Agency of the EU (henceforth FRA). The figure demonstrates the EU-wide relationship between prevalence of domestic violence and country-level gender equality (measured as country score on the EIGE Gender Equality Index). The relationship is in line with the predictions of male backlash: countries with higher levels of gender equality also experience higher levels of violence against women (FRA, 2014). The Nordic countries rank at the very top, with high levels of both gender equality and violence. This finding sparked a debate about 'The Nordic Paradox', as the gender-equal Nordic countries also appear to be unsafe places for women (Gracia and Merlo, 2016).

Figure 3: Domestic violence and gender equality



Notes: The figure plots the correlation between country-level gender equality and the fraction of women who have experienced physical or sexual violence by a partner (current or previous) since the age of 15. The gender equality measure is the combined Gender Equality Index of 2013, obtained from the European Institute for Gender Equality (EIGE, 2013). The data on violence prevalence is obtained from Table 2.1 in the Violence Against Women Report conducted by FRA (FRA, 2014).

Although the positive relationship in Figure 3 depicts mere correlations and should not be interpreted as causal, it does suggest the need for further research on how gender equality and female empowerment relate to domestic violence. But there are several empirical challenges associated with estimating this relationship. First, domestic violence is a sensitive issue that is most likely prone to selective underreporting (Ellsberg et al., 2001). Self-reported measures might not be representative of actual violence, but rather, the selection of who reports a violent incident. As the probability of reporting a violent incident is likely to increase with empowerment (Iyer et al., 2012), it is important to distinguish between changes in violence and changes in reporting behaviour.

Second, if we measure female empowerment using earnings, an important threat to identification is earnings endogeneity. Realised earnings likely reflect unobserved individual characteristics, which could be an outcome of, or correlate with, violence. Assortative matching will create selective marriages that are functions of earnings and of the underlying propensity for both perpetrating violence and for staying in a violent relationship (Pollak, 2004), which makes relative earnings a problematic measure of empowerment within households. Furthermore, the outside option of a marriage is not determined by a woman's realised earnings, but rather by the earnings potential she would face in case of a marriage dissolution (Aizer, 2010).

The conclusions from the existing literature on empowerment and domestic violence are mixed. Several well-identified empirical studies find support for the bargaining power hypothesis, i.e. that female empowerment reduces domestic violence (Stevenson and Wolfers, 2006; Aizer, 2010; Brassiolo, 2016; Anderberg et al., 2016; La Mattina, 2017). In contrast, several other studies find support for the male backlash theory, i.e. that female empowerment increases violence (Chin, 2012; Heath, 2014; Cools and Kotsadam, 2017; Guarnieri et al., 2018; Bhalotra et al., 2018). A strand of the literature connects domestic violence prevalence to the concept of gender identity, and finds that the effect of female economic empowerment depends on the gender norms of the husband and the relative status of the spouses (Atkinson et al., 2005; Tur-Prats, 2017; Alonso-Borrego and Carrasco, 2017; Svec and Andic, 2018). In addition, cross-country level evidence shows that domestic violence prevalence is higher in countries with more traditional gender norms (Heise and Kotsadam, 2015; González et al., 2018). Finally, a recent strand of the literature finds that domestic violence can be triggered by negative emotional cues or psychological stress (Card and Dahl, 2011; Cesur and Sabia, 2016; Beland and Brent, 2018).

The third chapter of this thesis is motivated by the puzzling results of the FRA survey. The chapter examines the causal relationship between female empowerment and domestic violence, while accounting for the empirical challenges noted above. While the existing literature on domestic violence has mainly relied on self-reported or aggregate measures of violence, the third chapter combines individual-level data with a non-self-reported measure of domestic violence.

2 Summary and contributions of the thesis

Next, I provide a short summary of each chapter of the thesis, briefly discussing their data, empirical strategies and results. Finally, I outline the contributions the chapters make to the existing literature.

2.1 Chapter I: Preschool Attendance, Schooling, and Cognitive Skills in East Africa

Co-authored with Jan Bietenbeck and Fredrick M. Wamalwa.
Published in *Economics of Education Review* (2019).

The first chapter of this thesis studies the effects of preschool attendance on children's school progression and cognitive skills in Kenya and Tanzania. Our empirical analysis draws on data from Uwezo, which conducts nationally representative household surveys of school-age children's education and their literacy and numeracy skills. The surveys also collect retrospective information on preschool attendance, which we can relate to current outcomes of respondents up to 16 years of age. The main part of our investigation focuses on impacts on the highest grade of school attended and a composite test score, which summarises a child's performance on the standardised literacy and numeracy assessments. The data contain information on these outcomes for more than half a million children across the two countries, independently of whether they are currently enrolled in school or not.

The main empirical challenge with identifying the impacts of preschool is that preschool attendance is not randomly assigned. For example, highly-educated parents may have a preference for sending their children to preschool and also foster their learning in other, unobserved ways. In this case, a simple regression of outcomes on preschool attendance would lead to a coefficient that is biased upward. We address this problem by comparing the outcomes of siblings who did and did not attend preschool. This way, we rely on within-household differences, thereby controlling for all determinants of outcomes and attendance that vary across families. We argue that the leftover variation between siblings is likely due to changes in the local availability of preschools, which came about because of an expansion of the pre-primary sector during our study period. In support of this claim, we show that even within households, children in later cohorts are much more likely to have attended preschool.

We find that the impact of preschool attendance on school progression follows an interesting dynamic pattern. In both Kenya and Tanzania, children often enrol in preschool late and only proceed to primary school once they have finished preschool. At early ages (7-9 years old), these children have therefore completed fewer school grades than their same-aged peers who did not attend preschool. However, we find that once enrolled in primary school, children who attended preschool progress through grades faster and are less likely to drop out. Eventually, they thus catch up with their peers and at ages 13-16 have accumulated the same number of grades of schooling in Kenya and about 0.1 more grades in Tanzania. In terms of cognitive skills, the estimates for the composite test score show that children who went to preschool outperform their peers already from the age of eight, and that these positive gains persist in the long run. Taken together, these findings show that there are important benefits from preschool attendance in both countries.

Our study contributes to the literature on the impacts of universal preschool programmes in developing countries. To the best of our knowledge, this study is the first large-scale assessment of preschool attendance for Sub-Saharan Africa, a region where preschool availability has increased rapidly in recent decades. Furthermore, in contrast to most existing literature, we are able to examine the longer-term effects of preschool attendance.

2.2 Chapter II: Cultural Gender Norms and Neighbourhood Exposure: Impacts on the Gender Gap in Math

The second chapter of this thesis investigates the interaction between cultural norms and neighbourhood characteristics, along with their impact on the gender gap in math. Specifically, I ask two research questions: first, is there an effect of cultural gender norms on the gender gap in math, and second, to what extent can this effect be mitigated by surrounding neighbourhood gender equality?

To answer the first research question, and to isolate the impact of cultural gender norms from that of more formal institutions, I estimate the effect of the gender norms in mothers' countries of origin on the gender gap in math among second-generation immigrants (who are all born in Sweden). Assuming that mothers transmit norms to their children, and that these norms differ systematically depending on the mother's source country, second-generation immigrants provide the ideal experiment to isolate the effect of cultural norms from the effect of formal institutions. To account for the fact that gender norms are not randomly assigned to parents, and therefore likely correlate with unobserved maternal characteristics, I follow Finseraas and Kotsadam (2017) and compare the gender gap in math only between opposite-sex siblings in a sibling fixed effects model. The sibling fixed effects control for everything that affects both siblings equally, including everything that correlates with source-country norms but that is unrelated to gender. By construction, the variation that remains is the *gender-specific* component of the cultural norms that affects opposite-sex siblings differently, i.e. gender norms.

To answer the second research question, I investigate the extent to which the cultural gender norm effect can be mitigated by municipality gender equality. To account for the possible bias caused by selection in where people choose to live, and to obtain exogenous variation in municipality characteristics, I exploit a refugee placement policy. Under this policy, government officials assigned asylum-seeking immigrants their initial location of residence. As these immigrants were not free to choose where they would be placed, their initial location of residence is independent of unobserved individual characteristics.

The study is based on high-quality Swedish administrative data on the universe of ninth-grade students who took the national standardised math test between 2004–2012. To proxy cultural gender norms and neighbourhood gender equality, I use female-over-male labour force participation rates, of both the immigrant mother's source country and of her assigned municipality of residence.

My results for the first research question show that mothers' cultural gender norms increase the sibling gender gap in math, such that girls with more gender-traditional mothers perform worse relative to their brothers. A one-standard-deviation increase in cultural gender norms (i.e. towards more traditional norms) increases the size of the math gender gap by 56%, in favour of boys. In addition, I find similar effects for final marks in other school subjects, which shows that the results are not driven by math-specific cultural norms, but rather by general gender stereotypes about girls and educational outcomes.

My results for the second research question show that municipality gender equality can almost completely mitigate the negative cultural gender norm effect. This result suggests that even though the sibling gender gap in math increases with mothers' adherence to traditional gender norms, this increase is smaller for siblings whose mothers were placed in more gender-equal municipalities. Taken together, my results show that while cultural gender norms play an important role for the gender gap in math, they are not immune to the influence of surrounding characteristics.

To the best of my knowledge, the second chapter of this thesis is the first study to estimate the interaction between neighbourhood characteristics and cultural norms. Thus, a novel and important contribution of my study is that it merges the literatures on cultural norms, neighbourhood exposure and educational outcomes. The chapter is also the first study to establish a causal link between cultural gender norms and the gender gap in math. By focusing on the gender gap between opposite-sex siblings, I am able to control for many potentially worrisome causes of variation that previous papers have not been able to control for, which allows me to more credibly isolate of the effect of cultural gender norms.

2.3 Chapter III: Backlash: Female Economic Empowerment and Domestic Violence

The third and final chapter estimates the effect of female economic empowerment on domestic violence, and accounts for the empirical challenges of selective underreporting and earnings endogeneity. As a proxy for female economic empowerment, I use a measure of women's potential earnings. I exploit the fact that women and men tend to sort into different industries and create a measure of prevailing local female earnings potential. This measure captures earnings variation that is not endogenous to domestic violence, and it provides a more accurate representation of the outside option of a marriage.

I measure domestic violence using hospital visits for assault, which I derive from third-party reported hospital records. The benefit of using hospital data is that my measure of violence suffers from very little selective reporting bias. Information on hospital visits for accidents allows me to examine possible misreporting of injury causes at the hospital, which I conclude does not pose a threat to my study. High-quality Swedish administrative data allows me to observe domestic violence and earnings on an individual level.

I show that the causal effect of increasing women's potential earnings, while keeping the earnings of their husbands constant, is positive. This result means that an exogenous increase in female economic empowerment causes an increase in the risk of domestic violence. The effect is substantial in magnitude, and does not depend on which spouse earns more than the other. Thus, my results are in line with the predictions of male backlash theory. As further support for the backlash mechanism, I show that increasing women's potential earnings, while keeping the earnings of their husbands constant, increases the risk of destructive behaviour by the husbands, such as visiting a hospital for reasons related to depression, anxiety, substance abuse and assault.

The richness of my data allows me to conduct a detailed heterogeneity analysis, where I show that the effect of increased potential earnings differs depending on the sub-group of the population. For the youngest women the effect of increased potential earnings is negative, but after the age of 40 the effect is consistently positive. Likewise, for the women with no more than high school education potential earnings reduce the risk of assault, but for women of higher education levels potential earnings increase the risk. The threat of leaving for a young woman may be more credible as it has yet to be tested. For the least educated women the outside option may be binding, such that increased potential earnings may significantly affect their ability to leave an abusive spouse. Thus, the heterogeneous results indicate that the women for whom a change in potential earnings actually affects the credibility of their threat of leaving, the results are in line with the predictions of bargaining power theory. But for the older women, and the women who may have the economic possibility to leave their spouse, but still do not, the effects are in line with male backlash theory. In line with this reasoning, I show that the backlash effect also increases with the duration of the marriage.

My contributions to the existing literature are threefold. First, my study is the first to use a (close to) objective measure of violence from individual level data when investigating the relationship between female empowerment and domestic violence. My study is also the first to investigate possible misreporting at the hospital, which allows me to conclude that my results are not suffering from reporting bias.

Second, I show that the effect of increased potential earnings on domestic violence differs sharply for different subgroups of the population. In this way, I show that both effects in line with the bargaining power hypothesis and effects in line with the theory of male backlash can co-exist, depending on the subgroup of the population and, speculatively, on how credible their threat of leaving an abusive spouse is.

Third, I investigate the mechanisms behind the positive effect of women's potential earnings on domestic violence. By estimating the effect of women's potential earnings on various measures of husbands' destructive behaviour, I show that the mechanisms are in line with a male backlash response to an improved relative economic position of the wife. This result is especially interesting to find in a gender-equal country like Sweden.

References

Aizer, A. (2010). The gender wage gap and domestic violence. *American Economic Review*, 100(4):1847–59.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Akerlof, G. A. and Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of economic literature*, 40(4):1167–1201.

Alesina, A., Giuliano, P., and Nunn, N. (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, 128(2):469–530.

Alonso-Borrego, C. and Carrasco, R. (2017). Employment and the risk of domestic violence: does the breadwinner's gender matter? *Applied Economics*, 49(50):5074–5091.

Anderberg, D., Rainer, H., Wadsworth, J., and Wilson, T. (2016). Unemployment and domestic violence: Theory and evidence. *The Economic Journal*, 126(597):1947–1979.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of economic perspectives*, 24(2):3–30.

Atkinson, M. P., Greenstein, T. N., and Lang, M. M. (2005). For women, breadwinning can be dangerous: Gendered resource theory and wife abuse. *Journal of Marriage and Family*, 67(5):1137–1148.

Bedard, K. and Cho, I. (2010). Early gender test score gaps across oecd countries. *Economics of Education Review*, 29(3):348–363.

Behrman, J. R., Cheng, Y., and Todd, P. E. (2004). Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach. *Review of Economics and Statistics*, 86(1):108–132.

Beland, L.-P. and Brent, D. A. (2018). Traffic and crime. *Journal of Public Economics*, 160:96–116.

Berlinski, S., Galiani, S., and Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, 93(1-2):219–234.

Berlinski, S., Galiani, S., and Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics*, 92(5-6):1416–1440.

Bhalotra, S. R., Kambhampati, U. S., Rawlings, S., and Siddique, Z. (2018). Intimate partner violence and the business cycle. Working Paper 11274.

Bietenbeck, J., Piopiunik, M., and Wiederhold, S. (2018). Africa's Skill Tragedy: Does Teachers' Lack of Knowledge Lead to Low Student Performance? *Journal of Human Resources*, 53(3):553–578.

Bisin, A. and Verdier, T. (2011). The economics of cultural transmission and socialization. In *Handbook of social economics*, volume 1, pages 339–416. Elsevier.

Bouguen, A., Filmer, D., Macours, K., and Naudeau, S. (2018). Preschool and Parental Response in a Second Best World: Evidence from a School Construction Experiment. *Journal of Human Resources*, 53(2):474–512.

Brassiolo, P. (2016). Domestic violence and divorce law: When divorce threats become credible. *Journal of Labor Economics*, 34(2):443–477.

Brinkman, S. A., Hasan, A., Jung, H., Kinnell, A., and Pradhan, M. (2017). The Impact of Expanding Access to Early Childhood Education Services in Rural Indonesia. *Journal of Labor Economics*, 35(S1):S305–S335.

Card, D. and Dahl, G. B. (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics*, 126(1):103–143.

Cesur, R. and Sabia, J. J. (2016). When war comes home: The effect of combat service on domestic violence. *Review of Economics and Statistics*, 98(2):209–225.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.

Chin, Y.-M. (2012). Male backlash, bargaining, or exposure reduction?: women's working status and physical spousal violence in india. *Journal of population Economics*, 25(1):175–200.

Cools, S. and Kotsadam, A. (2017). Resources and intimate partner violence in sub-saharan africa. *World Development*, 95:211–230.

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

Cunha, F., Heckman, J. J., Lochner, L., and Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. *Handbook of the Economics of Education*, 1:697–812.

Cunha, F., Heckman, J. J., and Schennach, S. M. (2010). Estimating the technology of cognitive and noncognitive skill formation. *Econometrica*, 78(3):883–931.

Currie, J. (2001). Early childhood education programs. *Journal of Economic perspectives*, 15(2):213–238.

Currie, J. and Thomas, D. (1995). Does Head Start Make a Difference? *American Economic Review*, 85(3):341–64.

Dahl, G. B., Felfe, C., Frijters, P., and Rainer, H. (2020). Caught between cultures: Unintended consequences of improving opportunity for immigrant girls. Working Paper 26674, National Bureau of Economic Research.

Dahl, G. B., Løken, K. V., and Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, 104(7):2049–74.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1(3):111–134.

Dietrichson, J., Kristiansen, I. L., and Nielsen, B. C. (2018). Universal preschool programs and long-term child outcomes: A systematic review. Working Paper 2018:19, IFAU - Institute for Evaluation of Labour Market and Education Policy.

DiPrete, T. A. and Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. Russell Sage Foundation.

European Institute for Gender Equality (2013). Gender equality index. Data retrieved from The EIGE website, `https://eige.europa.eu/gender-equality-index/2013/compare-countries/index/table` Accessed: 2020-03-06.

Ellsberg, M., Heise, L., Pena, R., Agurto, S., and Winkvist, A. (2001). Researching domestic violence against women: methodological and ethical considerations. *Studies in family planning*, 32(1):1–16.

Farmer, A. and Tiefenthaler, J. (1997). An economic analysis of domestic violence. *Review of social Economy*, 55(3):337–358.

Fernández, R. (2011). Does culture matter? In *Handbook of social economics*, volume 1, pages 481–510. Elsevier.

Fernandez, R. and Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American economic journal: Macroeconomics*, 1(1):146–77.

Finseraas, H. and Kotsadam, A. (2017). Ancestry culture and female employment—an analysis using second-generation siblings. *European sociological review*, 33(3):382–392.

Fundamental Rights Agency of the European Union (2014). Violence against women: An eu-wide survey. main results report. Available at: http://fra.europa. eu/en/publication/2014/violence-against-women-eu-widesurvey-main-results-report.

Fryer Jr, R. G. and Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2):210–40.

Garces, E., Thomas, D., and Currie, J. (2002). Longer-Term Effects of Head Start. *American Economic Review*, 92(4):999–1012.

García-Moreno, C., Pallitto, C., Devries, K., Stöckl, H., Watts, C., and Abrahams, N. (2013). *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence*. World Health Organization.

González, L., Rodríguez-Planas, N., et al. (2018). Gender norms and intimate partner violence. Working Paper 1620, Economics Working Paper Series.

Gracia, E. and Merlo, J. (2016). Intimate partner violence against women and the nordic paradox. *Social Science & Medicine*, 157:27–30.

Guarnieri, E., Rainer, H., et al. (2018). Female empowerment and male backlash. Working Paper 7009.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, gender, and math. *Science*, 320(5880):1164–1165.

Guryan, J., Hurst, E., and Kearney, M. (2008). Parental education and parental time with children. *Journal of Economic perspectives*, 22(3):23–46.

Guterres, A. (2020). The highest aspiration: A call to action for human rights. Available at: `https://www.un.org/sg/sites/www.un.org.sg/files/atoms/files/The_Highest_Asperation_A_Call_To_Action_For_Human_Right_English.pdf` Accessed: 2020-03-02.

Heath, R. (2014). Women's access to labor market opportunities, control of household resources, and domestic violence: Evidence from bangladesh. *World Development*, 57:32–46.

Heise, L. L. and Kotsadam, A. (2015). Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys. *The Lancet Global Health*, 3(6):e332–e340.

Hoeffler, A. (2017). What are the costs of violence? *Politics, Philosophy & Economics*, 16(4):422–445.

Hornung, C. A., McCullough, B. C., and Sugimoto, T. (1981). Status relationships in marriage: Risk factors in spouse abuse. *Journal of Marriage and the Family*, pages 675–692.

Iyer, L., Mani, A., Mishra, P., and Topalova, P. (2012). The power of political voice: women's political representation and crime in india. *American Economic Journal: Applied Economics*, 4(4):165–93.

Krafft, C. (2015). Increasing educational attainment in Egypt: The impact of early childhood care and education. *Economics of Education Review*, 46:127–143.

Kranton, R. E. (2016). Identity economics 2016: Where do social distinctions and norms come from? *American Economic Review*, 106(5):405–09.

La Mattina, G. (2017). Civil conflict, domestic violence and intra-household bargaining in post-genocide rwanda. *Journal of Development Economics*, 124:168–198.

Macmillan, R. and Gartner, R. (1999). When she brings home the bacon: Labor-force participation and the risk of spousal violence against women. *Journal of Marriage and the Family*, pages 947–958.

Martinez, S., Naudeau, S., and Pereira, V. (2012). The promise of preschool in africa: A randomized impact evaluation of early childhood development in rural mozambique.

Nollenberger, N., Rodríguez-Planas, N., and Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, 106(5):257–61.

Nores, M. and Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) Investing in the very young. *Economics of Education Review*, 29(2):271–282.

Olivetti, C., Patacchini, E., and Zenou, Y. (2018). Mothers, peers, and gender-role identity. *Journal of the European Economic Association*.

Program for International Student Assessment (2018). Oecd pisa scores in reading and mathematics, by gender. Data retrieved from The PISA website, `https://pisadataexplorer.oecd.org/ide/idepisa/dataset.aspx` Accessed: 2020-03-06.

Pollak, R. A. (2004). An intergenerational model of domestic violence. *Journal of Population Economics*, 17(2):311–329.

Pope, D. G. and Sydnor, J. R. (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives*, 24(2):95–108.

Rao, N., Sun, J., Wong, J. M. S., Weekes, B., Ip, P., Shaeffer, S., Young, M., Bray, M., Chen, E., and Lee, D. (2014). Early childhood development and cognitive development in developing countries: a rigorous literature review. London: Department for International Development.

Rodríguez-Planas, N. and Nollenberger, N. (2018). Let the girls learn! it is not only about math… it's about gender social norms. *Economics of Education Review*, 62:230–253.

Schüller, S. (2015). Parental ethnic identity and educational attainment of second-generation immigrants. *Journal of Population Economics*, 28(4):965–1004.

Stevenson, B. and Wolfers, J. (2006). Bargaining in the shadow of the law: Divorce laws and family distress. *The Quarterly Journal of Economics*, 121(1):267–288.

Svec, J. and Andic, T. (2018). Cooperative decision-making and intimate partner violence in peru. *Population and development review*, 44(1):63–85.

Tur-Prats, A. (2017). Unemployment and intimate-partner violence: A gender-identity approach. Working Paper 963, Barcelona Graduate School of Economics.

UNESCO (2012). *Global education digest 2012. Opportunities lost: The impact of grade repetition and early school leaving*. Paris: UNESCO.

Uwezo (2015). *Are Our Children Learning? Literacy and Numeracy Across East Africa 2014*. Twaweza East Africa, Nairobi.

World Bank (2018). *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

The World Bank (2018a). School enrollment, pre-primary (%gross). Data retrieved from World Development Indicators, `https://data.worldbank.org/indicator/SE.PRE.ENRR?locations=ZG` Accessed: 2020-03-02.

The World Bank (2018b). School enrollment, primary (%gross). Data retrieved from World Development Indicators, `https://data.worldbank.org/indicator/SE.PRM.ENRR?locations=ZG` Accessed: 2020-03-02.

Chapter 1

Preschool Attendance, Schooling, and Cognitive Skills in East Africa

Co-authored with Jan Bietenbeck[1] and Fredrick M. Wamalwa.[2]
Published in *Economics of Education Review* (2019).

Abstract

We study the effects of preschool attendance on children's schooling and cognitive skills in Kenya and Tanzania. We use a within-household estimator and data from nationally representative surveys of school-age children's literacy and numeracy skills, which include retrospective information on preschool attendance. In both countries, school entry rules are not strictly enforced, and children who attend preschool often start primary school late. At ages 7–9, these children have thus attended fewer school grades than their same-aged peers without pre-primary education. However, they catch up over time: at ages 13–16, children who went to preschool have attended about the same number of school grades and score about 0.10 standard deviations higher on standardized tests in both countries. They are also 3 (5) percentage points more likely to achieve basic literacy and numeracy in Kenya (Tanzania).

Keywords: preschool, education, cognitive skills, Sub-Saharan Africa
JEL Classifications: I21, J24

[1]Department of Economics, Lund University and IZA.
[2]School of Economics, Faculty of Commerce, University of Cape Town.

1 Introduction

School enrollment in Sub-Saharan Africa has increased substantially over the past two decades. However, many students fall behind the curriculum early on, and grade repetition and early dropout are widespread (UNESCO, 2012). Students also learn remarkably little in school: for example, only one in five third-grade students in East Africa has second-grade literacy and numeracy skills, and less than one third of sixth-grade students in Southern and Eastern Africa can solve a simple subtraction problem (Uwezo, 2015; Bietenbeck et al., 2018).

One possible reason why students in these countries perform so poorly is that they enter school unprepared. Specifically, to the extent that early and later learning are complementary, a lack of education before starting school reduces children's efficiency in learning once they arrive there (Cunha and Heckman, 2007). High-quality preschool programs which prepare children for school are therefore often seen as a promising way to enhance learning outcomes (e.g. World Bank, 2018). In Sub-Saharan Africa, pre-primary education has been expanding rapidly, with the gross enrollment ratio doubling from 15% to 32% between 2000 and 2017.[1] Whether preschools are actually effective at boosting student outcomes in this region is unclear, however, because rigorous empirical evidence is still scarce.

In this paper, we study the effects of preschool attendance on children's schooling and cognitive skills in Kenya and Tanzania. Our empirical analysis draws on data from Uwezo, which conducts nationally representative household surveys of school-age children's education and their literacy and numeracy skills. The surveys also collect retrospective information on preschool attendance, which we can relate to current outcomes of respondents up to 16 years of age. The main part of our investigation focuses on impacts on the highest grade of school attended and a composite test score, which summarizes a child's performance on the standardized literacy and numeracy assessments. The data contain information on these outcomes for more than half a million children across the two countries, independently of whether they are currently enrolled in school or not.

Our regression framework compares the outcomes of children who did and did not attend preschool. For identification, we rely on within-household differences, thereby controlling for all determinants of outcomes and attendance that vary across families.[2] We argue that the leftover variation between siblings is likely due to

[1]These figures were retrieved from the World Development Indicators (https://data.worldbank.org/indicator/SE.PRE.ENRR) on June 19, 2019.

[2]Similar strategies have been used by Currie and Thomas (1995), Garces et al. (2002), and Deming (2009) to estimate the impacts of the Head Start program in the United States, and by Berlinski

changes in the local availability of preschools, which came about because of an expansion of the pre-primary sector during our study period. In support of this claim, we show that even within households, children in later cohorts are much more likely to have attended preschool. To mitigate concerns about endogenous selection from the start, the regressions also control for a variety of predetermined characteristics that still vary between siblings.

The impact of preschool attendance on schooling follows an interesting dynamic pattern. In both Kenya and Tanzania, school entry rules are not strictly enforced, and for reasons discussed below, children who attend preschool often start primary school late. At ages 7–9, these children thus have attended fewer school grades than their same-aged peers without pre-primary education. However, once enrolled in primary school, children who attended preschool progress through grades faster and are less likely to drop out. Eventually, in both countries, they thus catch up and at ages 13–16 have attended about the same number of school grades as their peers.

In terms of cognitive skills, the estimates for the composite test score show that children who went to preschool outperform their peers in the long run. In Kenya, this effect fades in at early ages and soon stabilizes at a gain of about 0.10 standard deviations (SD). In Tanzania, in contrast, children who attended preschool outperform their peers from the age of school start, with long-term gains of the same order of magnitude as in Kenya. Separate regressions moreover reveal that among 13– to 16-year-olds, preschool attendance raises the likelihood of mastering basic, second-grade literacy and numeracy by 3 percentage points in Kenya and by 5 percentage points in Tanzania. Taken together, these findings show that there are important benefits from preschool attendance in both countries.

To ensure that these results are not driven by selection, we perform a variety of falsification tests and robustness checks. For example, we show that children who attended preschool do not differentially benefit from other educational inputs such as private tutoring, suggesting that our findings are not due to child-specific investments related to unobserved characteristics. Applying the method developed by Oster (2017), we also judge the importance of omitted variable bias more generally by observing the sensitivity of our regression results to the addition of controls. From this approach, selection on unobserved factors would need to be at least four times as large as selection on observed factors to explain away the long-term impacts on the composite test score.

et al. (2008) to estimate the effects of preschool in Uruguay. Throughout the paper, we use the terms "within-household differences" and "between-sibling differences" interchangeably.

Our paper contributes to a growing literature on the impacts of preschool education on children's outcomes in developing countries, which has focused mostly on Asia and Latin America and which is reviewed in detail in Nores and Barnett (2010) and in Rao et al. (2014). Using retrospective data on preschool enrollment and the same within-household estimator as we do, Berlinski et al. (2008) find that Uruguayan children who attended preschool accumulate 0.8 more years of education by age 15. Applying the same strategy to Egyptian data, Krafft (2015) finds that preschool attendance leads to an additional 0.4 years of schooling among 18-29 year-olds. Behrman et al. (2004), Berlinski et al. (2009), and Brinkman et al. (2017) similarly document positive short-term effects of preschool attendance on children's cognitive skills in Bolivia, Argentina, and Indonesia, respectively. In contrast, a randomized evaluation of a preschool construction program in Cambodia found negative short-term impacts on test scores of targeted children, a result that is partly explained by a shift from underage enrollment in primary school to enrollment in preschool (Bouguen et al., 2018).[3]

To the best of our knowledge, the only other rigorous study of preschool effects in Sub-Saharan Africa is the paper by Martinez et al. (2012), who report on an experimental evaluation of a model preschool program in the Gaza province of Mozambique. The authors find that two years after the start of the program, children were more likely to be enrolled in primary school and had higher cognitive and socio-emotional skills. In contrast to this small-scale evaluation, our study uses nationally representative data on preschool attendance and learning outcomes from two countries. Moreover, unlike most of the previous literature, we are able to examine the longer-term effects of preschool attendance.

2 Institutional background

2.1 Education and preschools in Kenya

Basic education in Kenya consists of three years of preschool, eight years of primary school, and four years of secondary school.[4] Preschool comprises three distinct

[3]Zuilkowski et al. (2012), Hazarika and Viren (2013), Cortázar (2015), and Aboud et al. (2016) also evaluate the impacts of preschool education on children's learning outcomes in developing countries. Unlike our analysis, these studies do not control for selection into preschool based on unobserved factors.

[4]The following description of the preschool and school systems in Kenya draws on Tooley et al. (2008), Bidwell et al. (2013), Heyneman and Stern (2014), Edwards Jr. et al. (2015), and Ngware et al. (2016).

grades – baby class (ages 3–4), nursery (ages 4–5), and pre-unit (ages 5–6) –, even though in practice children of different ages are often taught together in the same classroom. Attendance is not compulsory, and children who do not go to preschool typically stay home to help with household chores instead (Daniel, 2012). In the year after they turn six, all children are supposed to enter primary school, although this rule is not strictly enforced and in practice many children enter primary school late. Primary school has been free of charge since fees were abolished in 2003. At the end of primary school, students take a national leaving exam, which largely determines which secondary schools they can enter.

There are two broad types of preschools in Kenya. First, public preschools are run by the government and are usually attached to a primary school. In rural areas, these preschools are often the only available option. Second, private preschools are owned and run by a variety of providers, including non-governmental and faith-based organizations, community-based associations, and private-for-profit agents. They comprise a diverse range of institutions, including: highly unregulated and unregistered non-formal preschools, which are mainly located in informal urban settlements; formal private preschool academies in middle and high-income areas; and a small number of exclusive private preschools catering to very high-income households. All preschools (both public and private) charge tuition fees, which vary widely depending on the type of institution attended, with non-formal private preschools charging the lowest fees.

Unlike in many European countries and the United States, where pre-primary education for younger children typically focuses on play, preschool studies in Kenya are highly academic: students sit at desks and listen to the teacher teach in a classroom-like setting. Curricula, while not standardized, tend to emphasize the learning of basic numeracy and literacy skills via memorization and recitation. In contrast, only little attention is paid to the development of socio-emotional skills (Ngware et al., 2016). As children often spend more than 35 hours per week in preschool, most institutions offer a feeding program, which is financed via the tuition fees or via special meal fees (Bidwell et al., 2013).

Mirroring the diversity in institutional arrangements, the quality of preschools as perceived by parents varies widely (Bidwell et al., 2013), and is considered low on average by international standards (Ngware et al., 2016). Available metrics of preschool quality show no consistent trends during our study period: on the one hand, the average student-teacher ratio in Kenya rose from 23 to 26 between 2004 and 2014 (in comparison, the ratio in OECD countries stood at 15 in 2014). On the other hand, the share of pre-primary teachers with at least some formal training rose from 70% to 82% during those years. Moreover, the 2000s were characterized by a decline in the quality of primary education, as indicated by an increasing student-teacher ratio, a decreasing share of trained teachers, and decreased government spending per student.[5]

We end our discussion of the institutional context in Kenya by presenting some stylized facts on enrollment.[6] First, pre-primary education has expanded substantially over the last few decades, with 83% of the 2004 cohort attending preschool. Observers ascribe this increased enrollment both to a series of government efforts to increase preschool availability and quality (e.g. by training preschool teachers), and to increased demand due to more mothers entering the workforce (Nganga, 2009; Ngware et al., 2016). Second, late enrollment in primary school is common: for example, in 2013, 14% of 7-year-olds were not yet enrolled in primary school, with the vast majority still attending preschool. A key reason for this is that enrollment in pre-primary education is often late itself, and parents prefer their children to complete preschool before proceeding to primary school (Bidwell et al., 2013). Third, while school enrollment is high compared to other countries in Sub-Saharan Africa, some dropout occurs, with 5% of 13-year-olds reporting not to be enrolled in school in 2013.

2.2 Education and preschools in Tanzania

Basic education in Tanzania consists of two years of preschool, seven years of primary school, four years of lower ('ordinary') secondary school, and two years of upper ('advanced') secondary school.[7] Children can enter non-compulsory pre-

---

[5] These figures were retrieved from the World Development Indicators (https://data.worldbank.org) on June 27, 2019. No data on the share of pre-primary teachers with formal training are available for OECD countries. No separate data on spending on pre-primary education are available. In primary school, the student-teacher ratio increased from 40 in 2004 to 57 in 2012, the share of trained teachers decreased from 99% in 2003 to 97% in 2009, and government expenditure per primary student decreased from 21% in 2001 to 10% in 2012 (years chosen for data availability). In the 2010s, the student-teacher ratio started decreasing and spending started increasing again, but as will become clear below, the children in our sample were in primary school mostly during the 2000s.

[6] All enrollment statistics in this section are based on data from the nationally representative Uwezo surveys, which we describe in detail in the next section. For the preschool figures, we focus on cohorts who were at least 10 years old in our data in order to account for the frequent late enrollment. For dropout rates, we focus on 13-year-olds, who have typically not completed primary school yet in either country. When reporting statistics for individual years, we choose 2013 as this is the latest year in which nation-wide Uwezo surveys were conducted in both Kenya and Tanzania.

[7] The following description of the preschool and school systems in Tanzania draws on Kweka et al. (1997) and Mtahabwa and Rao (2010).

primary education at age 5 and are supposed to start primary school in the year after they turn 7, but this rule is not strictly enforced in practice. At the end of grade 7, they take a school leaving exam which regulates access to public secondary schools. Tuition fees for primary education were abolished in 2002, but secondary schools still levied fees during our study period.

Preschools in Tanzania are predominantly public: in 2016, 95% of preschool students were enrolled in a government-run institution (President's Office of the United Republic of Tanzania, 2016). These are often attached to a primary school and charge varying tuition fees. Unlike in Kenya, the Tanzanian government has adopted an official preschool curriculum, which emphasizes the development of both cognitive and socio-emotional skills. In practice, however, preschool teachers often have little or no knowledge of the official curriculum and tend to focus on formal instruction in basic literacy and numeracy (Mligo, 2016).

The quality of education in Tanzanian preschools is usually described as being low (e.g. Kweka et al., 1997; Mtahabwa and Rao, 2010). Student-teacher ratios are high and have increased from 54 to 77 between 2004 and 2014. This reflects a shortage of teachers, due to which pre-primary students are sometimes taught together with older, primary school students in multi-grade classrooms (Mghasse and William, 2016). Countering this trend, the share of preschool teachers with formal training increased from 18% to 36% during the same period. Finally, unlike in Kenya, primary school quality as measured by the student-teacher ratio, the share of trained teachers, and government spending per student appears to have stayed roughly constant during the 2000s and early 2010s.[8]

Like for Kenya, we now present some stylized facts on enrollment. First, there has been an expansion of pre-primary education in Tanzania, with the attendance rate rising from 61% for the 1995 cohort to 69% for the 2004 cohort. This rise is spurred by the Tanzanian government, whose expansion strategy has been to attach pre-primary classrooms to existing primary schools (Kweka et al., 1997; Mtahabwa and Rao, 2010).[9] Second, children frequently enroll in primary school late, and some never enroll at all: in 2013, for example, 9% of 8-year-olds reported to be enrolled neither in school nor in preschool. Third, dropout during primary school

[8] See footnote 5 for details on the sources of these figures.

[9] In particular, the Tanzanian government's 1995 Education and Training Policy mandated primary schools to establish pre-primary classes in partnership with local communities, with the Ministry of Education and Culture developing the pre-primary curriculum and facilitating teacher training (Mtahabwa and Rao, 2010). While local observers seem to agree that this policy has boosted the expansion of the pre-primary sector, to our knowledge no data exist on the actual number of classrooms formed or teachers trained.

is common, with 12% of 13-year-olds not enrolled in school in 2013.

3 Data

3.1 The Uwezo surveys

The Uwezo initiative has been conducting large-scale assessments of school-age children's literacy and numeracy skills in Kenya, Tanzania, and Uganda since 2009, with more than 1.3 million children tested until 2014. The assessments are administered as part of repeated cross-sectional household surveys, which are representative at the district level. An important advantage of this design is that skills are measured also for children who are currently not enrolled in school. The surveys collect information from children aged 6-16 (7-16 in Tanzania) on their current enrollment and highest grade attended as well as on a variety of child and household characteristics. Crucially for our purposes, in recent waves respondents were also asked whether they ever attended preschool.

The literacy and numeracy assessments measure core competencies that children should have learned after two years of schooling according to the national curriculum. Literacy tests given in both English and Swahili assess the following four competencies in order of rising difficulty: (1) recognition of letters, (2) recognition of words, (3) reading a paragraph, and (4) reading a short story. Numeracy tests measure the following six competencies in order of rising difficulty: (1) counting (the number of objects on a show card), (2) recognition of numbers, (3) rank ordering of numbers, (4) addition, (5) subtraction, and (6) multiplication. A student's score on each test equals the highest competency level achieved, with a zero indicating that she did not even master the simplest skill assessed. Previous analyses of Uwezo data have shown that even many higher-grade students do not master these second-grade competencies (Jones et al., 2014; Uwezo, 2015).

3.2 Variable definitions

The key explanatory variable in our regressions is an indicator for whether or not a child has attended preschool.[10] In the 2013 and 2014 waves of the Uwezo surveys, we additionally observe for how many years each child attended preschool, and we

[10]In the 2011 and 2012 waves, the indicator measures attendance of preschool or nursery, which includes less education-focused child care institutions. Our results are robust to excluding these waves from the sample. See Online Appendix A for an overview of the exact questions on preschool attendance asked in each country and wave.

use this information below to estimate effects at the intensive margin. Note that the data on preschool attendance are based on retrospectively reported information. This has the major advantage that we can estimate longer-term impacts by relating past enrollment to current outcomes. However, it comes with the drawback that we do not observe any information on the nature or quality of the preschool attended, which limits the analysis of potential mechanisms underlying our findings. Another concern with retrospectively recorded information is the possibility of recall error; specifically, if such recall error systematically depends on preschool attendance, this could bias our estimates (Garces et al., 2002). We therefore show in a robustness check that systematic recall error is unlikely to drive our results.

Our main analysis focuses on two outcomes. First, we study the highest school grade attended. We observe this variable both for children who are currently enrolled in school and for those who dropped out, with children who are still in preschool coded as having zero grades attended.[11] Because all our regressions include age dummies, this outcome is best interpreted as a measure of school progression. Second, we construct a composite test score as follows: we first standardize the English, Swahili, and numeracy scores by country, Uwezo survey wave, and age to have mean zero and standard deviation one; we then average these scores for each student and standardize the resulting composite test score again. Finally, in auxiliary analysis, we examine the effects of preschool attendance on indicators for current school enrollment and the possession of second-grade skills, as defined by achieving the highest competency level in the numeracy test and at least one of the two literacy tests.

The control variables include a variety of socio-demographic characteristics, such as age and gender, mother's education, and an index of current household wealth. Moreover, we construct two measures of early-life economic conditions at the district level from external data sources. The first measure is the log of average night light density, which is a proxy for economic activity (Henderson et al., 2012). The second measure consists of two separate dummies for positive and negative rainfall shocks, defined as rainfall above the 80th percentile and below the 20th percentile of the long-term district mean. Rainfall shocks have been used widely as a measure of income shocks in rural economies; see Shah and Steinberg (2017) for a recent example. We allow for differential impacts of economic conditions at dif-

[11]A previous version of this paper (Bietenbeck et al., 2017) reported results for the highest grade completed rather than the highest grade attended. The main advantage of focusing on the highest grade attended is that it allows us to distinguish between first graders and children who are still in preschool, both of which have zero grades completed. Results for both outcomes are qualitatively similar.

ferent ages by computing our two measures separately at each age between zero and the official school entry age for each child. We provide many more details on the construction of these and all other variables used in the empirical analysis in Online Appendix A.

3.3 Sample selection and descriptive statistics

We use data from all available waves of the Uwezo surveys with information on preschool attendance. These are the 2013 and 2014 waves in Kenya and the four waves conducted between 2011 and 2014 in Tanzania. In Uganda, the only nationally representative Uwezo survey which asked about preschool attendance was conducted in 2013. Unfortunately, this key information is missing for 49% of children in the data for this wave, which led us to exclude Uganda from the analysis (preschool attendance is observed for all children in Kenya and Tanzania). We restrict our attention to children aged 7 and above in Kenya and 8 and above in Tanzania because some younger children were still of preschool age at the time of the survey. In order to ensure that we focus on comparable siblings in our within-household analysis, we also drop any children who report never to have enrolled in preschool or school. After these restrictions, our final sample includes 517,096 children across both countries, of whom 38,685 have a sibling with different preschool status.

Table 1 reports summary statistics for key variables separately for each country. Almost a fifth of children have mothers without any formal education, and more than two thirds live in rural areas. In Kenya, 85% of children attended preschool for an average length of 2.1 years, with the corresponding figures for Tanzania being 62% and 1.3 years. The vast majority are currently enrolled in education, a statistic that is partly due to our focus on children who ever enrolled in preschool or school. On average, they have attended about four and half grades, but only 58% in Kenya and 43% in Tanzania possess second-grade skills. Finally, Appendix Table 1 presents enrollment statistics and outcome means separately by age. As shown there, children tend to be behind grade for age, with a non-negligible share of the younger school-age children still attending preschool. This finding will be important for the interpretation of our results below.

Table 1: Summary statistics

	Kenya	Tanzania
Socio-demographic characteristics		
Age	11.08	11.72
	(2.77)	(2.46)
Female	0.49	0.50
	(0.50)	(0.50)
Mother's education:		
None	0.17	0.19
	(0.37)	(0.39)
Some primary or more	0.83	0.81
	(0.37)	(0.39)
No. of children in household	3.09	2.47
	(1.55)	(1.26)
Current household wealth (index)	0.00	0.00
	(1.00)	(1.00)
Rural location	0.67	0.78
	(0.47)	(0.41)
Early-life economic conditions		
No. of negative rainfall shocks	1.44	1.87
	(0.85)	(0.87)
No. of positive rainfall shocks	1.10	1.27
	(0.79)	(0.72)
Log night light density	-1.44	-2.39
	(2.32)	(2.29)
Preschool attendance		
Attended preschool	0.85	0.62
	(0.36)	(0.48)
Years of preschool |attended	2.11	1.32
	(1.08)	(0.75)
Outcomes		
Highest grade attended	4.62	4.42
	(2.53)	(2.28)
Currently enrolled	0.99	0.94
	(0.10)	(0.24)
Composite test score	0.00	0.00
	(1.00)	(1.00)
Has 2nd-grade lit./num. skills	0.58	0.43
	(0.49)	(0.49)
Observations (children):		
Total	223,339	293,757
With within-household variation	7,532	31,153

Notes: The table reports means and standard deviations (in parentheses) of key variables separately for children in Kenya and Tanzania. In regressions, early-life economic conditions are proxied by district-level indicators for negative and positive rainfall shocks and district-level log night lights at each age between 0 and 5 in Kenya (0 and 6 in Tanzania); for conciseness, this table shows totals across all of these ages. Years of preschool are observed only in the 2013 and 2014 waves of the Uwezo survey (N=223,339 in Kenya and N=111,043 in Tanzania). Currently enrolled is an indicator for being currently enrolled in either preschool or school. Has 2nd-grad lit./num. skills is an indicator for achieving the highest competency level in the numeracy test and at least one of the two literacy tests. The final row reports the fraction of children living in households in which at least one child went to preschool and at least one child did not.

4 Empirical strategy

The main challenge in identifying the causal effects of preschool attendance on later outcomes is that selection into pre-primary education is likely non-random. For example, more educated parents may have a stronger preference for preschool education while also fostering their children's learning in other ways. In this case, any regression that does not control for this selection would yield estimates that are biased upward. To address this challenge, we follow a strand of previous literature (Currie and Thomas, 1995; Garces et al., 2002; Berlinski et al., 2008; Deming, 2009) and estimate models with household fixed effects, thus holding constant all determinants of preschool attendance and outcomes that do not vary between siblings. Our main OLS specification reads:

$$\begin{aligned} Y_{ij} = &\alpha + \beta_1 PRE_{ij} \times Age\ Group_{ij}^{7-9} + \beta_2 PRE_{ij} \times Age\ Group_{ij}^{10-12} \\ &+ \beta_3 PRE_{ij} \times Age\ Group_{ij}^{13-16} + \mathbf{AGE}'_{ij}\gamma + \mathbf{X}'_{ij}\theta + \eta_j + \varepsilon_{ij}. \end{aligned} \tag{1}$$

Here, i denotes individuals and j denotes households, Y_{ij} is the highest grade attended or the composite test score, PRE_{ij} is the indicator for preschool attendance, and $\mathbf{AGE}_{ij}$ is a vector of individual age dummies. $\mathbf{X}_{ij}$ is a vector of controls that includes dummies for birth order and gender and their interactions, dummies for cohort and their interactions with the individual age dummies, and the proxies for early-life economic conditions described above.[12] We allow for dynamic impacts of preschool attendance by interacting PRE_{ij} with three age group indicators: $Age\ Group_{ij}^{7-9}$ for ages 7–9, $Age\ Group_{ij}^{10-12}$ for ages 10–12, and $Age\ Group_{ij}^{13-16}$ for ages 13–16. Because these age groups encompass all individuals in our sample, β_1, β_2, and β_3 identify the main effect of preschool attendance for each age group (further below, we also report results from specifications in which PRE_{ij} is interacted with the ten individual age dummies in $\mathbf{AGE}_{ij}$ instead). We weight all of our regressions using the sampling weights provided with the Uwezo data, and cluster standard errors at the district level.

The main parameters of interest in equation 1 are β_1, β_2, and β_3. They identify the age-group specific causal effect of preschool attendance under the assumption that among siblings, selection into preschool is uncorrelated with any other determ-

[12]In alternative specifications which do not include household fixed effects, we also control for the other socio-demographic characteristics shown in Table 1.

inants of the outcome.[13] While comparatively weak, this assumption might be violated for several reasons, two of which are particularly salient. First, given that pre-primary education is costly, household income shocks around preschool age may be driving siblings' differential enrollment. Because such income shocks can influence children's educational success also in other ways (e.g. Shah and Steinberg, 2017), this could introduce bias into our estimates. We address this concern by including detailed district-level controls for early-life economic conditions in our regressions. Second, households with limited resources may choose to invest only in children with the "highest potential."[14] In this case, we would expect children who attended preschool to differentially benefit also from other investments such as private tutoring. However, in a robustness check below, we find no evidence of such behavior.

Which factors drive the between-sibling variation in preschool attendance in our data if not income shocks and differential investments based on relative "potential?" We investigate this question in Table 2, which reports results of regressions of the indicator for preschool attendance on the control variables. Columns 1 and 3 show estimates from a specification without household fixed effects for Kenya and Tanzania, respectively. In both countries, children of educated mothers and from wealthier households are more likely to have attended preschool, underlining the importance of controlling for between-family differences. Columns 2 and 4 show that once household fixed effects are included in the regressions, most of the factors that still vary between siblings are no longer predictive of preschool attendance, including the proxies for early-life economic conditions.[15]

The lower part of Table 2 reports the coefficients on the cohort dummies. There is a marked and nearly monotonic trend in both countries, with later cohorts being significantly more likely to have attended preschool. This trend is especially pro-

[13] Note that in equation 1, children without siblings with different preschool status do not contribute to the identification of β_1, β_2, and β_3. We nevertheless keep these children in our sample as this increases the precision of our estimates. In a robustness check below, we show that results are qualitatively and quantitatively similar when focusing on the restricted sample of households with within variation in preschool attendance.

[14] Such reinforcing behavior has been found in several previous studies on developing countries, see Almond and Mazumder (2013). Alternatively, compensatory behavior might lead to negative selection into preschool, biasing our estimates downward.

[15] To avoid cluttering, rather than separate dummies for early-life economic conditions at each age, specifications in Table 2 simply include the number of positive and negative rainfall shocks and the average log night lights before school entry. Results are qualitatively similar if we include the full set of controls instead; in particular, early-life economic conditions appear to be largely orthogonal to preschool attendance. Notably, this is not due to poor measurement, as these variables are highly predictive of children's literacy and numeracy skills (results available upon request).

nounced in Tanzania, where the attendance rate was much lower than in Kenya at baseline (see Section 2) and where it increased by 15 percentage points over the twelve cohorts in our sample. Our interpretation of these estimates is that the expansion of pre-primary education during our study period led to differences in preschool availability between siblings, which in turn are driving the differences in attendance. As long as these changes in availability are unrelated to changes in other determinants of educational outcomes, this implies that we identify the true causal effects of preschool attendance in the analysis below.[16]

[16] Ideally, we would like to further investigate this hypothesis using data on preschool openings by district and year. Unfortunately, however, such data do not appear to exist.

Table 2: Predicting preschool attendance

	Kenya		Tanzania	
	(1)	(2)	(3)	(4)
Female	0.003	-0.001	0.008**	0.003
	(0.003)	(0.002)	(0.003)	(0.003)
Firstborn	0.005	0.000	-0.002	0.002
	(0.004)	(0.003)	(0.005)	(0.004)
Female × firstborn	-0.001	0.001	-0.004	0.000
	(0.004)	(0.003)	(0.005)	(0.005)
Mother ≥ some primary edu.	0.016**		0.097***	
	(0.007)		(0.006)	
No. of children in household	-0.002		-0.012***	
	(0.002)		(0.002)	
Household wealth index	0.006**		0.052***	
	(0.003)		(0.004)	
Rural location	-0.000*		-0.016	
	(0.000)		(0.022)	
No. of negative rainfall shocks	-0.003**	-0.001	-0.007*	-0.000
	(0.001)	(0.001)	(0.004)	(0.004)
No. of positive rainfall shocks	-0.004	-0.000	0.006	0.007*
	(0.002)	(0.001)	(0.004)	(0.004)
Log night light density	0.009	0.002	0.016	0.011
	(0.006)	(0.005)	(0.016)	(0.012)
Cohort				
1996			0.023***	0.021
			(0.008)	(0.014)
1997			0.030***	0.035***
			(0.009)	(0.011)
1998	0.005	0.002	0.047***	0.050***
	(0.008)	(0.004)	(0.009)	(0.013)
1999	0.012*	0.003	0.062***	0.069***
	(0.007)	(0.004)	(0.010)	(0.014)
2000	0.011	0.006	0.088***	0.084***
	(0.007)	(0.004)	(0.011)	(0.015)
2001	0.019***	0.006	0.090***	0.086***
	(0.007)	(0.005)	(0.011)	(0.016)
2002	0.022***	0.008	0.098***	0.101***
	(0.007)	(0.005)	(0.011)	(0.015)
2003	0.026***	0.004	0.110***	0.106***
	(0.007)	(0.005)	(0.011)	(0.016)
2004	0.023**	0.007	0.113***	0.109***
	(0.009)	(0.005)	(0.012)	(0.017)
2005	0.033***	0.007	0.099***	0.099***
	(0.009)	(0.006)	(0.012)	(0.017)
2006	0.040***	0.014**	0.147***	0.149***
	(0.007)	(0.006)	(0.016)	(0.023)
2007	0.032***	0.013		
	(0.009)	(0.009)		
Household fixed effects	No	Yes	No	Yes
Observations	223,339	223,339	293,757	293,757

Notes: The table reports estimates from regressions of an indicator for preschool attendance on the variables listed in rows and Uwezo wave dummies. Standard errors in parentheses are clustered at the district level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

5 Results

We now present our main results. The following subsection reports estimates of the effect of preschool attendance on school progression, and the second subsection reports the corresponding impacts on literacy and numeracy skills. We then present results from regressions that probe for heterogeneity of these effects by children and household characteristics. After that, we present evidence on a potential mechanism and compare our findings to the results from the previous literature.

5.1 Effects of preschool attendance on school progression

Table 3 shows estimates of the effect of preschool attendance on the highest grade of school attended. Column 1 reports results from parsimonious specifications which only control for age dummies, their interaction with cohort dummies, and district fixed effects. Due to the frequent late enrollment in pre-primary and subsequently primary education, children who went to preschool initially have attended fewer grades than their peers who directly entered primary school. However, these children also progress through grades at a faster pace and eventually overtake their peers: at ages 13–16, they have attended 0.18 more grades in Kenya and 0.31 more grades in Tanzania. Columns 2–4 successively add controls for socio-demographic characteristics, early-life economic conditions, and household fixed effects to the regressions. Consistent with the idea of positive selection into pre-primary education, this tends to reduce the coefficients: from our preferred specification in column 4, children who went to preschool are now estimated to accumulate the same number of grades in Kenya and about 0.1 more grades in Tanzania by ages 13-16.[17]

An interesting question is whether children who went to preschool catch up with their peers in terms of grades completed because they skip more or repeat fewer grades while in school, or because they are less likely to drop out of school. To investigate this issue, column 5 of Table 3 presents estimates of the effect of preschool

[17] Besides household fixed effects accounting for selection on unobserved factors, there are at least three further potential explanations for the reduction of coefficients between columns 3 and 4. First, attenuation bias due to measurement error in attendance is aggravated in the between-sibling specification, an issue that we discuss in detail below. Second, the inclusion of household fixed effects nets out any positive sibling spillovers. Third, the effects in column 4 are identified only from households with within variation, which might differ from those in the full sample. Investigating this last possibility, we found that households with both attending and non-attending children were larger, poorer, and more likely to be located in a rural area. However, when we restricted the sample to these households only, the inclusion of household fixed effects similarly led to a decline in the estimated long-term benefits of preschool. This suggests that identification based on a different sample is not driving the change in coefficients between columns 3 and 4 of Table 3.

attendance on current enrollment based on our preferred specification with household fixed effects. The results indicate that children who went to preschool are indeed more likely to be enrolled, especially at higher ages and in Tanzania. Thus, lower dropout is at least partly underlying the catch-up observed in column 4.[18] Finally, Figure 1 plots estimates from regressions of the highest grade completed and enrollment status in which the effect of preschool attendance is allowed to differ at each age, rather than across age groups. The plots show that the impacts of attendance on these outcomes rise almost monotonically with age, confirming the results from Table 3.

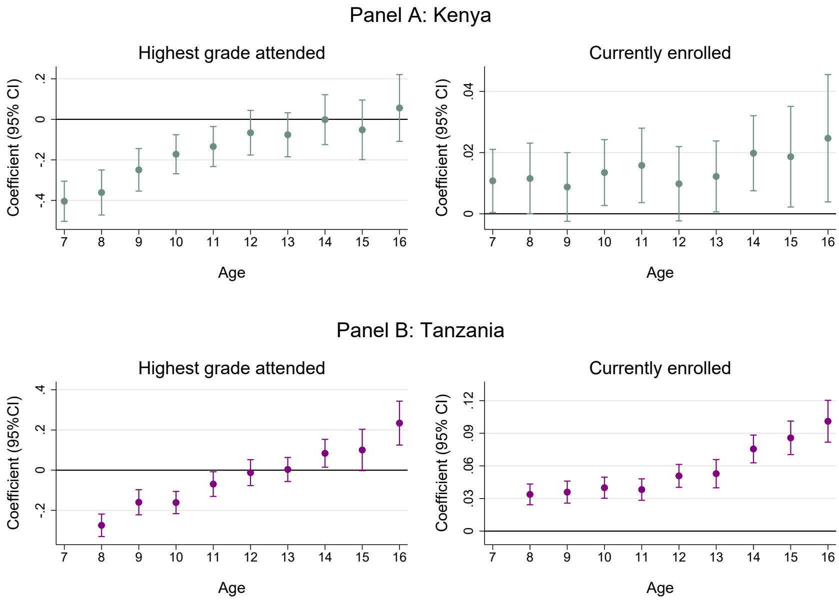
[18]Unfortunately, the Uwezo data do not contain information on school starting age and grade repetition, which prevents us from fully disentangling the mechanisms behind this catch-up. In column 5 of Table 3, note that for the younger two age groups, enrollment rates are very high and the predicted probabilities from the linear probability model sometimes exceed 100 percent. Addressing this issue, we confirmed that probit models and a simple comparison of means also suggest that children who attended preschool are more likely to be currently enrolled in school.

Table 3: Preschool attendance and school progression

	Highest grade attended				Currently enrolled
	(1)	(2)	(3)	(4)	(5)
	Panel A: Kenya				
Attended preschool					
7-9 years old	-0.234***	-0.252***	-0.242***	-0.336***	0.011*
	(0.040)	(0.042)	(0.042)	(0.045)	(0.005)
10-12 years old	0.017	-0.006	-0.006	-0.123***	0.013**
	(0.024)	(0.023)	(0.023)	(0.044)	(0.006)
13-16 years old	0.178***	0.161***	0.153***	-0.023	0.018***
	(0.034)	(0.034)	(0.033)	(0.055)	(0.006)
p(equal coefficients)	0.000	0.000	0.000	0.000	0.032
Observations	218,728	218,728	218,728	218,728	218,728
	Panel B: Tanzania				
Attended preschool					
8-9 years old	-0.071***	-0.127***	-0.115***	-0.212***	0.036***
	(0.018)	(0.018)	(0.018)	(0.025)	(0.005)
10-12 years old	0.081***	0.028	0.031*	-0.074***	0.044***
	(0.020)	(0.018)	(0.019)	(0.025)	(0.004)
13-16 years old	0.313***	0.258***	0.243***	0.096***	0.076***
	(0.026)	(0.026)	(0.024)	(0.031)	(0.006)
p(equal coefficients)	0.000	0.000	0.000	0.000	0.000
Observations	284,396	284,396	284,396	284,396	284,396
	Controls included in panels A and B				
Age × cohort effects	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	No	No
Socio-demographics	No	Yes	Yes	Yes	Yes
Early-life conditions	No	No	Yes	Yes	Yes
Household fixed effects	No	No	No	Yes	Yes

Notes: The table reports estimates from regressions of the highest grade attended and enrollment status on a dummy for preschool attendance and control variables as indicated in the lower panel. The dummy for preschool attendance is interacted with three age-group dummies (7/8-9, 10-12, and 13-16 years), and the table reports the estimated effect of preschool attendance separately for each group. Socio-demographic controls include the variables shown in Table 1, dummies for birth order and their interactions with gender, and sibling age span. Controls for early-life economic conditions include district-level indicators for negative and positive rainfall shocks and district-level log night lights at each age between 0 and 5 (0 and 6 in Tanzania), all of which are interacted with a dummy for rural location. Specifications in columns 1-3 additionally control for Uwezo wave dummies. See Appendix Table 1 for outcome means by age. Standard errors in parentheses are clustered at the district level. * p<0.10, ** p<0.05, *** p<0.01.

Figure 1: Preschool attendance and school progression, by age



Notes: The figure plots coefficient estimates and 95% confidence intervals from regressions of the highest grade attended and an indicator for current enrollment on preschool attendance. The indicator for preschool attendance is interacted with age dummies, and the figure shows the estimated effect of preschool attendance separately for each age. Specifications are otherwise equal to the household fixed effects regressions reported in columns 4 and 5 of Table 3.

5.2 Effects of preschool attendance on literacy and numeracy skills

Table 4 reports estimates of the effect of preschool attendance on children's literacy and numeracy skills. Column 1 shows results from a specification with only basic controls and the composite test score as outcome. In Kenya, children who went to preschool have slightly higher scores than their peers at early ages, and this advantage grows to 0.1 SD for the two older age groups. In contrast, in Tanzania, children with pre-primary education outperform their peers by 0.26 SD already early on, but this difference decreases to 0.22 SD for the group of 13–16-year-olds. Columns 2–4 successively add control variables and household fixed effects to these regressions. Similar to the pattern found for the highest grade attended, this substantially reduces the size of the estimates for Tanzania, where the impact for the oldest age group is now estimated at 0.08 SD. In contrast, the coefficients for Kenya are relatively stable across specifications, suggesting that there is little selection into preschool based on academic ability.

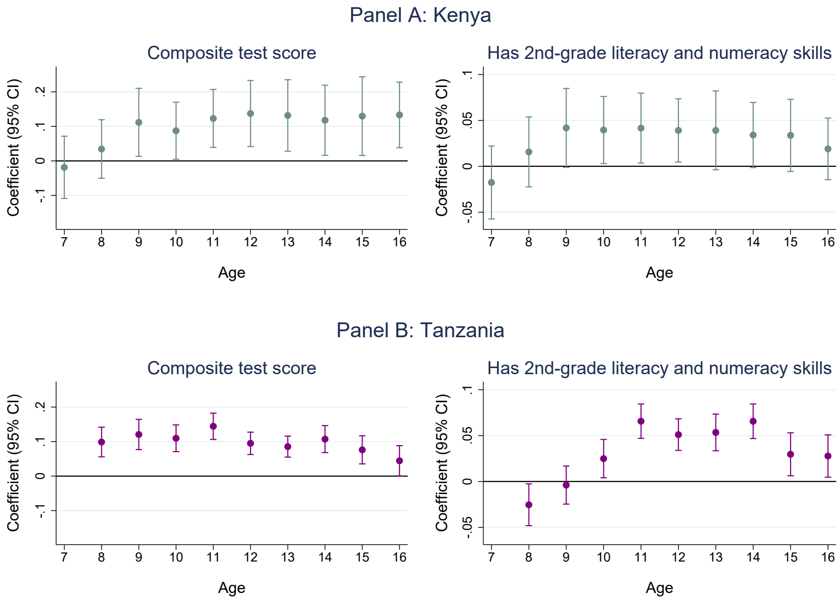
Column 5 presents results for a specification with the dummy for possessing second-grade literacy and numeracy skills as dependent variable. In both countries, the effect for the youngest age group is close to zero. This should come as no surprise because preschools are unlikely to teach children such advanced skills, and because attendance actually leads to a reduction in grades completed at these ages (see Table 3). Note that this finding implies that the positive impact on test scores for young children in Tanzania in column 4 must be due to pre-primary education boosting their very basic literacy and numeracy skills. Mirroring the long-term gains on the composite test score, the estimates for 13–16-year-olds show a 3.2 (4.7) percentage point increase in the likelihood to achieve second-grade skills in Kenya (Tanzania), which corresponds to a sizable 4 (7) percent over the mean. Finally, Figure 2 shows age-by-age impacts on test scores and the dummy for basic skills which confirm the patterns observed in Table 4.

Table 4: Preschool attendance and literacy and numeracy skills

	Composite test score				2^{nd}-grade lit./num.
	(1)	(2)	(3)	(4)	(5)
	Panel A: Kenya				
Effect of preschool					
at ages 7-9	0.036*	0.015	0.012	0.042	0.013
	(0.022)	(0.022)	(0.021)	(0.040)	(0.016)
at ages 10-12	0.097***	0.078***	0.077***	0.114***	0.040**
	(0.022)	(0.021)	(0.021)	(0.041)	(0.015)
at ages 13-16	0.104***	0.095***	0.098***	0.125***	0.032*
	(0.021)	(0.021)	(0.021)	(0.044)	(0.017)
p (equal coefficients)	0.010	0.005	0.003	0.001	0.093
Observations	218,134	218,134	218,134	218,134	218,134
	Panel B: Tanzania				
Effect of preschool					
at ages 8-9	0.256***	0.204***	0.195***	0.108***	-0.016*
	(0.017)	(0.016)	(0.015)	(0.018)	(0.009)
at ages 10-12	0.263***	0.212***	0.209***	0.113***	0.046***
	(0.014)	(0.013)	(0.013)	(0.014)	(0.007)
at ages 13-16	0.217***	0.163***	0.169***	0.081***	0.047***
	(0.016)	(0.015)	(0.014)	(0.014)	(0.008)
p (equal coefficients)	288,084	288,084	288,084	288,084	288,084
Observations	0.021	0.008	0.032	0.152	0.000
	Controls included in panels A and B				
Age × cohort effects	Yes	Yes	Yes	Yes	Yes
District fixed effects	Yes	Yes	Yes	No	No
Socio-demographics	No	Yes	Yes	Yes	Yes
Early-life conditions	No	No	Yes	Yes	Yes
Household fixed effects	No	No	No	Yes	Yes

Notes: The table reports estimates from regressions of the composite test score and the indicator for achieving second-grade literacy and numeracy skills on a dummy for preschool attendance and control variables as indicated in the lower panel. The dummy for preschool attendance is interacted with three age-group dummies (7/8-9, 10-12, and 13-16 years), and the table reports the estimated effect of preschool attendance separately for each group. Socio-demographic controls include the variables shown in Table 1, dummies for birth order and their interactions with gender, and sibling age span. Controls for early-life economic conditions include district-level indicators for negative and positive rainfall shocks and district-level log night lights at each age between 0 and 5 (0 and 6 in Tanzania), all of which are interacted with a dummy for rural location. Specifications in columns 1-3 additionally control for Uwezo wave dummies. See Appendix Table 1 for outcome means by age. Standard errors in parentheses are clustered at the district level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Figure 2: Preschool attendance and school progression, by age



Notes: The figure plots coefficient estimates and 95% confidence intervals from regressions of the highest grade attended and an indicator for current enrollment on preschool attendance. The indicator for preschool attendance is interacted with age dummies, and the figure shows the estimated effect of preschool attendance separately for each age. Specifications are otherwise equal to the household fixed effects regressions reported in columns 4 and 5 of Table 3.

5.3 Heterogeneity

In Table 5, we explore the heterogeneity of the preschool impacts along several dimensions. Columns 1–3 report estimates from specifications that allow the effects to differ by length of attendance. The results reveal that children who went to preschool for two or three years tend to have attended fewer school grades than those who went for only one year, likely because they entered primary school later. This difference shrinks over time, however, suggesting that children with more years of preschool progress through grades at a faster pace. The impacts on the composite test score similarly tend to be less positive for children with several years of pre-primary education, at least in Kenya. This could be due to the fact that they are further behind in school, or due to negative selection at the intensive margin, with parents keeping children with lower academic ability in preschool for longer. Moreover, the fact that children of different ages are often taught together in preschool implies that attending for more years does not necessarily mean learning higher-level skills, a reality that might also contribute to the lack of a positive effect at the intensive margin.

Columns 4–5 show that in both countries, the effects of preschool attendance are consistently more positive for girls, even though the differences are not always statistically significant at conventional levels. Furthermore, columns 6–7 reveal that the improvements in literacy and numeracy skills among 13– to 16-year-olds tend to be larger for those with uneducated mothers: in Tanzania, for example, preschool attendance raises the composite test score for these children by 0.14 SD, compared to 0.06 SD for children with mothers who have at least some formal education. Finally, columns 8–9 show that in Tanzania only, longer-term gains in the number of grades attended and test scores are substantially larger in more urban, high-economic-activity areas, as proxied by living in a district with night light density above the 85th percentile of the national distribution.

Table 5: Heterogeneity

	by number of years attended			by gender		by mother's education		by night light density	
	main effect	× 2 years	× 3 years	main effect	× female	main effect	× some educ.	main effect	× > 85%
Panel A: Kenya									
1. Highest grade attended									
7-9 years old	-0.211***	-0.163***	-0.292***	-0.341***	0.013	-0.270***	-0.024	-0.336***	0.053
	(0.050)	(0.036)	(0.041)	(0.049)	(0.046)	(0.073)	(0.096)	(0.044)	(0.184)
10-12 years old	-0.043	-0.141***	-0.192***	-0.164***	0.083**	-0.094	-0.033	-0.127***	0.009
	(0.048)	(0.030)	(0.044)	(0.047)	(0.036)	(0.070)	(0.086)	(0.037)	(0.211)
13-16 years old	-0.031	-0.050	-0.032	-0.071	0.100**	-0.024	-0.046	-0.054	0.093
	(0.058)	(0.031)	(0.053)	(0.066)	(0.049)	(0.073)	(0.092)	(0.047)	(0.225)
Observations		218,728		218,728		218,728		218,728	
2. Composite test score									
7-9 years old	0.045	-0.051*	0.029	0.034	0.019	0.076	-0.041	0.022	0.104
	(0.041)	(0.027)	(0.034)	(0.046)	(0.032)	(0.072)	(0.091)	(0.043)	(0.098)
10-12 years old	0.134***	-0.049**	-0.017	0.086**	0.055**	0.143*	-0.069	0.116**	-0.013
	(0.043)	(0.021)	(0.030)	(0.041)	(0.024)	(0.079)	(0.094)	(0.045)	(0.104)
13-16 years old	0.168***	-0.064**	-0.068**	0.100**	0.054	0.235**	-0.148	0.136***	-0.037
	(0.046)	(0.024)	(0.033)	(0.046)	(0.038)	(0.100)	(0.120)	(0.049)	(0.103)
Observations		288,084		288,084		288,084		288,084	
Panel B: Tanzania									
1. Highest grade attended									
7-9 years old	-0.237***	-0.273***		-0.256***	0.087***	-0.255***	0.070	-0.205***	-0.000
	(0.054)	(0.076)		(0.028)	(0.029)	(0.043)	(0.049)	(0.027)	(0.068)
10-12 years old	-0.156**	-0.237***		-0.080***	0.012	-0.127***	0.058	-0.095***	0.102
	(0.060)	(0.064)		(0.028)	(0.029)	(0.046)	(0.047)	(0.026)	(0.072)
13-16 years old	-0.052	-0.125*		0.100***	-0.010	-0.047	0.167***	0.048*	0.310***
	(0.058)	(0.069)		(0.032)	(0.039)	(0.055)	(0.062)	(0.027)	(0.104)
Observations	107,825			284,396		284,396		284,396	
2. Composite test score									
7-9 years old	0.080**	0.077**		0.080***	0.054**	0.080**	0.031	0.096***	0.029
	(0.034)	(0.037)		(0.022)	(0.025)	(0.039)	(0.050)	(0.018)	(0.060)
10-12 years old	0.127***	0.017		0.089***	0.048***	0.140***	-0.039	0.097***	0.090***
	(0.025)	(0.027)		(0.016)	(0.018)	(0.037)	(0.044)	(0.015)	(0.032)
13-16 years old	0.124***	-0.056**		0.069***	0.024	0.144***	-0.082*	0.075***	0.105**
	(0.028)	(0.031)		(0.017)	(0.018)	(0.040)	(0.046)	(0.014)	(0.049)
Observations	107,605			288,084		288,084		288,084	

Notes: The table reports estimates from specifications in which the dummy for preschool attendance is interacted with three age-group dummies as in Tables 3 and 4 and additionally with two dummies for attending preschool for two and three years (columns 1-3), a female dummy (columns 4-5), a dummy for having a mother with at least some formal education (columns 6-7), and a dummy for living in a district with night light density above the 85^{th} national percentile in the year 2000 (columns 8-9). The table reports the main effects for each age group as well as the coefficients on the interactions with these characteristics. Base levels in columns 1/4/6/8: attended for 1 year/male/no formal education/district night lights below the 85^{th} national percentile. Outcome variables are indicated in cursive in the rows above the respective regressions. All regressions include household fixed effects and controls as in column 4 of Tables 3 and 4. Standard errors in parentheses are clustered at the district level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

5.4 "Head start" as a potential mechanism

An interesting open question is how exactly preschools improve learning outcomes for attending children. One obvious explanation is that with their focus on teaching basic literacy and numeracy, preschools give children a head start that makes it easier for them to follow the primary school curriculum. An implication of this explanation is that these students should have higher skills already at school start. To what extent this is indeed the case is not immediately obvious from our results above, which are based on same-age comparisons that blend the potential effect of a skill boost from preschool attendance with any effect on learning due to later school entry and lower grade attainment. In order to disentangle these two channels, one would ideally want to compare children who differ in terms of their preschool attendance but who started school at the same time. Unfortunately, such a comparison is not feasible here because school starting age is not observed in the Uwezo data.

As an alternative way to separate the test score impacts due to learning in preschool from the impacts due to later school entry, Table 6 reports results from specifications that control for the highest grade attended. If children do indeed get a head start from attending preschool, we would expect the coefficients in these regressions to increase compared to our main results, especially for the youngest children who are furthest behind in school. This turns out to be the case: for example, the effect for 7– to 9-year-olds in Kenya is now estimated to be 0.16 SD, compared to only 0.04 SD when their lower grade attainment is not taken into account. Thus, these results suggest that preschools do in fact give children a head start in the learning of literacy and numeracy skills, and that this might drive the long-term improvements in learning outcomes.

Table 6: Effects of preschool attendance on composite test scores after controlling for highest grade attended

	Composite test score	
	Baseline	Controlling for highest grade attended
	(1)	(2)
	Panel A: Kenya	
Effect of preschool		
at ages 7-9	0.042	0.157***
	(0.040)	(0.036)
at ages 10-12	0.114***	0.148***
	(0.041)	(0.035)
at ages 13-16	0.125***	0.177***
	(0.044)	(0.039)
Observations	218,134	218,134
	Panel B: Tanzania	
Effect of preschool		
at ages 7-9	0.108***	0.157***
	(0.018)	(0.017)
at ages 10-12	0.113***	0.126***
	(0.014)	(0.014)
at ages 13-16	0.081***	0.104***
	(0.014)	(0.013)
Observations	288,084	288,084

Notes: Column 1 replicates the estimates shown in column 4 of Table 4. Specifications in column 2 add separate dummies for the number of grades attended to these regressions as controls. See the notes to Table 4 for further details on included control variables. Standard errors in parentheses are clustered at the district level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

5.5 Comparison with results from previous studies

We now compare our results to the previous literature on preschool effectiveness in developing countries. Focusing first on schooling, our estimates are qualitatively similar to those by Berlinski et al. (2008) for Uruguay. Like us, the authors find that children who go to preschool initially fall behind in terms of grades attended but that they progress through school faster later on, leading to an increase in educational attainment by 0.8 years of schooling at age 15. For Egypt, Krafft (2015) similarly shows that children with pre-primary education accumulate 0.4 more years of schooling by ages 18–29. Unlike her, we do not observe final educational attainment, which may explain our smaller estimates at ages 13–16. However, the monotonic positive trend in the impact on grades attended in Figure 2 suggests that children who attended preschool might eventually acquire more years of schooling also in Kenya and Tanzania.

Turning to cognitive skills, Berlinski et al. (2009) find that in Argentina, preschool attendance increases third-grade students' math and language test scores by 0.23 SD. For Indonesia, Brinkman et al. (2017) similarly show that three years after the establishment of early childhood services, children from poor households improved by 0.20 SD on an index of language and cognitive development.[19] One potential reason why our estimates are smaller is the presence of ceiling effects: as discussed in Section 3, the Uwezo assessments measure second-grade skills, and although most children in our sample do not reach the highest competency levels on the tests, a non-negligible fraction of the older children in particular does (see Appendix Table 1). As ceiling effects lead to an attenuation of regression estimates, this could explain our lower point estimates compared to the previous literature, as well as the slight fade-out of the effects on cognitive skills visible for the older age groups in Figure 2.

Finally, a recent study by Bouguen et al. (2018) shows that in Cambodia, 6-year-old children scored 0.19 SD lower on an index of cognitive development one year after preschools were constructed in their villages. Investigating potential channels, the authors find that preschool construction led to a shift from early enrollment in primary school to enrollment in preschool, suggesting that the decrease in cognitive skills was partly due to the lower emphasis on literacy and numeracy skills in the preschool curriculum.[20] Interpreted in the light of our results, this shift in en-

[19] Behrman et al. (2004) and Martinez et al. (2012) also document positive effects on measures of child development that include cognitive skills, but these are difficult to compare directly to our outcomes.

[20] Also in our setting, early enrollment in primary school might be part of the counterfactual. How-

rollment led to a decrease in grades attended early on. While Bouguen et al. (2018) cannot investigate the longer term consequences of this change, our estimates reveal that despite low or negative initial returns, children who attend preschool can catch up in terms of grades attended and strongly benefit in terms of learning later on. These longer-run estimates on cognitive skills in particular are a key contribution of our paper over the previous literature, which has only been able to study short-run effects.

6 Robustness

6.1 Addressing selection concerns

In Section 3 above, we argue that the between-sibling variation in preschool attendance is likely due to changes in availability, which came about because of the expansion of the pre-primary sector during our study period. One might worry, however, that this variation instead reflects child-specific investments that are correlated with unobservables. As an example, households with limited resources may choose to invest only in children with the "highest potential." If this is indeed the case, one would expect that families differentially spend on children who are sent to preschool also in other ways. We test this hypothesis in a falsification exercise by examining whether children who attended preschool are more likely to benefit from two other costly educational inputs observed in our data: private after-school tutoring and enrollment in private school.[21] Table 7 shows that in regressions of indicators for receiving these inputs, the coefficients on preschool attendance are close to zero and precisely estimated. Although these are just two out of many ways in which parents invest in their children, these estimates thus suggest that differential investments based on child unobservables are not driving our main results.

ever, whereas Bouguen et al. (2018) report that 60 percent of children in control villages attended primary school in the year before they reached the official school starting age, only 26 percent of 5-year-olds in the 1999 Kenyan census and 8 percent of 6-year-olds in the 2002 Tanzanian census did so. Note also that respondents in the Uwezo surveys answer separate questions about current and past preschool attendance and current and past school attendance, allowing us to separate the impacts of preschool attendance from those of early enrollment in primary school.

[21] See Wamalwa and Burns (2017) for an analysis of private school effectiveness in Kenya using the Uwezo survey data.

Table 7: Preschool attendance and other educational investments

	After-school tutoring (1)	Private school (2)
	Panel A: Kenya	
Attended preschool		
7-9 years old	-0.002	-0.004
	(0.011)	(0.010)
10-12 years old	0.005	-0.014
	(0.011)	(0.011)
13-16 years old	0.020	-0.015
	(0.017)	(0.011)
Observations	223,339	208,424
	Panel B: Tanzania	
Attended preschool		
8-9 years old	-0.000	-0.005
	(0.025)	(0.004)
10-12 years old	0.018	-0.006*
	(0.024)	(0.003)
13-16 years old	0.020	-0.001
	(0.019)	(0.003)
Observations	25,346	264,810

Notes: Columns 1 and 2 report estimates from regressions in which the dependent variables are an indicator for receiving private after-school tutoring and and indicator for currently attending private school, respectively. The specifications are otherwise identical to the one in column 4 of Tables 3 and 4. Means of the dependent variables for the youngest/middle/oldest age group: Kenya: after-school tutoring: 0.27/0.31/0.37; Tanzania: after-school tutoring: 0.22/0.25/0.25; Kenya: private school: 0.24/0.15/0.09; Tanzania: private school: 0.06/0.05/0.08. After-school tutoring is only observed in 2014 in Tanzania. Private school is only observed for students who are currently enrolled in school. Standard errors in parentheses are clustered at the district level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

To judge the importance of selection bias more generally, we next ask how large such bias would need to be in order to explain away our main effects. Our analysis builds on the approach presented in Oster (2017), which relies on comparing the coefficients of interest and the R-squared between regressions with and without control variables to gain insights into the influence of omitted variables. Here, we focus on the calculation of δ, which is the ratio of the impact of unobservables to the impact of observable controls that would drive the coefficient on the treatment variable to zero. As a point of reference, Oster (2017) suggests that effects for which $\delta > 1$ can be considered robust. Applying this method to our case, we contrast estimates of the impact of preschool attendance from a specification with only basic controls (as in column 1 of Tables 3 and 4) with those from our preferred specification with household fixed effects. We restrict our sample to households with variation in preschool attendance for this analysis because for all other households, the fixed effects fully explain preschool attendance, leaving no role for selection on unobservables.[22]

Table 8 reports the results from this exercise. Columns 1 and 3 show estimates from specifications with only basic controls, with regressions underlying columns 2 and 4 adding further controls and household fixed effects. As would be expected, the estimates from our preferred specification are generally very similar in the restricted sample compared to the full sample used in the main analysis (we test the equality of these coefficients more formally below). Moving from the basic specification to our preferred specification substantially increases the R-squared and tends to decrease the coefficient estimates, in line with what is observed in Tables 3 and 4. Based on these differences, we report the implied δ for preschool impacts for 13– to 16-year-olds, for whom we find the most positive impacts and where selection on unobservables is thus the most relevant concern. In three out of four specifications, δ is greater than one, implying that selection on unobservables would have to be greater than selection on the observed control variables to drive the preschool impacts to zero. For the case of the composite test score, this value is even above four for both countries, which strongly suggests that omitted variable bias is not driving these results.[23]

[22] We thank Emily Oster for this suggestion.

[23] There is a strong mechanical relationship between highest grade attended and age, as reflected by the high values of R-squared in column 1 of Table 8. This limits the additional explanatory power any observed controls can have in these regressions, which partly explains the lower values for δ in column 2 compared to column 4 of Table 8.

Table 8: Judging the importance of selection on unobservables

	Highest grade attended		Composite test score	
	(1)	(2)	(3)	(4)
	Panel A: Kenya			
7-9 years	-0.703***	-0.423***	-0.104	0.023
	(0.072)	(0.070)	(0.063)	(0.066)
10-12 years	0.110	-0.090	0.251***	0.165**
	(0.070)	(0.078)	(0.064)	(0.066)
13-16 years	0.144*	-0.020	0.092	0.111
	(0.086)	(0.085)	(0.069)	(0.075)
Observations	7,229	7,229	7,141	7,141
R-squared	0.688	0.869	0.114	0.619
δ (13-16 years)		-0.119		11.317
	Panel B: Tanzania			
8-9 years	-0.115***	-0.147***	0.108***	0.093***
	(0.030)	(0.041)	(0.028)	(0.032)
10-12 years	-0.154***	-0.088**	0.073***	0.111***
	(0.034)	(0.038)	(0.023)	(0.022)
13-16 years	0.110***	0.083**	0.128***	0.120***
	(0.042)	(0.038)	(0.022)	(0.024)
Observations	29,981	29,981	30,427	30,427
R-squared	0.622	0.831	0.072	0.637
δ (13-16 years)		1.150		4.553
	Controls included in panels A and B			
Age × cohort effects	Yes	Yes	Yes	Yes
District fixed effects	Yes	No	Yes	No
Socio-demographics	No	Yes	No	Yes
Early-life conditions	No	Yes	No	Yes
Household fixed effects	No	Yes	No	Yes

Notes: The table reports regression estimates that provide the inputs into the computation of Oster's (2017) δ, that is, the ratio of the impact of unobservables to the impact of observable controls that would drive the coefficient on preschool attendance for 13- to 16-year-olds to zero. For details on the underlying method, see text and Oster (2017). For this analysis, the sample is restricted to households with variation in preschool attendance. The regressions are otherwise identical to those in columns 1 and 4 in Table 3 (columns 1 and 2 in the current table) and columns 1 and 4 in Table 4 (columns 3 and 4 in the current table). To calculate δ, we use the Stata command `-psacalc-`, setting the maximum achievable R-squared ('Rmax') to 1.3 times the R-squared in the regression with household fixed effects (and at most 1). * p<0.10, ** p<0.05, *** p<0.01.

6.2 Further robustness checks

We now address a number of further potential concerns regarding our empirical analysis. First, as discussed in Section 3, our results may be affected by recall error in the retrospectively reported preschool variable. Note that such recall error can lead to upward bias only if it systematically varies between siblings who did and did not attend preschool; any general, idiosyncratic recall error will simply drive our estimates towards zero. We investigate this issue by taking advantage of the repeated cross-sectional nature of our data, which lets us follow cohorts over time. Appendix Table 2 shows the fraction of children in each cohort reporting to have attended preschool separately for each Uwezo survey wave.[24] If our data were contaminated by recall error, we would expect these fractions to change over time. This is not the case for the majority of cohorts, though, which suggests that recall error does not bias our results.

Second, our estimates might be attenuated by measurement error in the preschool attendance variable. Such downward bias is particularly relevant in the context of sibling fixed effects models, in which the signal to noise ratio of the measurement may be greatly reduced (Ashenfelter and Krueger, 1994). Because of the absence of repeated individual-level information on preschool attendance in our data, we are unable to establish the extent of this problem conclusively. However, the strong similarity of the aggregate, cohort-level measure over time in Appendix Table 2 suggests that measurement error is not a major issue in our context.

Third, one might worry that our results are driven by siblings who are very different in age, and who thus grew up under very distinct circumstances. For example, the impacts on cognitive skills might vary between age groups only because different cohorts attended preschools of very different quality (although as discussed in Section 2, the available metrics do not show clear trends in preschool quality over time). To address this issue, Online Appendix Table 1 presents estimates from regressions in which the sample is restricted to siblings who are born at most five years apart. As can be seen there, the results are qualitatively and quantitatively similar to our main findings.

[24] For this exercise, we focus on a comparable sample of districts that were visited in all waves of the Uwezo surveys. We disregard the 2014 wave in Tanzania because only a small subsample of districts were included in that year's survey. In the raw data, we observe level shifts in preschool attendance rates *for all cohorts* between some of the waves, likely because the question asking about preschool attendance changed. In Appendix Table 2, we therefore report regression-adjusted attendance rates after taking out wave fixed effects. Note that level shifts in preschool attendance do not influence our within-household results, which use variation within survey waves.

Finally, we conduct additional analyses that address a range of further potential concerns. Thus, in Appendix Table 3, we compare our main results with those obtained from a sample which is restricted to households with within variation in preschool attendance. The sub-sample results are qualitatively and, for the most part, quantitatively similar to the main results, but, as expected, less precisely estimated. In Online Appendix Figure 1, we further document that preschool impacts on the individual English, Swahili, and numeracy scores are very similar to the ones on the composite test scores used in the main analysis. Our main results are also robust to not using sampling weights, as indicated in Online Appendix Table 1. Lastly, Online Appendix Figure 2 presents Kaplan-Meier estimates as an alternative way to investigate the impacts of preschool attendance on grade progression. These estimates are in line with our main findings in Section 5.

7 Conclusion

Most children in Sub-Saharan Africa enroll in school nowadays, but they learn remarkably little there. One possible reason is that they enter school unprepared, which makes preschool programs that aim to get children ready for school a promising way to improve learning outcomes. While pre-primary education is becoming increasingly common within the region, to date very little is known about its effectiveness.

In this paper, we provide some of the very first evidence of preschool impacts on learning outcomes in Sub-Saharan African. We use data from large-scale surveys of children's educational attainment and cognitive skills from Kenya and Tanzania, which also collect retrospective information on preschool attendance. Our analysis compares the highest school grade attended as well as achievement on standardized literacy and numeracy tests of siblings who did and did not attend preschool. This strategy allows us to control for any determinants of pre-primary enrollment and outcomes that do not vary within households. We provide evidence that the leftover between-sibling variation in attendance is due to changes in availability, which came about because of a large expansion of preschool education during our study period.

Our results show that preschool education leads to important long-term learning benefits: at ages 13–16, children who went to preschool are three and five percentage points more likely to achieve basic, second-grade literacy and numeracy in Kenya and Tanzania, respectively. These gains materialize relatively late because children who attend preschool tend to enter primary school late and thus fall behind early on. However, the skills learned in preschool give them a head start in school, meaning that they can progress through grades faster and eventually catch up with their peers who did not attend preschool in terms of grades attended. Overall, our analysis shows that increasing access to pre-primary education can be an effective instrument to improve learning outcomes in Sub-Saharan Africa.

References

Aboud, F. E., Proulx, K., and Asrilla, Z. (2016). An impact evaluation of Plan Indonesia's early childhood program. *Canadian Journal of Public Health*, 107(4-5):e366–e372.

Almond, D. and Mazumder, B. (2013). Fetal Origins and Parental Responses. *Annual Review of Economics*, 5(1):37–56.

Ashenfelter, O. and Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *American Economic Review*, 84(5):1157–1173.

Behrman, J. R., Cheng, Y., and Todd, P. E. (2004). Evaluating Preschool Programs When Length of Exposure to the Program Varies: A Nonparametric Approach. *Review of Economics and Statistics*, 86(1):108–132.

Berlinski, S., Galiani, S., and Gertler, P. (2009). The effect of pre-primary education on primary school performance. *Journal of Public Economics*, 93(1-2):219–234.

Berlinski, S., Galiani, S., and Manacorda, M. (2008). Giving children a better start: Preschool attendance and school-age profiles. *Journal of Public Economics*, 92(5-6):1416–1440.

Bidwell, K., Parry, K., and Watine, L. (2013). *Exploring Early Education Programs in Peri-urban Settings in Africa: Nairobi Report*. Report, Innovations for Poverty Action.

Bietenbeck, J., Ericsson, S., and Wamalwa, F. M. (2017). Preschool attendance, school progression, and cognitive skills in east africa. IZA Discussion Paper 11212.

Bietenbeck, J., Piopiunik, M., and Wiederhold, S. (2018). Africa's Skill Tragedy: Does Teachers' Lack of Knowledge Lead to Low Student Performance? *Journal of Human Resources*, 53(3):553–578.

Bouguen, A., Filmer, D., Macours, K., and Naudeau, S. (2018). Preschool and Parental Response in a Second Best World: Evidence from a School Construction Experiment. *Journal of Human Resources*, 53(2):474–512.

Brinkman, S. A., Hasan, A., Jung, H., Kinnell, A., and Pradhan, M. (2017). The Impact of Expanding Access to Early Childhood Education Services in Rural Indonesia. *Journal of Labor Economics*, 35(S1):S305–S335.

Cortázar, A. (2015). Long-term effects of public early childhood education on academic achievement in chile. *Early Childhood Research Quarterly*, 32:13–22.

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

Currie, J. and Thomas, D. (1995). Does Head Start Make a Difference? *American Economic Review*, 85(3):341–64.

Daniel, K. (2012). Educational Behavior of Children not in Preschool. The Case study of Machakos District, Kenya. *International Journal of Applied Psychology*, 2(1):1–5.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, 1(3):111–134.

Edwards Jr., D. B., Klees, S. J., and Wildish, J. (2015). Dynamics of Low-Fee Private Schools in Kenya: Governmental Legitimation, School-Community Dependence, and Resource Uncertainty. *Teachers College Record*, forthcoming.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., and Michaelsen, J. (2015). The climate hazards infrared precipitation with stations – a new environmental record for monitoring extremes. *Scientific Data*, 2:150066.

Garces, E., Thomas, D., and Currie, J. (2002). Longer-Term Effects of Head Start. *American Economic Review*, 92(4):999–1012.

Hazarika, G. and Viren, V. (2013). The effect of early childhood developmental program attendance on future school enrollment in rural north india. *Economics of Education Review*, 34:146–161.

Henderson, J. V., Storeygard, A., and Weil, D. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, 102(2):944–1028.

Heyneman, S. P. and Stern, J. M. (2014). Low cost private schools for the poor: What public policy is appropriate? *International Journal of Educational Development*, 35(Supplement C):3 – 15.

Jones, S., Schipper, Y., Ruto, S., and Rajani, R. (2014). Can your child read and count? Measuring learning outcomes in East Africa. *Journal of African Economies*, 23(5):643–672.

Krafft, C. (2015). Increasing educational attainment in Egypt: The impact of early childhood care and education. *Economics of Education Review*, 46:127–143.

Kweka, A., Binagi, E., and Kainamula, V. (1997). The situation of early childhood education in Tanzania: the case of Temeke District. A draft report prepared for UNESCO Dar es Salaam. Technical report, UNESCO Office Dar es Salaam.

Martinez, S., Naudeau, S., and Pereira, V. (2012). The promise of preschool in africa: A randomized impact evaluation of early childhood development in rural mozambique.

Mghasse, N. E. and William, F. (2016). Practices and Challenges in the Provision of Pre-Primary Education in Tanzania. *African Research Review*, 10(1):1–16.

Mligo, I. R. (2016). Teachers' perceptions and concerns about the implementation of the 2005 preschool curriculum in Tanzania. *Early Years*, 36(4):353–367.

Mtahabwa, L. and Rao, N. (2010). Pre-primary education in Tanzania: Observations from urban and rural classrooms. *International Journal of Educational Development*, 30(3):227–235.

Nganga, L. W. (2009). Early childhood education programs in Kenya: challenges and solutions. *Early Years*, 29(3):227–236.

Ngware, M., Hungi, N., Kitsao-Wekulo, P., Mutisya, M., and Muhia, N. (2016). The Tayari Pre-Primary Program in Kenya: Getting Children Ready for Primary School. Baseline Report. Technical report, African Population and Health Research Center, Nairobi.

Nores, M. and Barnett, W. S. (2010). Benefits of early childhood interventions across the world: (Under) Investing in the very young. *Economics of Education Review*, 29(2):271–282.

Oster, E. (2017). Unobservable Selection and Coefficient Stability: Theory and Evidence. *Journal of Business and Economic Statistics*, pages 1–18.

President's Office of the United Republic of Tanzania (2016). Pre-Primary, Primary and Secondary Education Statistics in Brief. Dar es Salaam: President's Office of the United Republic of Tanzania, Regional Administration and Local Government.

Rao, N., Sun, J., Wong, J. M. S., Weekes, B., Ip, P., Shaeffer, S., Young, M., Bray, M., Chen, E., and Lee, D. (2014). Early childhood development and cognitive development in developing countries: a rigorous literature review. London: Department for International Development.

Schady, N., Behrman, J. R., Caridad Araujo, M., Azuero, R., Bernal, R., Bravo, D., Lopez-Boo, F., Macours, K., Marshall, D., Paxson, C., and Vakis, R. (2015). Wealth Gradients in Early Childhood Cognitive Development in Five Latin American Countries. *Journal of Human Resources*, 50(2):446–463.

Shah, M. and Steinberg, B. M. (2017). Drought of Opportunities: Contemporaneous and Long-Term Impacts of Rainfall Shocks on Human Capital. *Journal of Political Economy*, 125(2):527–561.

Tooley, J., Dixon, P., and Stanfield, J. (2008). Impact of Free Primary Education in Kenya: A Case Study of Private Schools in Kibera. *Educational Management Administration & Leadership*, 36(4):449–469.

UNESCO (2012). *Global education digest 2012. Opportunities lost: The impact of grade repetition and early school leaving*. Paris: UNESCO.

Uwezo (2015). *Are Our Children Learning? Literacy and Numeracy Across East Africa 2014*. Twaweza East Africa, Nairobi.

Wamalwa, F. M. and Burns, J. (2017). Private Schools and Student Learning Achievements in Kenya. SALDRU Working Paper No. 202.

World Bank (2018). *World Development Report 2018: Learning to Realize Education's Promise*. Washington, DC: World Bank.

Zuilkowski, S. S., Fink, G., Moucheraud, C., and Matafwali, B. (2012). Early childhood education, child development and school readiness: Evidence from zambia. *South African Journal of Childhood Education*, 2(2):20.

Appendix

Appendix Table 1: Grade progression, enrollment, and literacy and numeracy skills, by age

	Highest grade attended	Currently enrolled	Still in preschool	Raw numeracy score	Raw English lit. score	Raw Swahili lit. score	Has 2nd-grade lit./num.
				Panel A: Kenya			
Age 7	1.49	1.00	0.13	3.33	2.00	2.01	0.14
Age 8	2.27	1.00	0.06	4.12	2.43	2.46	0.26
Age 9	3.00	1.00	0.03	4.70	2.82	2.87	0.40
Age 10	3.80	1.00	0.02	5.05	3.10	3.15	0.51
Age 11	4.66	0.99	0.00	5.40	3.36	3.41	0.65
Age 12	5.39	0.99	0.00	5.56	3.53	3.57	0.73
Age 13	6.17	0.99	0.00	5.70	3.66	3.70	0.80
Age 14	6.83	0.98	0.00	5.75	3.75	3.78	0.84
Age 15	7.42	0.97	0.00	5.78	3.80	3.83	0.88
Age 16	8.15	0.95	0.00	5.81	3.84	3.85	0.90
				Panel B: Tanzania			
Age 8	1.73	0.98	0.04	3.09	1.01	1.64	0.10
Age 9	2.43	0.98	0.02	3.71	1.29	2.06	0.18
Age 10	3.19	0.98	0.01	4.17	1.58	2.41	0.27
Age 11	3.96	0.97	0.01	4.62	1.87	2.75	0.39
Age 12	4.73	0.96	0.00	4.88	2.11	2.97	0.46
Age 13	5.55	0.94	0.00	5.12	2.38	3.20	0.56
Age 14	6.19	0.91	0.00	5.30	2.65	3.35	0.64
Age 15	6.71	0.86	0.00	5.39	2.81	3.44	0.69
Age 16	7.28	0.82	0.00	5.47	2.99	3.50	0.73

Notes: The table reports means of the variables indicated in the column heads across children of the age indicated in rows. Raw numeracy scores range from 0 to 6. Raw English and Swahili literacy scores range from 0 to 4. For definitions of all other variables, see the notes to Table 1 and Online Appendix A. Note that the figures reported here can differ from the enrollment statistics reported in Section 2 of the paper because those statistics are based on the unrestricted sample of children in the Uwezo data.

Appendix Table 2: Reported preschool attendance by cohort and wave

Wave:	2011	2012	2013	2014
	Panel A: Kenya			
Cohort				
1997			0.83	
1998			0.84	0.83
1999			0.84	0.84
2000			0.83	0.84
2001			0.84	0.85
2002			0.85	0.84
2003			0.84	0.85
2004			0.85	0.84
2005			0.85	0.85
2006			0.87	0.85
2007				0.85
	Panel B: Tanzania			
Cohort				
1995	0.57			
1996	0.59	0.58		
1997	0.59	0.58	0.60	
1998	0.61	0.60	0.61	
1999	0.63	0.61	0.62	
2000	0.65	0.64	0.64	
2001	0.66	0.65	0.63	
2002	0.69	0.66	0.65	
2003	0.69	0.67	0.64	
2004		0.68	0.65	
2005			0.64	

Notes: The table shows the fractions of children reporting to have attended preschool by country, cohort, and Uwezo survey wave. The sample is restricted to districts that were visited in all waves of the Uwezo survey. We disregard the 2014 wave in Tanzania because only a subsample of districts were sampled (45 districts versus more than 120 districts in the three previous waves). In the raw data, we observe level shifts in preschool attendance rates for all cohorts between some of the waves, likely because the question asking about preschool attendance changed. The table therefore shows regression-adjusted attendance rates after taking out wave fixed effects.

Appendix Table 3: Results for the full sample vs. the sample of households with within variation in preschool attendance

	Highest grade attended			Composite test score		
	Main sample	Sub-sample	p (equal effects)	Main sample	Sub-sample	p (equal effects)
			Panel A: Kenya			
Effect of preschool						
at ages 7-9	-0.336***	-0.423***	0.121	0.042***	0.023	0.670
	(0.021)	(0.051)		(0.015)	(0.041)	
at ages 10-12	-0.123***	-0.090*	0.547	0.114***	0.165***	0.283
	(0.021)	(0.049)		(0.015)	(0.041)	
at ages 13-16	-0.023	-0.020	0.967	0.125***	0.111***	0.785
	(0.021)	(0.047)		(0.015)	(0.039)	
Observations	218,728	7,229		218,134	7,141	
			Panel B: Tanzania			
Effect of preschool						
at ages 8-9	-0.212***	-0.147***	0.069	0.108***	0.093***	0.589
	(0.013)	(0.033)		(0.008)	(0.021)	
at ages 10-12	-0.074***	-0.088***	0.625	0.113***	0.111***	0.889
	(0.011)	(0.023)		(0.006)	(0.015)	
at ages 13-16	0.096***	0.083***	0.602	0.081***	0.120***	0.034
	(0.010)	(0.022)		(0.006)	(0.014)	
Observations	284,396	29,981		288,084	30,427	
			Controls included in panels A and B			
Age × cohort effects	Yes	Yes		Yes	Yes	
Socio-demographics	Yes	Yes		Yes	Yes	
Early-life conditions	Yes	Yes		Yes	Yes	
Household fixed effects	Yes	Yes		Yes	Yes	

Notes: The table shows estimates of the impact of preschool attendance on highest grade attended and the composite test score, separately for the full sample used in the main analysis and for the restricted sample of households with within variation in preschool attendance. Columns 1 and 4 re-estimate the main specifications of Tables 3 and 4. Columns 2 and 5 present the corresponding estimates for the restricted sample. Columns 3 and 6 present p values from Wald tests of the hypothesis that the age-group specific effects in the two columns to the left are equal. These cross-specification tests are conducted using seemingly unrelated regression using Stata's -suest- command. Due to computational constraints, rather than estimating the household fixed effects models directly, we first demean the data using -xtdata-, and then estimate regressions that exclude the household fixed effects on these data. This results in the same coefficient estimates, but lower standard errors. This in turn increases the chance of rejecting the null of equal coefficients in the Wald tests, which means that we err on the conservative side in our interpretation of the p values.

Online Appendix A: Data Appendix

The Uwezo surveys: sampling and test design

Uwezo, which means 'capability' in Kiswahili, is a non-governmental organization that aims to improve competencies in literacy and numeracy among school-aged children in East Africa. Since 2009, Uwezo has conducted annual assessments of the basic literacy and numeracy skills of children in Kenya, Tanzania, and Uganda. The assessments are administered as part of repeated cross-sectional household surveys, which also collect information on a variety of child and household characteristics and education outcomes. Households are selected in a two-stage sampling design: first, in each census district of each country, 30 enumeration areas (which typically correspond to one or several villages) are sampled with probability proportional to size; then, 20 households in each of these enumeration areas are randomly selected to participate in the survey.[25] The resulting sample is representative at both the national and the district level. Weights which reflect this sampling design and which implement a number of ex-post corrections are provided with the data; we use these weights throughout our analysis.[26]

In participating households, all children aged 6-16 (7-16 in Tanzania) are assessed on core literacy and numeracy competencies that should be achieved after two years of schooling according to the national curriculum. Two separate literacy tests in English and Swahili measure the following four competencies in order of rising difficulty: (1) recognition of letters, (2) recognition of words, (3) reading a paragraph, and (4) reading a short story. The numeracy test measures the following six competencies in order of rising difficulty: (1) counting (the number of objects on a show card), (2) recognition of numbers, (3) rank ordering of numbers, (4) addition, (5) subtraction, and (6) multiplication.[27] For each assessment, there are several test

[25] A few districts were excluded in some rounds of the survey due to security concerns. In 2014, Tanzania selected households from a random subsample of districts only.

[26] Unfortunately, the weights included with the Tanzanian data over-emphasize the importance of observations in 2014. Specifically, as reported in the previous footnote, only a random subsample of districts was surveyed in 2014, and this wave correspondingly includes less than a third of the observations compared to any previous wave. Nevertheless, the weights in 2014 add up to about 125% of the weights in all previous waves. We attempt to correct for this irregularity by re-scaling the 2014 weights at the district level, using the relative importance of each district in the 2013 wave as a scaling factor. Our results are however robust to using the original weights, not using any weights at all, or dropping the 2014 wave for Tanzania altogether.

[27] In Kenya, children who master multiplication are also assessed on their division skills. We ignore this seventh, higher competency here in order to ensure comparability of test scores across Kenya and Tanzania.

booklets in order to prevent children within the same household from copying each other's answers. A child's score on each test equals the highest competency level achieved, with a zero indicating that she did not even master the simplest skill assessed.

Variable definitions

Household identifier. The data contain a household identifier, which we use to construct household-level variables such as number of children and wealth. Because polygamy is common in some communities, a few households contain children from different mothers. For each child, we observe his/her mother's age and education, which we use to construct a unique mother identifier. Our within-household specifications are based on this more conservative mother identifier rather than the household identifier, even though in practice this makes little difference.

Socio-demographic characteristics. We define a child's cohort as Uwezo survey wave minus age. Mother's education is recorded differently between countries and survey waves; we make this variable comparable by collapsing it into two categories: no education and at least some primary education. To construct the index of current household wealth, we follow Schady et al. (2015) and aggregate the following dwelling characteristics and assets using the first principal component: wall materials, source of lighting, tv, radio, computer (only Kenya), mobile phone, bicycle, motorbike, and motor vehicle. We compute this index separately for each country and normalize it to have mean zero and standard deviation one.

The rural indicator describes the location of the enumeration area. For Kenya, this variable is not included with the publicly available data, but we were able to obtain it directly from Uwezo. For Tanzania, the variable is included in the publicly available data for the 2011 and 2012 survey waves; as we were not able to obtain the variable for the 2013 and 2014 waves, it is missing for children observed in these years.[28]

Early-life economic conditions. We construct two proxies for district-level economic conditions using external satellite data on night lights and rainfall. For Kenya, district definitions in the Uwezo data are based on the 2009 census. For Tanzania, we create a crosswalk which maps districts in the Uwezo data to districts in the 2002

[28] As usual in survey data, there are some missing values also in other control variables. In order not to unnecessarily reduce sample size, we impute missing values at the sample mean and include separate dummies for missing values on each control variable in all of our regressions.

census. We use GIS census district boundary files from IPUMS International to compute summary statistics for our two proxies for each district and year.[29]

We obtain the night lights data from the Defense Meteorological Satellite Program's Operational Linescan System (DMSP-OLS).[30] The data provide yearly measurements of average light density at a fine geographical level, with light density ranging from 0 to 63. For a detailed description of these data, we refer to Henderson et al. (2012). Our rainfall measures are derived from the Climate Hazards group Infrared Precipitation with Stations (CHIRPS) data.[31] These data provide annual measures of precipitation since 1981; for details, see Funk et al. (2015).

From the satellite data, we construct a variable measuring average log night lights and indicators for positive and negative rainfall shocks at each age before school entry (ages 0-5 in Kenya and ages 0-6 in Tanzania). In line with recent literature (e.g. Shah and Steinberg, 2017), we define rainfall shocks as precipitation above the 80th percentile and below the 20th percentile of the long-term district mean.

Preschool attendance. Recent waves of the Uwezo survey ask respondents whether they ever attended preschool and whether they are currently still enrolled in preschool.[32] From the answers to these two question, we construct our key explanatory variable as an indicator which takes value 1 if a child ever attended preschool and 0 otherwise. In the 2013 and 2014 waves, we moreover have information on length of attendance in years. As a few respondents indicate lengths of attendance far beyond the usual, we winsorize this variable at 3 years in Kenya and 2 years in Tanzania (i.e. at the maximum "normal" length according to the national education system).

Outcome variables. Our first main outcome is the highest grade of school attended. Children who are currently enrolled in preschool are coded as having zero grades attended. Children who are currently in school report the grade they are attend-

[29]The district boundary files for the Kenyan 2009 census and the Tanzanian 2002 census are available here: https://international.ipums.org/international/ gis_yrspecific_2nd.shtml.

[30]We use the Average Visible, Stable Lights, and Cloud Free Coverages series, which is available here: https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html.

[31]We use the CHIRPS-2.0 global annual yearly data series, which is available here: ftp://ftp.chg.ucsb.edu/pub/org/chg/products/CHIRPS-2.0/global_annual/tifs/.

[32]The exact questions asked differ slightly across countries and waves. In the 2013 and 2014 waves for Kenya, respondents were asked to indicate whether the child currently attends preschool, with a separate question asking them "How many years of preschool did the child attend?" The 2013 and 2014 waves in Tanzania similarly asked respondents to indicate whether the child currently attends preschool and "If attended, for how many years?" The 2011 and 2012 waves of the survey in Tanzania instead asked "Did you attend preschool (nursery) before joining primary school?"

ing. Children who have dropped out of school report the grade during which they dropped out; for them, the highest grade attended equals the dropout grade. We winsorize the resulting variable such that children can be ahead at most two grades; for example, a 10 year-old child can have attended at most grade six in Kenya and grade five in Tanzania.

The second main outcome variable is the composite test score. We construct this score by first standardizing the English, Swahili, and numeracy scores by country, Uwezo survey wave, age, and test booklet to have mean zero and standard deviation one. In a second step, we then average these standardized scores for each student and normalize the resulting composite again to obtain the score used in the regressions.

In auxiliary regressions, we also use a number of further outcomes. These include an indicator for current enrollment, which takes value 1 if the child reports to be currently enrolled in preschool or school and 0 otherwise. We also construct an indicator for achieving second-grade literacy and numeracy, which takes value 1 if a child achieves the highest competency level in the numeracy test and at least one of the two literacy tests and 0 otherwise.

Sample selection

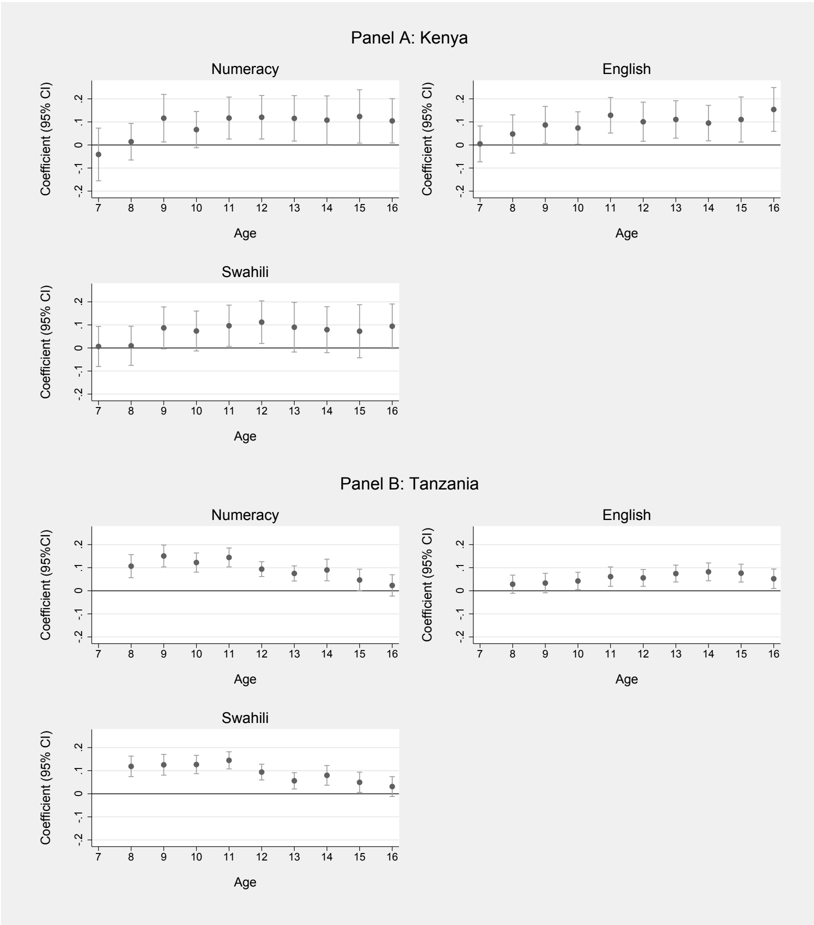
We use data from all available waves of the Uwezo surveys with information on preschool attendance. These are the 2013 and 2014 waves in Kenya and the four waves conducted between 2011 and 2014 in Tanzania. We decided to drop Uganda from the analysis because the only survey with national scope and which collected information on preschool attendance there was fielded in 2013, and information on preschool attendance is missing for 49% of respondents in the corresponding data.

We restrict our attention to children aged 7 and above (8 and above) in Kenya (Tanzania) because some younger children were still of preschool age at the time of the survey. In order to ensure that we focus on comparable siblings in our within-household analysis, we also drop from the sample any children who report never to have enrolled in preschool or school. Our final sample comprises more than half a million children with information on preschool attendance and at least one of the two main outcomes described above. Note that because a few children are observed with only one of these outcomes, observation numbers in regression tables vary slightly.[33]

[33] All results are robust to focusing on a slightly smaller sample of children observed with both outcomes.

Online Appendix B: Figures and Tables

Online Appendix Figure 1: Preschool attendance and numeracy, English, and Swahili skills, by age



Notes: The figure plots coefficient estimates and 95% confidence intervals from regressions of numeracy, English, and Swahili scores on preschool attendance. Scores are standardized by country and age to have mean zero and standard deviation one. The indicator for preschool attendance is interacted with age dummies, and the figure shows the estimated effect of preschool attendance separately for each age. Specifications are otherwise equal to the household fixed effects regressions reported in column 4 of Table 4.

Online Appendix Table 1: Further robustness checks

	Only siblings born ≤5 years apart		No sampling weights	
	Highest grade attended (1)	Composite test score (2)	Highest grade attended (3)	Composite test score (4)
	Panel A: Kenya			
Attended preschool				
7-9 years old	-0.257***	0.067	-0.377***	0.033
	(0.064)	(0.052)	(0.038)	(0.034)
10-12 years old	-0.083	0.136***	-0.146***	0.124***
	(0.068)	(0.051)	(0.034)	(0.035)
13-16 years old	-0.038	0.123**	-0.026	0.148***
	(0.076)	(0.054)	(0.042)	(0.039)
Observations	158,132	158,408	218,728	218,134
	Panel B: Tanzania			
Attended preschool				
8-9 years old	-0.207***	0.101***	-0.207***	0.103***
	(0.027)	(0.021)	(0.024)	(0.014)
10-12 years old	-0.081**	0.120***	-0.074***	0.112***
	(0.031)	(0.016)	(0.022)	(0.012)
13-16 years old	0.058*	0.077***	0.053**	0.058***
	(0.032)	(0.018)	(0.024)	(0.013)
Observations	235,967	238,852	284,396	288,084

Notes: In columns 1 and 2, the sample is restricted to families with children born at most 5 years apart. Columns 3 and 4 report estimates from regressions that do not use the sampling weights provided with the data. Standard errors in parentheses are clustered at the district level. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

Chapter 11

Cultural Gender Norms and Neighbourhood Exposure: Impacts on the Gender Gap in Math

Abstract

This paper estimates the effect of cultural gender norms on the gender gap in math, and explores whether this effect is mitigated by municipality gender equality. I use high-quality Swedish administrative data on the results of national standardised math tests. To separate the effect of cultural gender norms from formal institutions, I estimate the effect of mothers' source-country gender norms on the gender gap in math for second-generation immigrants. By contrasting the outcomes of opposite-sex siblings, I show that the sibling gender gap in math increases with mothers' adherence to traditional gender norms; such that girls with more gender-traditional mothers perform worse relative to their brothers. To investigate whether the cultural gender norm effect can be mitigated by municipality gender equality, I exploit a refugee placement policy to obtain random variation in municipality characteristics. I show that municipality gender equality can almost completely mitigate the negative cultural norm effect. Taken together, my results imply that while cultural gender norms play an important role for the gender gap in math, they are not immune to the effects of neighbourhood exposure.

Keywords: cultural gender norms, math gender gap, epidemiological approach, refugee placement policy, sibling fixed effects
JEL Classifications: I21, I24, J15, J16, Z13

1 Introduction

Girls and boys differ in terms of their educational achievement. In most cases, a gender gap in education implies one that favours girls, as girls outperform boys along most educational dimensions (DiPrete and Buchmann, 2013). However, one exception is math. Girls systematically perform worse than boys on math tests, particularly at the top of the performance distribution (Bedard and Cho, 2010; Pope and Sydnor, 2010). The gender gap in math has been shown to correlate strongly with gender equality and norms regarding women's role in society, which suggests that social forces may, at least in part, be driving the differential performance of boys and girls (Guiso et al., 2008; Pope and Sydnor, 2010; Nollenberger et al., 2016).[1] In addition, Rodríguez-Planas and Nollenberger (2018) show that this relationship is not driven by math-specific norms, but rather, by general gender stereotypes about girls and educational outcomes.

One channel through which norms could affect educational outcomes is the formation of identities. Norms shape our expectations regarding the social group we identify with, which in turn affect our beliefs of what we are capable of, and our preferences for what we spend time on. Akerlof and Kranton (2000) develop a theoretical framework in which individuals choose to identify with different social categories, and derive utility from complying with the behaviour prescribed by these chosen categories.[2] In this framework, as well as in empirical studies testing its implications, identity concerns have a significant impact on educational outcomes (Akerlof and Kranton, 2002; Schüller, 2015).

But from where do we perceive the norms that, through identity formation, shape our behaviour and impact our economic outcomes? Several studies find that culture and historical traditions have a significant impact on both the attitudes and the economic outcomes of individuals today (see e.g. Fernández, 2011; Alesina et al., 2013; Nollenberger et al., 2016; Finseraas and Kotsadam, 2017; Rodríguez-Planas and Nollenberger, 2018; Dahl et al., 2020), which demonstrates that cultural norms are important determinants of our behaviour. However, another strand of the literature documents significant behavioural impacts of neighbourhood exposure and

[1] Geographical variation in the math gender gap indicates that it is not driven solely by innate ability differences between boys and girls (Bedard and Cho, 2010). In addition, the gender gap in math does not exist at the point of school entry, but rather emerges over time when children are socialized into school (Fryer Jr and Levitt, 2010).

[2] One salient category is gender, where everyone is assigned to either being a "man" or a "woman", and where there are prescribed attributes and behaviours that are considered "manly" or "womanly". This way, gender identity changes the pay-off of different actions, and the choice of identity may impact our economic outcomes, including educational performance.

peer effects (see e.g. Chetty et al., 2016; Chetty and Hendren, 2018; Dahl et al., 2014; Olivetti et al., 2018), which indicates that our on-going exposure to institutions, peers, and other surrounding factors may also be an important determinant of our behaviour.[3] Taken together, a large literature shows that both family culture and neighbourhood characteristics affect our economic, including educational, outcomes. It is likely that these two channels do not operate independently of each other, however, there is limited empirical evidence combining the two channels.

This paper investigates the interaction between cultural norms and neighbourhood characteristics, in the context of their impact on the gender gap in math. Specifically, I ask two research questions: first, is there an effect of cultural gender norms on the gender gap in math, and second, to what extent can this effect be mitigated by surrounding neighbourhood gender equality? This way, I explore the effect of cultural norms on individuals' behaviour (measured by their educational outcomes), and I investigate how this effect interacts with exposure to neighbourhood characteristics and peers.

There are two main empirical challenges associated with estimating the effect of cultural gender norms. First, norms are correlated with institutional settings, which likely have an impact on educational outcomes. Second, parent's cultural norms are not randomly assigned, but instead are likely correlated with other parental characteristics that affect the educational outcomes of their children. Therefore, to answer the first research question, it is crucial to disentangle the impact of cultural norms from the impact of both formal institutions and of parental characteristics.

To isolate cultural gender norms from formal institutions, I estimate the effect of gender norms in mothers' countries of origin on the gender gap in math among second-generation immigrants. Second-generation immigrants, in this context, are all born and raised in Sweden and encounter the same formal institutions but potentially differ in their cultural heritage. Assuming that mothers transmit norms to their children, and that these norms differ systematically depending on the mother's source country, second-generation immigrants provide the ideal experiment to isol-

[3]See Kranton (2016) for more examples of determinants of identify formation. The distinction between cultural values and neighbourhood exposure is similar to the framework developed by Bisin and Verdier (2011), who contrast vertical and horizontal transmission of norms. Vertically transmission of norms occurs within the family, from parents to children, and happens if parents believe that their children will benefit from certain cultural traits. Horizontal transmission denotes the socialisation of norms that takes place within a community context, where norms are transmitted from peers and surroundings. However, the culture/neighbourhood distinction noted above is broader compared to that of Bisin and Verdier (2011), as the cultural (i.e. "vertical") channel includes also parents' peers and networks (sharing the same cultural beliefs), and the neighbourhood (i.e. "horizontal") channel includes not only the effect of norms, but also that of more formal institutions.

ate the effect of cultural norms from the effect of formal institutions.[4]

To account for the fact that gender norms are not randomly assigned to mothers, and therefore likely correlate with unobserved maternal characteristics, I follow Finseraas and Kotsadam (2017) and compare the gender gap in math only between opposite-sex siblings in a sibling fixed effects model. The sibling fixed effects control for everything that affects both siblings equally, including everything that correlates with source-country norms but that is unrelated to gender. By construction, the variation that remains is the *gender-specific* component of the cultural norms that affects opposite-sex siblings differently, i.e. gender norms.

To answer the second research question, I investigate the extent to which the cultural gender norm effect can be mitigated by municipality gender equality. The main empirical challenge of estimating the effect of neighbourhood characteristics is that there is selection in where people choose to live. Families will choose to reside in places that have certain desirable characteristics, and, in doing so, they themselves contribute to these characteristics. To account for this selection, as well as to obtain exogenous variation in municipality characteristics, I exploit a refugee placement policy. Under this policy, government officials assigned asylum-seeking immigrants their initial location of residence. As these immigrants were not free to chose where they would be placed, their initial location of residence is independent of unobserved individual characteristics.

I rely on high-quality Swedish administrative data on the universe of ninth-grade students who took the national standardised math test between 2004–2012. To proxy cultural gender norms and neighbourhood gender equality, I use female-over-male labour force participation rates, of both the immigrant mother's source country and of her assigned municipality of residence.

I show that mothers' cultural gender norms increase the sibling gender gap in math, such that girls with more gender-traditional mothers perform worse relative to their brothers. A one-standard-deviation increase in cultural gender norms (i.e. towards more traditional norms) increases the size of the math gender gap by 56%, in favour of boys. In addition, I find similar effects for final marks in other school subjects, which shows that the results are not driven by math-specific cultural norms, but rather by general gender stereotypes about girls and educational outcomes.

However, I also show that municipality gender equality can almost completely mitigate the negative cultural gender norm effect. This result suggests that even

[4]This method, refereed to as the epidemiological approach, aims to identify the effect of culture through variation in outcomes among individuals who share economic and formal institutions, but who potentially differ in their social beliefs (see e.g. Fernández, 2011).

though the sibling gender gap in math increases with mothers' adherence to traditional gender norms, this increase is smaller for siblings whose mothers were placed in more gender-equal municipalities. Taken together, my results show that while cultural gender norms play an important role for the gender gap in math, they are not immune to the influence of surrounding characteristics.

The first research question of my study relates to the growing body of literature on gender norms and educational outcomes.[5] Several studies show that girls' relative educational performance correlates positively with gender equality, both across countries and US states (Guiso et al., 2008; Fryer Jr and Levitt, 2010; Pope and Sydnor, 2010). One important issue here is the risk of reverse causality, as these studies cannot determine whether girls perform better because of increased gender equality or whether high-performing girls grow up to themselves contribute to increased gender equality. Nollenberger et al. (2016) and Rodríguez-Planas and Nollenberger (2018) account for this reverse causality, as they estimate the effect of mothers' cultural gender norms on the gender gap in math, reading and science, by using PISA data on the test scores of second-generation immigrants. The authors find that the educational gender gaps between boys and girls vary with mothers' cultural gender norms, such that girls with more gender-traditional mothers perform worse in all three subjects relative to boys with mothers from the same source country.

Dossi et al. (2019) show that girls who grow up in families that exhibit a preference for boys perform worse on standardised math tests, when compared to girls growing up in other families. Furthermore, the authors use survey data to show that mothers' attitudes regarding women's role in society correlate with girls' performance in mathematics, but not with boys'. Dahl et al. (2020) show that birthright citizenship for immigrant girls lowers their life satisfaction and self-esteem, a result they argue is due to the conflicting identities of German citizenship and parents' traditional cultural norms. They do not find the same effect for boys, which indicates that girls are pushed comparatively harder by their parents to conform with traditional gender norms, whereas boys are allowed to take advantage of the citizenship opportunities. Finally, using an estimation strategy similar to mine, Aldén and Neuman (2019) show that cultural gender norms affect the probability of girls choosing STEM, or other male-dominated fields, as their major in high school or university.[6]

[5]In addition, my paper contributes to the literature on the effects of source country culture. Using the standard epidemiological approach, studies show that source country culture affects gender roles, women's work and fertility, social trust, political regulation, domestic violence, corruption, migration etc. See Fernández (2011) for a literature review.

[6]Unrelated to gender norms, but related to the gender gap in educational outcomes, Figlio et al. (2019) show that family disadvantage is disproportionally detrimental to the educational outcomes of

The second research question of my study relates to the literature on the effects of neighbourhood characteristics on children's educational outcomes. Two well-identified papers by Chetty et al. (2016) and Chetty and Hendren (2018) show that neighbourhoods shape children's earnings, college attendance rates, marriage and fertility patterns, and that the effect increases linearly with time of exposure. In addition, they show that boys and girls are affected differently in neighbourhoods that are particularly beneficial for either gender.[7] Damm and Dustmann (2014) exploit random variation in neighbourhood exposure, caused by a refugee placement policy in Denmark, to show that children who were placed in high-crime neighbourhoods are more likely to themselves commit crimes. They find that social interaction is the key channel through which this neighbourhood exposure effect operates.

My study also relates to the literature on peer effects, where several papers show that the behaviour of peers affect individuals' economic outcomes, such as labour market decisions (Maurin and Moschion, 2009; Dahl et al., 2014; Olivetti et al., 2018). Finally, my study contributes to the literature that use the Swedish refugee placement policy to evaluate neighbourhood exposure effects, such as, for example, Edin et al. (2003) who show that living in ethnic enclaves improves immigrants' labour market outcomes, and Åslund et al. (2011) who show that immigrants' school performance is increasing with the number of highly educated individuals of shared ethnicity residing in the same neighbourhood.

To the best of my knowledge, my study is the first to estimate the interaction between neighbourhood characteristics and cultural norms. Thus, a novel and important contribution of my paper is that I merge the literatures on cultural norms, neighbourhood exposure and educational outcomes. Moreover, my paper is also the first to establish a causal link between cultural gender norms and the gender gap in math. Because my study focuses on the gender gap between opposite-sex siblings, I am able to control for many potentially worrisome causes of variation that previous papers have not been able to control for, which allows me to more credibly isolate the effect of cultural gender norms.

boys relative to girls, and that this result is robust to specifications within neighbourhoods, schools and families.

[7] Chetty et al. (2016) estimate the effect of neighbourhood exposure using variation caused by the Moving To Opportunity experiment. They find that only children who moved when young experience positive effects of moving to a low-poverty neighbourhood. Chetty and Hendren (2018) exploit variation in age of children when families move, finding that the outcomes of children who move converge towards the outcomes of the people in the new neighbourhood at a rate of 4% per year of exposure. Furthermore, they show that neighbourhoods matter because of differences in childhood environment rather than because of differences in labour market conditions.

The remainder of the paper is organised as follows. Section 2 provides some institutional background on the Swedish marking system and the refugee placement policy. Section 3 describes the data, after which Section 4 outlines the empirical strategy. Section 5 presents the results for both research questions, and further investigates heterogeneous effects and alternative outcomes. Section 6 ensures that my results are robust. Finally, Section 7 concludes this paper.

2 Institutional background

2.1 The Swedish marking system and the national standardised tests

Ninth grade is the last year of mandatory schooling in Sweden, and students at this level are between 15 and 16 years old. Students take 16 different subjects and receive a final mark for each subject. Marks are goal-oriented and not relative. The marks during this time period are IG (fail), G (pass), VG (pass with distinction) and MVG (pass with special distinction), which correspond to 0, 10, 15 and 20 points, all of which count towards the final total mark. The final total mark is the sum of all subject marks, with a maximum of 320 points.

The national standardised tests are issued in math, Swedish, English, social science and natural science. These tests, which students take during the spring semester of the ninth grade, are nationally standardised and mandatory for all students. The tests are developed to give all students an equal opportunity to demonstrate their knowledge, and act as means of supporting the teacher in making marking more fair. Students receive a mark on each subject test, which weigh heavily on the final mark for that subject.

The national standardised test in math consists of about 40 'pass-level' questions and about 35 'pass with distinction-level' questions. Students receive a test result, which is the sum of all correctly answered questions on each level, and an overall mark for the test. One significant benefit of using the math test score as the outcome variable is that it contains more, and continuous, variation than the final marks, which can only take one of four values.

2.2 The Swedish refugee placement policy

The refugee placement policy was introduced in 1985 and lasted until 1994. Its aim was to relieve pressure on the larger cities in times of large refugee inflows. By placing the asylum seekers in those municipalities with the most suitable reception characteristics, the government hoped to speed up the integration process.[8]

During the policy period the Immigration Board assigned asylum seekers their initial municipality of residence. Initially 60 municipalities participated in the policy, but due to the increased inflow of asylum seekers during the late 1980's the number of municipalities involved increased until 277 of the total 284 municipalities were participating. The strictest application of the policy was implemented during 1987 to 1991, when the assignment rate was almost 90% and the asylum seekers had very little ability to influence where they were assigned. For this reason, I focus my analysis on the refugees who arrived during the period 1987–1991.

The asylum seekers were placed in refugee centres while they waited for a decision from the immigration authorities. Mean duration before receiving a residence permit was between three and twelve months. While the assignment process did take the asylum seekers' preferences into account, most applied for residence in Stockholm, Gothenburg or Malmö. The Swedish housing market was booming at the time, and housing opportunities in these locations were very scarce; in practice this meant that the immigrants had very little influence over their placement.

When the number of applicants exceeded the number of available slots, municipal officers had the opportunity to choose which asylum seekers would be offered a residence permit. However, all selection was based on observable characteristics, as there was no interaction between the asylum seekers and the municipal officers. Priority was given to individuals who had attained higher education and to those who spoke the language as some of the residing immigrants. Furthermore, housing availability was dependent on family size. Edin et al. (2003) argues that municipality assignment can be viewed as random, conditional on observable immigrant characteristics.

However, Nekby and Pettersson-Lidbom (2017) identify some important empirical challenges regarding the use of the refugee placement policy as a proxy for random placement. They argue that the municipalities had more to say in the placement process than what had been previously understood. Refugees could not be placed without consent from the municipality, and the placement within a municipality was carried out by local municipality immigration agencies. For example, the municipality of Bollnäs did not place immigrants in the first available apartment, but rather, waited until housing opened up in areas with few social problems.

[8]The information in this section is obtained from Edin et al. (2003) and Åslund et al. (2011), who in turn base their information regarding the practical implementation of the policy on interviews with Immigration Board placement officials.

These types of decisions were not nationally standardised and could differ sharply between municipalities. As a result, there is a risk that municipality characteristics may correlate with placement and subsequent treatment, and therefore also with the outcomes of the asylum seekers. I address these empirical challenges in Section 4.2.

Following initial assignment, there were no restrictions on moving to a different location, provided that the immigrants themselves could find housing. Leaving the assigned municipality did not affect the welfare payments; the main cost of moving was delayed enrolment in Swedish courses.

3 Data

I rely on data from The Swedish Interdisciplinary Panel, which is administered by the Centre for Economic Demography. This is a two-generational dataset consisting of merged administrative registers. Using unique parental identifiers, I identify the parents and siblings of each individual in the data. The data contain information on the national standardised test results from years 2004–2012. Students take the tests when they are 16 years old; thus, my study contains cohorts born between 1987 and 1996. I restrict my sample to second-generation immigrants, which I define as individuals born in Sweden whose mothers were born in another country.[9]

Table 1 presents the summary statistics for three sample specifications: the total population of students taking the national standardised math test (1), the sample used for the first research question (henceforth the RQ1 sample) (2), the sample used for the second research question (henceforth the RQ2 sample) (3). The sample in column (1) acts merely as a reference point for the other two samples.

[9] As a robustness check I use a sample of second-generation immigrants defined by having a foreign-born father; the results can be found in Appendix Tables A4–A6.

Table 1: Summary statistics

	Full sample	RQ1 sample	RQ2 sample
2nd generation	0.23		
	(0.42)		
Girl	0.49	0.50	0.50
	(0.50)	(0.50)	(0.50)
Cohort	1991.76	1992.11	1992.51
	(2.43)	(2.40)	(2.19)
9th grade total mark	212.01	205.35	199.92
	(59.24)	(61.12)	(60.26)
Math test score	36.20	33.03	29.84
	(15.65)	(15.54)	(14.93)
Math gender gap (std)	-0.01	-0.07	-0.07
	(0.00)	(0.01)	(0.01)
Family size	1.77	2.46	2.46
	(0.72)	(0.70)	(0.64)
HH income > median	0.50	0.32	0.17
	(0.50)	(0.47)	(0.38)
Mothers: post-secondary ed.	0.40	0.29	0.19
	(0.49)	(0.45)	(0.39)
Fathers: post-secondary ed.	0.32	0.27	0.22
	(0.47)	(0.44)	(0.41)
Mother's age	44.82	44.31	43.63
	(5.04)	(4.95)	(4.85)
Father's age	47.76	48.19	48.75
	(6.01)	(6.08)	(6.40)
Mother's norms (1-FLFP)	0.16	0.38	0.56
	(0.13)	(0.24)	(0.24)
Father's norms (1-FLFP)	0.17	0.35	0.56
	(0.14)	(0.27)	(0.25)
Municipality FLFP	0.91	0.92	0.93
	(0.05)	(0.05)	(0.07)
Observations	885,745	24,632	3,926

Notes: The table reports means and standard deviations of the main variables of my study. Column (1) includes all ninth grade students taking the national standardised tests between 2004–2012. Column (2) includes siblings whose mothers are immigrants, and column (3) includes siblings whose mothers were subjected to the refugee placement policy.

3.1 Sample restrictions for RQ1

As my empirical strategy is based on comparisons between opposite-sex siblings, I implicitly restrict my sample to only families who have at least one child of each gender. I match each child to all of their biological opposite-sex siblings. This means that an individual with more than one opposite-sex sibling will be observed more than once in the data. All siblings are required to be born in Sweden and share both biological parents.[10]

I observe the source country for mothers born in 20 different countries, Sweden excluded.[11] Column (4) of Table 2 lists the frequency of children with mothers born in the various source countries of my sample. As I define a child as a second-generation immigrant based on their mother's country of birth, the children of my sample all have mothers born in one of the 20 foreign countries I can observe, but their fathers, however, can be missing from the data or born anywhere, including Sweden or one of the unobserved foreign countries. In total, 51% of the children have parents born in the same foreign country, and 78% of those with parents from different countries have a father born in Sweden.

Column (2) of Table 1 presents the summary statistics for this sample of second-generation immigrants. There are 24,632 observations, consisting of 12,316 unique sibling pairs. Compared to the full population of ninth grade students in column (1), the second-generation immigrants receive lower final marks and score lower on the math tests. In addition, they are more likely to live in households with below-median income, their parents have less education, and, by construction, their family size is larger than the national average.

3.2 Sample restrictions for RQ2

To answer the second research question, I impose additional sample restrictions, which creates a subset of the sample for the first research question. I require mothers to have immigrated to Sweden sometime between 1987–1991, the years in which the refugee placement policy was active. Unfortunately I cannot observe refugee status directly, but, following Edin et al. (2003) and Åslund et al. (2011), I assign refugee status using country of birth. The Swedish Migration Board lists the countries from which Sweden received asylum seekers between 1984–1999 (Migrationsverket, 1999).

[10] For some children I do not have any information on the identity of the father; if that is the case, I compare the siblings who all have the same mother and a missing father.

[11] For reasons of anonymity, countries from which Sweden received few immigrants are aggregated up to regions by Statistics Sweden. To be a part of my final sample, the mother must have been born in a country I can identify in the data.

Anyone immigrating from one of the countries mentioned in this list is defined as an asylum seeker in my sample. I exclude mothers whose spouses immigrated before them, as family immigrants were not subjected to the placement reform. This definition of refugee status leaves me with 10 different source countries. Column (5) of Table 2 shows the frequency of children with mothers coming from these countries as asylum seekers. In total, 80% of the children in this sub-sample have parents born in the same foreign country, and 35% of those with parents from different countries have a father born in Sweden.

Column (3) of Table 1 presents the summary statistics for the children whose mothers were subjected to the refugee placement policy. The total number of observations is 3,926, comprising 1,963 unique sibling pairs. Compared to the full population of ninth grade students in column (1), and to the second-generation sample in column (2), the children with mothers affected by the refugee placement policy receive lower final marks and math scores. Their parents have lower education levels and household income, and they immigrated from source countries with more traditional gender norms.

3.3 Dependent variable: standardized math test score

The main outcome variable is students' test scores on the national standardised math test. The raw test scores range between 0 to 75; about 40 of these are pass-level points and about 35 are pass with distinction-level points. As the tests differ slightly over the years, I standardise the scores by year to obtain a mean of zero and a standard deviation of one.[12]

Figure 1 illustrates the average gender gap in math (defined as girls' scores - boys' scores) for each year of my study. The top panel contains the RQ1 sample of second-generation siblings, while the bottom panel contains the RQ2 sample of second-generation siblings whose mothers were affected by the placement policy. For both samples, girls' test scores are about 7% of a standard deviation lower than boys' test scores, which is stable over the years.

[12] I standardise test scores for each sub-sample separately.

Figure 1: Average gender gap in math over time

Notes: The figure plots the average gender gap between boys and girls for each year between 2004 - 2012. The average gender gap is defined as (girls' average scores - boys' average scores).

3.4 Independent variable: cultural gender norms

I proxy the source-country gender norms using female-to-male labour force participation ratios, defined in in the following way:

$$FLFP_c = \frac{FemaleLFP_c}{MaleLFP_c} \tag{1}$$

This way, I only measure the relative labour market changes for women as compared to men, and account for any general labour market shifts that affect men and women equally. The measure varies only by source-country *c*. I use modelled ILO estimates from the World Bank Indicators Database (World Bank, 2019). I construct an average of the female-to-male labour force participation ratios over the years 1990–1993, in order to avoid capturing temporary fluctuations.

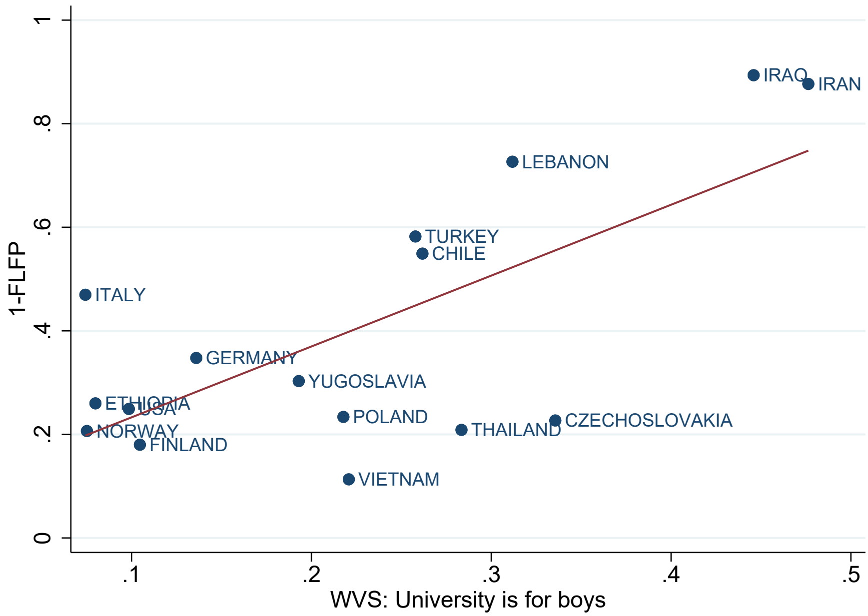
It is not obvious when in time I should measure labour force participation rates in order to best reflect mothers' current norms. On the one hand, norms measured at the time when the mother emigrated from her source country may best capture the norms she carries with her. On the other hand, norms measured around the birth of the second generation may be a better proxy for the norms transmitted from the first to the second generation. Due to data availability, my earliest consistent measure of labour force participation is from 1990 and onwards, which corresponds to around the time of the second generation's birth (as well as the time of emigration for the asylum-seeking mothers). This way, I follow common practice in the literature to measure fairly contemporary norms (Fernandez and Fogli, 2009; Rodríguez-Planas and Nollenberger, 2018; Finseraas and Kotsadam, 2017). However, as a robustness check, I obtain earlier data from the International Labour Organisation's ILOSTAT Database for a subset of source countries where this is possible (ILO, 2019). I construct a measure of source-country norms that varies by decade of immigration and show that my results are not sensitive to the time at which norms are measured.

I define the source-country gender norms as $1 - FLFP$, such that the measure increases as the source-country female-to-male labour force participation ratio decreases; thus, the measure increases as gender norms become more traditional. The main benefit of using female-to-male labour force participation ratios to proxy source-country gender norms is that data on gender-specific labour force participation is consistently available both at the source-country level and for Swedish municipalities at time of immigration.

The purpose of the $1-FLFP$ proxy is to capture variation that reflects source-country norms regarding the relative importance of boys' and girls' educational achievements. For this reason, I test the correlation between the $1 - FLFP$ proxy and the source-country level of agreement with the statement *"A university education is more important for a boy than for a girl."*, derived from the World Values Survey. Figure 2 presents the correlation, which is positive and strong, indicating that $1 - FLFP$ is a relevant proxy for source-country norms regarding gender and educational outcomes.[13]

[13]The correlation coefficient is 0.89. Excluding Iran and Iraq decreases the correlation to 0.71, but does not alter my results.

Figure 2: Correlation between 1-FLFP and WVS statement



Notes: The figure plots the correlation between source-country 1-FLFP and level of agreement with the statement "A university education is more important for a boy than for a girl." from the World Values Survey.

Table 2 presents the average $1 - FLFP$, the level of agreement with the statement *"A university education is more important for a boy than for a girl."* from the World Values Survey, and the score on the World Economic Forum Gender Gap Index (GGI), for each source country in my study.[14]

[14]The GGI is commonly used in studies similar to mine. In a robustness check, I define the source-country norms as $1 - GGI$ and ensure that my results are robust to this specification. I derive the source-country GGI score from the World Economic Forum Global Gender Gap report of 2011 (WEF, 2011), as this is the year in which the most of the source countries of my study are included in the report.

Table 2: Included countries: frequency and norm values

Source country	1-FLFP	WVS	1-GGI	RQ1	RQ2
CHILE	0.55	0.26	0.30	990	312
CZECHOSLOVAKIA	0.23	0.34	0.31	144	
DENMARK	0.16		0.22	918	
ETHIOPIA	0.26	0.08	0.39	976	420
FINLAND	0.18	0.11	0.16	6,718	
GERMANY	0.35	0.14	0.24	500	
GREECE	0.47		0.31	142	
IRAN	0.88	0.48	0.41	838	382
IRAQ	0.89	0.45		814	248
ITALY	0.47	0.07	0.32	90	
LEBANON	0.73	0.31	0.4	2,464	886
NORWAY	0.21	0.08	0.16	1,370	
POLAND	0.23	0.22	0.3	1,186	314
SOMALIA	0.71			774	218
THAILAND	0.21	0.28	0.31	408	
THE SOVIET UNION	0.23		0.3	88	
TURKEY	0.58	0.26	0.41	2,586	668
USA	0.25	0.1	0.26	360	
VIETNAM	0.11	0.22	0.33	388	156
YUGOSLAVIA	0.30	0.19	0.3	2,878	322
Observations	24,632	22,710	23,044	24,632	3,926

Notes: The table reports the source-country levels of 1-FLFP and 1-GGI, as well as the country level of agreement with the statement "A university education is more important for a boy than for a girl." from the World Values Survey. Columns (4) and (5) report frequencies of children with mothers from each source country for both sample definitions.

3.5 Independent variable: municipality gender equality

To investigate the mitigation of traditional cultural gender norms by neighbourhood characteristics, I derive a measure of municipality-level gender equality. I proxy municipality gender equality with a similar measure of female-to-male labour force participation ratios as for the source countries:

$$FLFP_{mt} = \frac{FemaleLFP_{mt}}{MaleLFP_{mt}} \tag{2}$$

The measure varies with assigned municipality m and year of immigration t. I obtain information on gender-specific labour force participation rates from the Statistics Sweden Labour Statistics Database (SCB, 2019). To make the mitigation estimates of my study easier to interpret, I define this neighbourhood measure to work in the opposite direction of the cultural norm measure. For this reason, I define the municipality gender equality proxy as $FLFP$, such that it increases as the female-to-male labour force participation ratio increases. I measure municipality gender equality at the time of immigration and within the assigned municipality, as this is the only time it can be assumed to be exogenously given.

4 Empirical strategy

4.1 Do cultural gender norms affect the math gender gap?

To investigate the effect of cultural norms, it is necessary to disentangle the effects of culture from the effect of formal institutions. To do so, I rely on an extended version of the epidemiological approach. The epidemiological approach aims to identify the effect of culture by examining variation in outcomes among individuals who share formal and economic institutions, but who potentially differ in their social beliefs (Fernandez, 2007; Fernandez and Fogli, 2009; Fernández, 2011). Second-generation immigrants provide a useful experimental cohort for such a study. The assumptions behind this strategy are that 1) mothers transmit cultural beliefs to their children, 2) these cultural beliefs vary in a systematic way that reflects the culture of the mother's source country, and 3) individuals growing up in the same country encounter similar economic and formal institutions.

In an ideal experimental setup mothers' cultural norms would be randomly assigned, such that the gender norms are orthogonal to everything unobserved that relates to the childhood environment. Unfortunately, such ideal setups are hard to come by, and a significant drawback of the epidemiological approach is that it cannot account for unobserved factors that may also vary in a systematic way by mothers' source country. For example, different immigrant groups may have different reasons for migrating, they may be more or less likely to live in ethnic enclaves, and they may face different levels and types of discrimination in the migrant country. In addition, family structure may be endogenous to the sex of children, as suggested by Dahl and Moretti (2008), and this endogeneity could correlate with source-country culture.

In lack of the ideal set-up, I follow Finseraas and Kotsadam (2017) and include sibling-pair fixed effects in the empirical epidemiological model. The sibling-pair fixed effect absorbs any variation that is constant across siblings; hence, they control for everything that affects both siblings equally, such as childhood environment, unobserved parental characteristics and endogenous family structure. This way, the sibling-pair fixed effect controls for everything that correlates with source-country cultural norms but that is unrelated to gender, such that the only variation that remains is the component of source-country norms that affects opposite-sex siblings in different ways. Thus, by construction, the model identifies only the effect of any *gender-specific* components of culture, i.e. gender norms, as anything that is not gender-specific will not vary across siblings and will therefore be absorbed by the sibling-pair fixed effect. I estimate the following model:

$$\begin{aligned} MathScore_{ij} = \alpha Girl_{ij} + \beta Girl_{ij} \times Norms_j \\ + \gamma Birthorder_{ij} + Cohort_{ij} + \eta_j + \epsilon_{ij} \end{aligned} \quad (3)$$

where i refers to individuals and j to sibling pairs. α captures the baseline math score difference between opposite-sex siblings. The baseline cultural gender norm measure does not vary within families and will be absorbed by the sibling-pair fixed effect η_j. The coefficient of interest is β, which captures how the sibling gender differences in math performance vary depending on the gender norms of the mother's source-country culture. I control for birth order and cohort fixed effects, and I cluster standard errors at the source-country times cohort level.[15] β identifies the causal effect of cultural gender norms under the identifying assumption that any latent gap in childhood outcomes between brothers and sisters is as good as randomly assigned between families from different source-countries.[16]

[15] Treatment is technically at the source-country level, but too few clusters may decrease standard errors (Angrist and Pischke, 2009). My results are robust to using a wild bootstrap procedure to obtain standard errors clustered at the source-country level.

[16] This assumption would be violated if, for example, source-country culture was correlated with

4.2 Is the effect mitigated by municipality gender equality?

When investigating any causal mitigation effect of neighbourhood characteristics, the ideal setup would be to randomise gender norms among children's mothers and then randomise the types of municipalities in which the children grow up. While, again, the ideal setup is not available, the refugee placement policy described in Section 2.2 offers quasi-random variation in municipality characteristics, as asylum seekers were not free to choose their assigned municipalities.

An important issue regarding the refugee placement policy, which was raised by Nekby and Pettersson-Lidbom (2017), is that placement into municipalities may correlate with how the refugees were treated, and therefore also with the outcomes of the second generation. For example, if some municipalities systematically placed the refugees in areas with fewer social problems, this would most likely imply that the children of those refugees would attend schools of higher quality, which may affect their educational outcomes. If the likelihood of placing refugees in neighbourhoods of "higher quality" correlates with municipality FLFP, this would bias my results, as I would be comparing boys and girls attending different types of schools and living in different types of neighbourhoods. By including sibling-pair fixed effects, I increase the likelihood that my model identifies only the mitigation effect of neighbourhood gender equality, rather than the effect of some unobserved municipality characteristic that correlates with placement, as siblings are placed in the same neighbourhood, school and house, and therefore experience the same special treatment (if such treatment exists). Again, the sibling fixed effect absorbs everything that affects both siblings equally, and I estimate the effect of the *gender-specific* component of municipality characteristics. I estimate the following equation:

$$\begin{aligned} MathScore_{ij} &= \alpha Girl_{ij} + \beta Girl_{ij} \times Norms_j \\ &+ \delta Girl_{ij} \times MunicipalityFLFP_{ij} + \lambda Girl_{ij} \times Norms_j \times MunicipalityFLFP_j \\ &+ \gamma Birthorder_{ij} + Cohort_{ij} + \eta_j + \epsilon_{ij} \end{aligned} \tag{4}$$

where i refers to individuals and j to sibling pairs. α captures the baseline math gender gap between brothers and sisters, and β captures how this gender gap varies with mothers' cultural gender norms. The coefficient of interest is λ, which

in-utero factors that affect opposite-sex foetuses differently. Following Figlio et al. (2019), I address this assumption by assessing whether cultural gender norms can predict neonatal characteristics, such as birth weight. Reassuringly, Appendix Table A1 shows that there is only a weak and negligible relationship between cultural gender norms and the gender gap in birth weight.

measures how the effect of cultural norms on the gender gap in math varies with the gender equality of the mothers' assigned municipality. In other words, λ captures the relative effect for girls growing up with more traditional mothers who were assigned to more gender-equal municipalities, compared to girls whose traditional mothers were assigned to less gender-equal municipalities. The baseline cultural norm and municipality characteristic effects will be absorbed by the sibling-pair fixed effect η_j. I cluster standard errors at assigned municipality times year of immigration.[17]

Another threat to identification is the possibility that asylum seekers were systematically placed in gender-equal municipalities based on the country they migrated from, which would cause correlation between municipality FLFP and the mothers' cultural gender norms. To investigate this, I test whether mothers' cultural gender norms can predict municipality gender equality. Table 3 presents the results of this balance test.[18] The top panel shows the correlation between asylum-seeking mothers' cultural gender norms and the FLFP of the assigned municipality. Reassuringly, the correlation is very small and statistically insignificant, which indicates that the allocation of asylum-seeking future mothers were random with respect to cultural gender norms. The share of municipality residents from the same source country as the refugee (cultural density) and family size predicts municipality FLFP; this is expected as these are characteristics that influenced placement decisions. However, these correlations are not problematic, as all of these differences will be controlled for by the sibling fixed effects.

The middle panel illustrates the relationship when I allow selection, i.e. the correlation between asylum-seeking mothers' cultural gender norms and the FLFP of the municipality they live in when their child is in ninth grade. Similarly, the bottom panel shows the correlation between mothers' cultural gender norms and FLFP of the ninth grade municipality for the full sample of all second-generation immigrants. The correlations shown in the middle and bottom panel are positive and statistically significant, which indicates that families select into municipalities

[17] My results are robust to clustering the standard errors at source country times assigned municipality times year of immigration level.

[18] Column (1) presents the raw correlation between cultural gender norms and municipality gender equality. When municipalities could choose which asylum seekers were allocated to them, priority was given to the more highly educated and those who spoke the language of some of the resident immigrant stock. In addition, family size determined housing availability. Column (2) controls for these relevant placement characteristics, and column (3) adds immigration year fixed effects. Cultural density is calculated as the share of municipality residents with the same nationality. Immigration year fixed effects should be included as the number of participating municipalities increased over time. Finally, column (4) adds the cohort of the mother.

partly based on gender norms and gender equality. If this selection is related to the relative educational outcomes of boys and girls, this would bias my results. Thus, these results highlight the importance of using the refugee placement policy in order to obtain exogenous variation in municipality characteristics.

As previously stated, refugees were not required to stay in their assigned municipality. Following initial assignment, there were no restrictions on moving to a different location, and refugees could move to another municipality if they found housing on their own. For this reason, I am only able to estimate the intention-to-treat effect of the assigned municipality characteristics. However, 56% of the asylum-seeking mothers still lived in their assigned municipality at the time when their children graduated from ninth grade.[19]

[19] As expected, younger and more highly educated individuals are more likely to change municipality. Some measurement error may exist if individuals move within the first year of assignment, as I observe municipality of residence in the end of the year. This issue is investigated thoroughly by Edin et al. (2003), who use a weighting scheme based on the aggregate data on municipality reception of refugees. The weighting does not change their estimates significantly, which suggests that measurement error is not a substantial concern.

Table 3: Balance test for refugee placement policy

	(1)	(2)	(3)	(4)
	Outcome: assigned mun FLFP (RQ2 sample)			
Mother's norms	-0.014	-0.001	-0.000	0.000
	(0.030)	(0.030)	(0.029)	(0.029)
Cultural density		0.122***	0.124***	0.126***
		(0.035)	(0.035)	(0.035)
Mother's education		-0.008	-0.006	-0.010
		(0.013)	(0.014)	(0.015)
Family size		-0.128***	-0.122***	-0.133***
		(0.034)	(0.035)	(0.036)
Mother's cohort				-0.007
				(0.008)
R-squared	0.00	0.03	0.04	0.05
Observations	3,926	3,926	3,926	3,926
	Outcome: current mun FLFP (RQ2 sample)			
Mother's norms	0.079***	0.075**	0.072**	0.073**
	(0.030)	(0.031)	(0.030)	(0.030)
Cultural density		-0.100**	-0.106***	-0.105***
		(0.040)	(0.039)	(0.039)
Mother's education		-0.002	-0.006	-0.007
		(0.014)	(0.014)	(0.015)
Family size		-0.030	-0.023	-0.025
		(0.029)	(0.029)	(0.032)
Mother's cohort				-0.002
				(0.006)
R-squared	0.01	0.02	0.02	0.02
Observations	3,926	3,926	3,926	3,926
	Outcome: current mun FLFP (RQ1 sample)			
Mother's norms	0.163***	0.183***	0.184***	0.196***
	(0.015)	(0.014)	(0.013)	(0.013)
Cultural density		0.028	0.028	0.028
		(0.032)	(0.032)	(0.032)
Mother's education		0.033***	0.033***	0.025***
		(0.005)	(0.005)	(0.005)
Family size		-0.077***	-0.077***	-0.080***
		(0.015)	(0.015)	(0.015)
Mother's cohort				-0.020***
				(0.002)
R-squared	0.03	0.03	0.05	0.06
Observations	24,544	24,544	24,544	24,544
	Indicators			
Immigration year FE	No	No	Yes	Yes

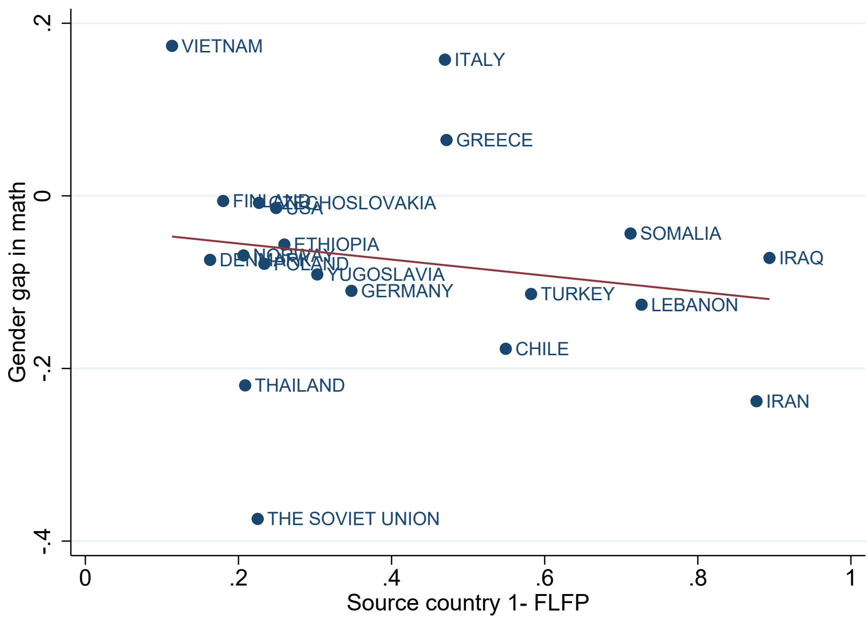
Notes: The table reports the correlation between municipality FLFP and mothers' source-country gender norms and individual characteristics. The dependent variable in the top panel is the FLFP of the assigned municipality. The dependent variable in the second and third panel is the FLFP of the municipality the family lives in when the child graduates ninth grade. All parental characteristics (for the RQ2 sample) are measured at the time of immigration. Cultural density is the share of municipality residents from the same source country as the mother who is being placed there. Standard errors are clustered at assigned municipality × immigration year level. * p<0.1, ** p<0.05, *** p<0.01.

5 Results

5.1 Do cultural gender norms affect the math gender gap?

Figure 3 illustrates the descriptive correlation between mothers' cultural gender norms and the average gender gap in math performance of the second generation. The fitted relationship is negative, indicating that the gender gap in math increases as cultural gender norms become more traditional.

Figure 3: Correlation between 1-FLFP and average gender gap in math



Notes: The figure plots the correlation between parents' source-country 1-FLFP and average gender gap in math (girls' average scores - boys' average scores) for second-generation immigrants.

Table 4 shows the results for the first research question: the effect of mothers' cultural gender norms on the gender gap in math. All columns control for birth order and cohort fixed effects. Column (2) adds an indicator of whether both parents are immigrants, as well as source-country and municipality of residence fixed effects. Column (3) adds controls for family size, household income, education level and cohort of the mothers. Column (4) contains the preferred model specification, which includes sibling-pair fixed effects.

Table 4 demonstrates that girls score lower on the math tests compared to boys; the baseline math score sibling gender gap is about 7% of a standard deviation.[20] The estimates in column (1) show that both boys' and girls' math performance decrease as mothers' gender norms become more traditional, but more so for the girls' than for the boys'. The interaction effect of the girl indicator and the cultural gender norms measure shows that the gender gap in math increases with mothers' cultural gender norms, such that girls whose mothers have more traditional gender norms fall behind their brothers by an additional 4% of a standard deviation. Compared to the baseline sibling gender gap, a one-standard-deviation increase in cultural gender norms (about 0.24, which corresponds to going from Norway to Italy, or from Greece to Somalia) increases the gender gap in math by 56%.

Table 4: Effect of mothers' gender norms on the math gender gap

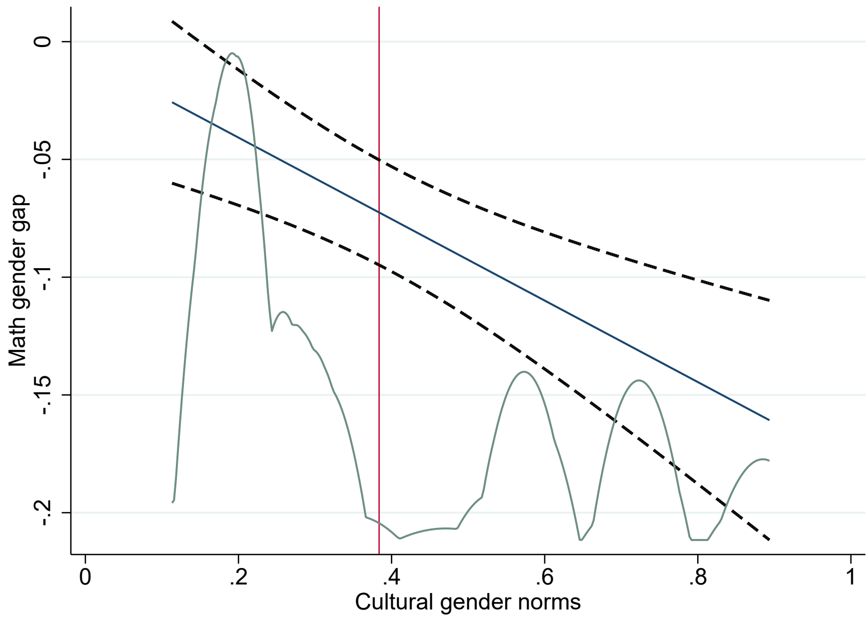
	(1)	(2)	(3)	(4)
Girl	-0.073***	-0.073***	-0.073***	-0.073***
	(0.014)	(0.013)	(0.013)	(0.011)
Mother's norms	-0.147***			
	(0.020)			
Girl × mother's norms	-0.042***	-0.041***	-0.042***	-0.041***
	(0.014)	(0.013)	(0.013)	(0.011)
R-squared	0.03	0.12	0.18	0.71
Observations	24,632	24,632	24,632	24,632
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Source country & mun FE	No	Yes	Yes	No
Mother characteristics	No	No	Yes	No
Sibling FE	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24). Mothers' characteristics are cohort, education level, household income and family size. Standard errors in parentheses are clustered at the source-country × cohort level. * p<0.1, ** p<0.05, *** p<0.01.

[20] The girl indicator captures the gender gap in math for siblings whose mothers' gender norm proxy takes the value zero (i.e. complete gender equality). However, the mean gender gap in math for the entire sample is very similar, at about 7%.

Figure 4 graphically illustrates the interaction effect from the preferred specification, with mothers' cultural gender norms on the x-axis and the sibling math gender gap on the y-axis. The distribution of the gender norms proxy is plotted in the background. The interaction effect is negative; going from the most gender-equal source country to the least gender-equal source country corresponds to an increase in the math gender gap of about 15% of a standard deviation.

Figure 4: Effect of mothers' gender norms on the math gender gap



Notes: The figure plots the relative effect of mothers' more traditional cultural gender norms (girl × source-country 1-FLFP) on the gender gap in math. The dotted lines depict the 95% confidence interval. The kernel density distribution of the source-country 1-FLFP is plotted in the background. The vertical line indicates the mean of 1-FLFP.

Compared to previous literature, my estimates are of the same sign, but smaller in magnitude. Nollenberger et al. (2016) and Rodríguez-Planas and Nollenberger (2018) find that a one-standard-deviation increase in the source-country World Economic Forum Gender Gap Index is associated with a reduction in the math gender gap, and an increase in the reading and science gender gap, of about 30% of a standard deviation of the aggregate gender gaps. After being recalculated to match their definition, my effect sizes correspond to an effect of 7.6% of the standard deviation of the aggregate math score gender gap.[21] A smaller effect size is in line with the results of Finseraas and Kotsadam (2017), who find that the effect of cultural gender norms on the gender gap in employment among second-generation immigrants in Norway is about 50% of the size of corresponding US estimates.

5.2 Is the effect mitigated by municipality gender equality?

Table 5 presents the results for the second research question: whether municipality gender equality can mitigate the negative effect of cultural gender norms. Column (1) replicates the preferred specification of the first research question, and estimates the effect of gender norms on the sibling gender gap in math for the subset of children whose mothers were asylum seekers under the placement policy. Column (2) presents the mitigation effect, controlling only for birth order and cohort fixed effects. Column (3) adds an indicator of whether both parents are immigrants, source-country fixed effects, and assigned municipality times immigration year linear trends. Column (4) adds controls for predetermined maternal characteristics that may have influenced placement. Finally, column (5) contains the preferred specification, which controls for birth order, cohort and sibling-pair fixed effects.

Similar to the results for the first research question, the baseline gender gap in math is about 7% of a standard deviation. Mothers' cultural gender norms have a negative effect, such that more traditional gender norms increase the sibling gender gap in math in favour of boys. Most importantly for the second research question, the mitigation effect of municipality gender equality is both positive and statistically significant. This result suggests that while girls who have more traditional mothers do relatively worse in math, this negative effect is mitigated for those girls whose mothers were assigned to more gender-equal municipalities. Increasing the assigned municipality FLFP by one standard deviation (0.07) leads to a mitigation effect of about 5% of a standard deviation of the math score, which corresponds to 82% of the negative cultural norm effect. Thus, municipality gender equality can almost completely mitigate the negative effect of mothers' traditional gender norms.

[21]Following Rodríguez-Planas and Nollenberger (2018), I calculate effect size as $\frac{NormSD \times \beta}{MathSD} = \frac{0.24 \times 0.041}{0.13} = 0.076$.

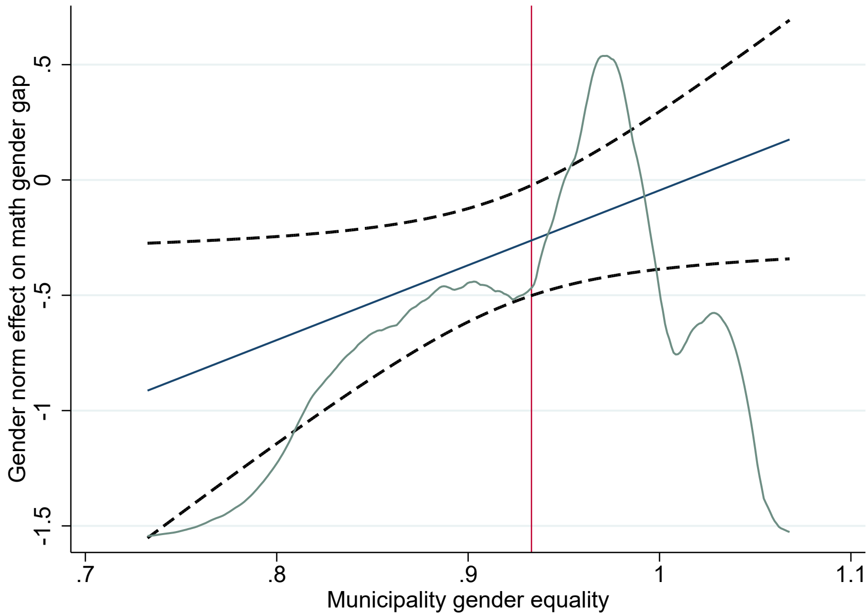
Table 5: Municipality mitigation effect

	(1)	(2)	(3)	(4)	(5)
Girl	-0.074***	-0.071***	-0.070***	-0.070***	-0.073***
	(0.026)	(0.027)	(0.027)	(0.027)	(0.027)
Girl × mother's norms	-0.064**	-0.063**	-0.064**	-0.064**	-0.062**
	(0.029)	(0.029)	(0.029)	(0.029)	(0.029)
Girl × mun FLFP		-0.039	-0.042	-0.041	-0.040
		(0.027)	(0.027)	(0.027)	(0.026)
Girl × mother's norms × mun FLFP		0.051**	0.049*	0.050*	0.051**
		(0.026)	(0.026)	(0.026)	(0.025)
R-squared	0.71	0.02	0.32	0.33	0.71
Observations	3,926	3,926	3,926	3,926	3,926
	Indicators				
Birth order	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
Source country FE	No	No	Yes	Yes	No
Mun. × Im.Year FE	No	No	Yes	Yes	No
Mother characteristics	No	No	No	Yes	No
Sibling FE	Yes	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. Municipality FLFP is measured at the assigned municipality, in the year in which the mother immigrated to Sweden. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24), and a one standard deviation increase in municipality FLFP (0.07). Mothers' characteristics are cohort, education level, family size and cultural density of assigned municipality, all of which are measured at time of immigration. Standard errors are clustered at the assigned municipality × immigration year level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Figure 5 illustrates the graphical representation of the mitigation effect, with municipality gender equality ($FLFP$) on the x-axis and the effect of more gender-traditional mothers on the y-axis. The distribution of assigned municipality $FLFP$ is plotted in the background. Going from the least gender-equal municipality to the most gender-equal municipality completely mitigates the negative effect of mothers' traditional gender norms.

Figure 5: Municipality mitigation effect



Notes: The figure plots the mitigation effect (girl × source-country 1-FLFP × municipality FLFP) on the gender gap in math. The dotted lines depict the 95% confidence interval. The kernel density distribution of the municipality-level FLFP is plotted in the background, and the vertical line indicates the mean municipality FLFP. Municipality FLFP is measured at the assigned municipality, in the year in which the mother immigrated to Sweden.

My estimates are comparable to those of Chetty and Hendren (2018), who estimate the effects of neighbourhood exposure on children's outcomes. The authors find that the outcomes of children who move to a new neighbourhood converge to the outcomes of the residents of the new neighbourhood at a rate of 4% per year of exposure. Extrapolating this result implies that children who move to a new neighbourhood at birth and stay there until they are 20 years old would pick up about 80% of the difference in residents' outcomes between their origin and destination neighbourhood. My estimates correspond to an assimilation of mothers' cultural norms to the neighbourhood setting, from before the birth of the child to the age of 16, and are of a similar magnitude to the extrapolated estimates of Chetty and Hendren (2018).

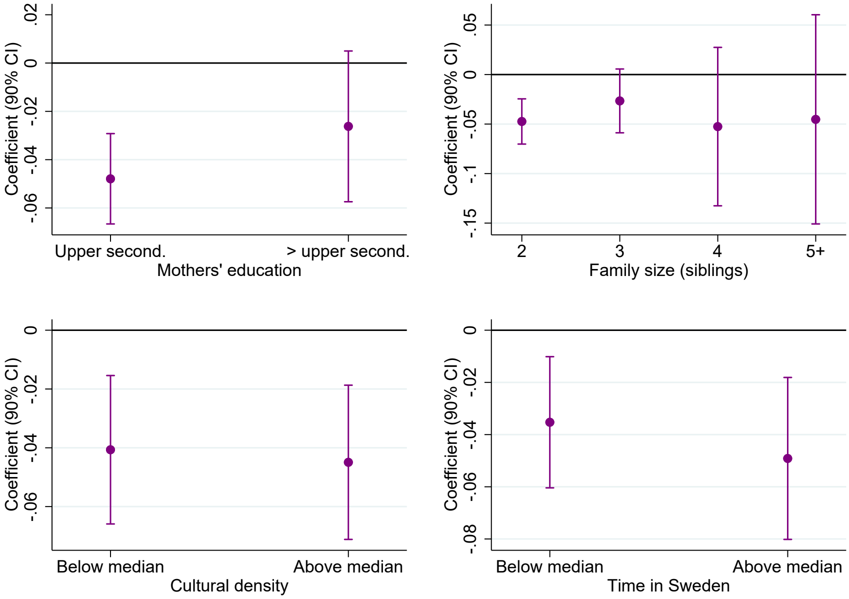
5.3 Heterogeneity

Both the effect of cultural gender norms and the mitigation effect may differ depending on the characteristics of the mother and of the assigned municipality. The assimilation of a mother's cultural values to the neighbourhood setting could be influenced by her education level — for example, because learning Swedish and finding employment may be easier for the more highly educated — or by how long it has been since she migrated to Sweden. The effects could also be influenced by municipality characteristics, such as cultural density (i.e. fraction of residents of the same ethnicity) or by the type of labour market she encounters upon immigration.

Furthermore, the relative math performance of girls may be influenced by the number of siblings she has. As my sample consists of girls that are matched to all their opposite-sex siblings, more weight is given to those girls who have many brothers. Hence, there is a risk of over-estimating the effect of culture if mothers with more traditional cultural gender norms tend to have more children, and if girls in larger families tend to perform worse in math due to, for example, increased responsibilities at home.

Figure 6 presents the heterogeneous effects of mothers' cultural gender norms for the RQ1 sample. The effect appears stronger for less educated mothers, although the estimates do not differ significantly from each other. Likewise, I find no heterogeneous effects by family size, cultural density, or time since the mother's immigration.

Figure 6: Heterogeneity in mothers' gender norm effect



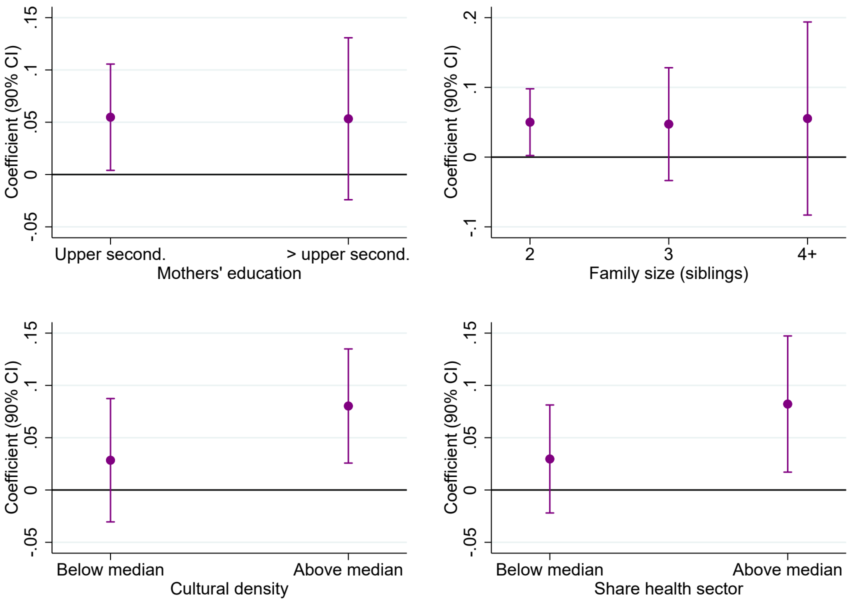
Notes: The figure plots the effect of mothers' gender norms on the gender gap in math, when the effect is allowed to differ by mothers' education level (attended up to, or more, than upper secondary education), family size (measured as the number of siblings a girl has), cultural density of the municipality of residence (measured by share of same-ethnicity residents, below or above the median share), and how long it has been since the mother immigrated to Sweden (below or above the median number of years (22)).The figure plots 90% confidence intervals. The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1.

Figure 7 presents the heterogeneous mitigation effects for the RQ2 sample. The figure shows that the mitigation effect is stronger for siblings with mothers who were assigned to municipalities in which the health sector is responsible for an above-median share of the total municipality employment.[22] Likewise, the mitigation effect is stronger for siblings with mothers who were assigned to municipalities with an above-median share of same-ethnicity residents. Although the estimates are not statistically different from each other, which may be an issue of power, these results are in line with the results of Åslund and Rooth (2007), who find that initial labour market exposure has lasting effects for arriving refugees, and Edin et al. (2003), who

[22] Almost 20% of the working population of previously asylum-seeking mothers are working within the health sector; hence, the relative size of this sector may be important for future labour market outcomes and assimilation.

show that living in an ethnic enclave improves labour market outcomes. I find no heterogeneous effects by mothers' education level or family size.

Figure 7: Heterogeneity in mitigation effect



Notes: The figure plots the mitigation effect of municipality gender equality, when the effect is allowed to differ by mothers' education level (attended up to, or more, than upper secondary education), family size (measured as the number of siblings a girl has), cultural density of the assigned municipality (measured by share of same-ethnicity residents, below or above median share), or relative share of the health care sector as compared to total industry composition in the assigned municipality. The figure plots 90% confidence intervals. The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1.

Finally, the effects may be heterogeneous over the math performance distribution. It is reasonable to believe that high-achieving students are not affected in the same way as struggling students are. However, as math performance is an outcome of gender norms, and because the sibling gender gap mechanically depends on the siblings' absolute performance, this is a difficult issue to investigate. As an example, I allocate the siblings into quintiles depending on their mean math performance, and allow the effect of cultural gender norms and the mitigation effect to differ depending on where on the performance distribution the siblings are located. Appendix Figure A1 shows that the results appear to be driven by students in the lower and middle part of the performance distribution. However, these heterogeneous effects may also be mechanically driven by the amount of variation in the sibling gender gap in each quintile, and should therefore be interpreted with caution.

5.4 Alternative outcomes: final marks

Rodríguez-Planas and Nollenberger (2018) show that the effect of parents' cultural gender norms affect not only the gender gap in math, but also the gender gaps in reading and science. This result implies that the cultural norms affecting girls' educational outcomes are not only math-specific, but rather reflect general stereotypes about gender and educational outcomes. To investigate whether this is also true for my setting, I estimate the effect of mothers' cultural gender norms on the sibling gender gap in final marks in Swedish, English, math and the total ninth grade mark.

Table 6 presents the effect of mothers' cultural gender norms on the sibling gender gaps in these final marks. Sisters outperform their brothers in all subjects, and they get a higher total ninth grade mark.[23] The effect of mothers' cultural gender norms is negative for all outcomes, such that girls with more traditional mothers perform worse relative to their brothers. These results confirm the findings of Rodríguez-Planas and Nollenberger (2018), and show that cultural gender norms affect not only girls' test scores in math, but also their relative school performance more generally. Thus, the norms at play are not math-specific, but rather gender stereotypes about education in general.

Table 7 shows the mitigation effect of assigned municipality gender equality for the final subject and total marks. The mitigation effect is positive for math, Swedish and the total ninth grade mark, but is only statistically significant for math. However, given the small sample size and lack of variation in outcomes (subject marks can only take the values 0, 10, 15 and 20) the model may not have enough power to estimate coefficients with precision.

[23]Interestingly, even though girls score lower on the national standardised tests in math, they still seem to get a slightly higher final mark.

Table 6: Effect of mothers' gender norms on gender gap in final marks

	(Math)	(Eng)	(Swe)	(Total)
Girl	0.021*	0.170***	0.524***	0.332***
	(0.011)	(0.010)	(0.014)	(0.009)
Girl × mother's norms	-0.034***	-0.007	-0.089***	-0.026***
	(0.011)	(0.009)	(0.016)	(0.010)
R-squared	0.68	0.70	0.68	0.73
Observations	24,443	24,443	19,725	24,632
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Sibling FE	Yes	Yes	Yes	Yes

Notes: The dependent variables are the student's mark on the national standardised tests in math (1), English (2), Swedish (3) and the total final ninth grade mark (4). All outcomes are standardised to have a mean of 0 and standard deviation of 1. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24). All regressions are controlling for birth order, and sibling-pair and cohort fixed effects. Standard errors in parentheses are clustered at the source country × cohort level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 7: Municipality mitigation effect on gender gap in final marks

	(Math)	(Eng)	(Swe)	(Total)
Girl	-0.002	0.190***	0.469***	0.331***
	(0.029)	(0.029)	(0.039)	(0.028)
Girl × mother's norms	-0.038	-0.015	-0.027	-0.017
	(0.028)	(0.025)	(0.036)	(0.025)
Girl × mun FLFP	-0.031	-0.009	-0.010	-0.017
	(0.028)	(0.025)	(0.039)	(0.026)
Girl × mother's norms × mun FLFP	0.049*	-0.002	0.006	0.013
	(0.026)	(0.024)	(0.033)	(0.024)
R-squared	0.68	0.72	0.67	0.73
Observations	3,909	3,909	2,499	3,926
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Sibling FE	Yes	Yes	Yes	Yes

Notes: The dependent variables are the student's mark on the national standardised tests in math (1), English (2), Swedish (3) and the total final ninth grade mark (4). All outcomes are standardised to have a mean of 0 and standard deviation of 1. Municipality FLFP is measured at the assigned municipality, in the year in which the mother immigrated to Sweden. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24), and a one-standard-deviation increase in municipality FLFP (0.07). All regressions are controlling for birth order, as well as sibling-pair and cohort fixed effects. Standard errors in parentheses are clustered at the assigned municipality × immigration year level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

6 Robustness checks

The aim of my model is to isolate any gender-specific components of culture, i.e. gender norms. By construction, anything that is not gender-specific will not vary across siblings and will therefore be absorbed by the sibling-pair fixed effect, such that only the gender-specific component remains. However, one possible concern is that cultural gender norms could vary systematically with other, unobserved source-country characteristics that could affect boys and girls in different ways. If that were true, I may not be estimating the effect of gender norms, but rather the effect of some other unobserved and source-country-specific characteristic. To mitigate this concern, I estimate a model in which I investigate the relative sibling impact of source-country placebo norms. I construct an index of desired child characteristics from the World Values Survey that I expect to relate to math performance but that do not necessarily have a gender-specific impact.[24] Table 8 presents the results from this placebo test. The placebo norm is positively related to math performance, but has no gender-specific impact on girls relative to boys. This result demonstrates that not all source-country norms have a gender-specific component that differs between boys and girls; accordingly, this supports the argument that my main model identifies the effect of gender norms rather than some other unobservable characteristic that varies by source-country.

Likewise, to test the assumption that my model for the second research question picks up only the effect of neighbourhood gender equality, rather than some other unobserved factor that varies systematically by municipality, I investigate whether the parental norm effect is mitigated by other municipality characteristics, such as average population income. Living in municipalities with higher average income could have a positive effect on children's educational performance, as (for example) high-income municipalities may be able to provide schools of higher quality. However, municipality income may not have a gender-specific component such that it can mitigate the cultural gender norm effect. Table 9 shows that municipality income does not mitigate the effect of mothers' gender norms. This result supports the argument that my main model captures the effect of surrounding gender equality, and not some other unobserved municipality characteristic.

[24] The index consists of country-level agreement regarding the various characteristics that parents wish to foster in their children. The characteristics are independence, responsibility and hard work, which are chosen based on their intuitive, and empirically proven, positive impact on math performance.

Table 8: Placebo: effect of other source-country norms

	(1)	(2)	(3)	(4)
Girl	-0.074***	-0.073***	-0.074***	-0.073***
	(0.016)	(0.015)	(0.015)	(0.015)
Mother's placebo norms	0.092***			
	(0.026)			
Girl × mother's placebo norms	0.015	0.015	0.015	0.015
	(0.016)	(0.016)	(0.015)	(0.012)
R-squared	0.01	0.13	0.18	0.71
Observations	22,710	22,710	22,710	22,710
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Source country & mun FE	No	Yes	Yes	No
Mother characteristics	No	No	Yes	No
Sibling FE	No	No	No	Yes

Notes: The dependent variable is the student's tests score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. The placebo norm is constructed as an index of various desired child characteristics, derived using source-country level of agreement with traits one would like children to learn from the World Values Survey. The characteristics are Responsibility, Independence and Hard work. These characteristics are chosen because they are intuitively, and empirically, positively related to math performance. The effects are estimated for a one-standard-deviation increase in the placebo norm index (0.07). Mothers' characteristics are cohort, education level, household income and family size. Standard errors are clustered at the source-country × cohort level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 9: Placebo: mitigation effect of other municipality characteristics

	(1)	(2)	(3)	(4)
Girl	-0.072***	-0.071***	-0.071***	-0.074***
	(0.027)	(0.027)	(0.027)	(0.026)
Mun income	0.012			
	(0.025)			
Girl × mother's norms	-0.066**	-0.067**	-0.067**	-0.065**
	(0.029)	(0.029)	(0.029)	(0.029)
Girls × mun income	-0.016	-0.018	-0.018	-0.018
	(0.023)	(0.023)	(0.023)	(0.023)
Girl × mother's norms × mun income	-0.004	-0.003	-0.003	-0.003
	(0.028)	(0.028)	(0.028)	(0.028)
R-squared	0.02	0.32	0.33	0.71
Observations	3,926	3,926	3,926	3,926
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Source country FE	No	Yes	Yes	No
Mun. × Im.Year FE	No	Yes	Yes	No
Mother characteristics	No	No	Yes	No
Sibling FE	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised tests in math, standardised to have a mean of 0 and standard deviation of 1. Municipality income is measured as average labour income of the assigned municipality, in the year in which the mother immigrated to Sweden. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24), and a one-standard-deviation increase in municipality average income (SEK 15,097). Mothers' characteristics are cohort, education level, household income, family size and cultural density of assigned municipality. Standard errors are clustered at the assigned municipality × immigration year level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

My results are not sensitive to relaxing the opposite-sex sibling restriction, which keeps also families without one child of each gender in the sample. Table 10 (11) show the results for the first (second) research question, without restricting the sample to only opposite-sex siblings.[25] For both research questions, the sample size triples, but the results remain of similar magnitude. Neither are the results sensitive to measuring source-country norms at time of immigration. For a sub-sample of the data, I have information on gender-specific labour force participation rates from earlier than 1990, and I define a mother's cultural gender norms as the source country $1 - FLFP$ during the decade in which the she migrated to Sweden. Table 12 shows that the effects of mothers' cultural gender norms on the sibling gender gap in math remain of similar sign and magnitude.[26]

Table 10: Mothers' gender norm effect: without opposite-sex sibling restriction

	(1)	(2)	(3)	(4)
Girl	-0.069***	-0.064***	-0.061***	-0.068***
	(0.008)	(0.007)	(0.007)	(0.011)
Mother's norms	-0.100***			
	(0.021)			
Girl × mother's norms	-0.031***	-0.030***	-0.032***	-0.043***
	(0.008)	(0.008)	(0.007)	(0.012)
R-squared	0.01	0.09	0.16	0.70
Observations	79,264	79,264	79,264	79,264
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	No
Source country & mun FE	No	Yes	Yes	No
Mother characteristics	No	No	Yes	No
Mother FE	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised tests in math, standardised to have a mean of 0 and standard deviation of 1. The sample consists of second-generation immigrants with a foreign-born mother, but without restricting the sample to families with at least one child of each gender. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24). Mothers' characteristics are cohort, education level, household income and family size. Standard errors in parentheses are clustered at the source country × cohort level. * p<0.1, ** p<0.05, *** p<0.01.

[25] In this setup mother fixed effects are included in the last columns instead of sibling-pair fixed effects, which implies that girls are compared to the average of all their brothers' test scores.

[26] For the second research question, norms are already measured around the time of migration, as all the asylum-seeking mothers arrived during 1987–1991.

Table 11: Mitigation effect: without opposite-sex sibling restriction

	(1)	(2)	(3)	(4)	(5)
Girl	-0.075***	-0.067***	-0.053***	-0.053***	-0.072***
	(0.024)	(0.018)	(0.017)	(0.017)	(0.024)
Girl × mother's norms	-0.066**	-0.039*	-0.034*	-0.034*	-0.066**
	(0.028)	(0.020)	(0.021)	(0.020)	(0.029)
Girl × mun FLFP		-0.041**	-0.048***	-0.048***	-0.040
		(0.017)	(0.017)	(0.017)	(0.025)
Girl × mother's norms × mun FLFP		0.034*	0.053**	0.050**	0.055**
		(0.020)	(0.022)	(0.021)	(0.026)
R-squared	0.69	0.01	0.17	0.19	0.69
Observations	12,303	12,303	12,303	12,303	12,303
	Indicators				
Birth order	Yes	Yes	Yes	Yes	Yes
Cohort FE	No	Yes	Yes	Yes	No
Source country FE	No	No	Yes	Yes	No
Mun. × Im.Year FE	No	No	Yes	Yes	No
Mother characteristics	No	No	No	Yes	No
Mother FE	Yes	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised tests in math, standardised to have a mean of 0 and standard deviation of 1. The sample consists of second-generation immigrants with an asylum-seeking mother who had been affected by the refugee placement policy, but without restricting the sample to families with at least one child of each gender. Municipality FLFP is measured at the assigned municipality, in the year in which the mother immigrated to Sweden. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24), and a one-standard-deviation increase in municipality FLFP (0.07). Mothers' characteristics are cohort, education level, family size and cultural density of assigned municipality, which are all measured at time of immigration. Standard errors are clustered at the assigned municipality × immigration year level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 12: Effect of source-country norm measured at time of immigration

	(1)	(2)	(3)	(4)
Girl	-0.080***	-0.079***	-0.080***	-0.079***
	(0.015)	(0.015)	(0.015)	(0.015)
Mother's norms	-0.087	-0.125**	-0.051	
	(0.059)	(0.054)	(0.044)	
Girl × mother's norms	-0.029**	-0.028**	-0.028*	-0.029**
	(0.014)	(0.014)	(0.014)	(0.014)
R-squared	0.01	0.12	0.18	0.71
Observations	16,548	16,548	16,548	16,548
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Source country & mun FE	No	Yes	Yes	No
Mother characteristics	No	No	Yes	No
Sibling FE	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. Source-country norm is measured as 1-FLFP of the decade in which the mother immigrated to Sweden. For Somalia, labour force participation data from this period is not available; moreover, for Czechoslovakia and Lebanon I observe very few mothers with a migration date. Thus, these three source countries are excluded from the regressions. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.37). Mothers' characteristics are cohort, education level, household income and family size. Standard errors in parentheses are clustered at the source country × immigration decade level. * p<0.1, ** p<0.05, *** p<0.01.

Table 13 and 14 show the results for a battery of robustness checks. Column (1) shows that my results are robust to dropping the countries contributing to the largest immigrant flows (Finland and Yugoslavia) and to dropping the largest and most urban municipalities in Sweden (Stockholm, Gothenburg and Malmö). Column (2) shows that the results are similar when I estimate my model on a sample of children whose parents are both immigrants. To increase the likelihood of the siblings being exposed to the same family environment, I also restrict my sample to siblings who are at most five years apart in age. Column (3) shows that the results are robust to this specification.

I also replicate my main results using alternative measures of both cultural gender norms and municipality gender equality. The first alternative measure of cultural gender norms is source-country level of agreement with the statement *"A university education is more important for a boy than for a girl."* derived from the World Values Survey. The second alternative measure of source-country gender norms is the World Economic Forum's Gender Gap Index.[27] The alternative municipality gender equality measure is average female-to-male wage ratio, measured within the

[27] The WEF GGI is used by Nollenberger et al. (2016), Rodríguez-Planas and Nollenberger (2018) and Rodríguez-Planas and Sanz-de Galdeano (2019). I define parents' cultural gender norms as 1 -

assigned municipality in the immigration year. As before, the measure of cultural norms increases with gender-traditional norms, while the municipality measure increases with gender equality. Columns (4) and (5) show that both the results for the cultural norm effect and the municipality mitigation are robust to using these alternative definitions of gender norms and gender equality.

Figlio et al. (2019) find that boys from families with lower socio-economic status perform worse relative to their sisters. If family disadvantage correlates with cultural gender norms, there is a risk that my study may capture a relative effect of family disadvantage for boys rather than a relative effect of gender norms for girls. In column (6), I estimate my main model with controls for a gender-specific effect of mothers' socio-economic status; reassuringly, it does not affect my estimates.[28] I address the concern of selective migration by controlling for a gender-specific effect of geographical distance between mothers' source-country and Sweden, which, as column (7) shows, also does not alter my estimates.[29]

GGI.

[28] I measure socio-economic status using the mothers' education levels.

[29] Migrating parents may be disproportionally drawn from the lower or upper part of the source country distribution for preferences regarding girls' relative educational achievement. If this selection varies systematically with the relative female labour force participation of the source country, it could bias my estimates. Belot and Hatton (2012) study the selection of migration among OECD countries and show that the selection on skills is more negative in proximate source countries. If the selection process on preferences against girls' educational achievements is similar, this would lead me to overstate the effect of cultural gender norms. Furthermore, Appendix Figure A2 addresses the concern that migrants might leave their source country precisely because they do not agree with the gender norms there; however, such selection would lead me to understate the effect of cultural gender norms and is therefore not a threat to my study. Appendix Figure A2 shows that the correlation between source country $1 - FLFP$ and the $1 - FLFP$ of the first generation in Sweden is positive.

Table 13: Other robustness checks: mothers' gender norm effect

	Drop FI + YU	Both immigrants	Max 5 years	WVS Uni	GGI norm	SES	Distance
Girl	-0.085***	-0.092***	-0.073***	-0.073***	-0.074***	-0.088***	-0.073***
	(0.014)	(0.016)	(0.011)	(0.011)	(0.011)	(0.020)	(0.011)
Girl × mother's norms	-0.030**	-0.025*	-0.041***			-0.039***	-0.038***
	(0.013)	(0.014)	(0.011)			(0.011)	(0.012)
Girl × mother's WVS norms				-0.047***			
				(0.011)			
Girl × mother's GGI norms					-0.043***		
					(0.011)		
Girl × mother's SES						0.004	
						(0.005)	
Girl × distance							-0.008
							(0.011)
R-squared	0.72	0.70	0.71	0.71	0.71	0.71	0.71
Observations	15,036	15,326	23,759	22,710	23,044	24,632	24,632
	Indicators						
Birth order	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24). Column (1) drops Finland and Yugoslavia, the countries contributing to the largest immigrant inflow. Column (2) estimates the effect for only students' whose parents are both immigrants. Column (3) includes only siblings born a maximum of five years apart. Column (4) uses the source-country level of agreement with the statement "A university education is more important for a boy than for a girl." derived from the World Values Survey as an alternative measure of cultural gender norms, while column (5) uses the World Economic Forum's Gender Gap Index as a second alternative measure (the norm measure is defined as 1-GGI). Column (6) controls for a potential gender gap effect of mothers' socio-economic status (measured by education level), while column (7) controls for a potential gender gap effect of geographical distance between source-countries and Stockholm, Sweden. All regressions are controlling for birth order, sibling-pair and cohort fixed effects. Standard errors in parentheses are clustered at the source country × cohort level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 14: Other robustness checks: mitigation effect

	Drop mun	Both im	Max 5 years	WVS Uni	Wage ratio	SES	Distance
Girl	-0.084***	-0.077***	-0.073***	-0.068**	-0.072***	-0.069	-0.073***
	(0.032)	(0.028)	(0.026)	(0.028)	(0.026)	(0.055)	(0.027)
Girl × mother's norms	-0.023	-0.052*	-0.061**		-0.061**	-0.063**	-0.063**
	(0.037)	(0.030)	(0.030)		(0.029)	(0.029)	(0.029)
Girl × mother's norms × mun FLFP	0.083**	0.039	0.052*			0.052**	0.051**
	(0.035)	(0.027)	(0.027)			(0.025)	(0.025)
Girl × mother's WVS norms				-0.058**			
				(0.029)			
Girl × mother's WVS norms × mun FLFP				0.053*			
				(0.028)			
Girl × mother's norms × mun wage ratio					0.058**		
					(0.025)		
Girl × mother's SES						-0.001	
						(0.015)	
Girl × mother's SES × mun FLFP						-0.016	
						(0.014)	
Girl × distance							-0.013
							(0.029)
Girl × distance × mun FLFP							0.008
							(0.024)
R-squared	0.71	0.70	0.71	0.70	0.71	0.71	0.71
Observations	2,904	3,656	3,816	3,708	3,926	3,926	3,926
	Indicators						
Birth order	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Sibling FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24), and a one-standard-deviation increase in municipality FLFP (0.07). Column (1) drops Stockholm, Gothenburg and Malmö. Column (2) estimates the effect for only students' whose parents are both immigrants. Column (3) includes only siblings born a maximum of five years apart. Column (4) uses source-country level of agreement with the statement "A university education is more important for a boy than for a girl." derived from the World Values Survey as an alternative measure of gender norms. Column (5) estimates the mitigation effect with an alternative measure of municipality gender equality: the average female-to-male wage ratio. Column (5) controls for a potential gender gap effect of mothers' socio-economic status (measured by education level), while column (7) controls for a potential gender gap effect of geographical distance between source countries and Stockholm, Sweden. Standard errors in parentheses are clustered at the assigned municipality × immigration year level.
* p<0.1, ** p<0.05, *** p<0.01.

The norms a child is exposed to at home could depend on the cultural values of both parents. To investigate whether the effects differ depending on the source country of the father, I estimate the gender norm and mitigation effect using a norm measure that combines the cultural gender norms of both parents. Appendix Tables A2 and A3 show that the results using the average parental norm are very similar to my main results.

Finally, I create a sample of second-generation immigrants defined by having a foreign-born, or asylum-seeking, father instead of mother. Appendix Table A4 shows that the gender norm effect is very similar for fathers' cultural norms, compared to the results using mothers. However, Appendix Table A6 reveals that municipality gender equality does not mitigate the effect of fathers' traditional gender norms. It seems intuitive that municipality gender equality has a stronger relative impact on women then on men. However, the mitigation results for fathers should be interpreted with caution, as Appendix Table A5 shows that fathers' cultural gender norms are correlated with the gender equality of the assigned municipality. This correlation indicates that fathers were not randomly allocated to municipalities with respect to gender norms.[30]

7 Conclusion

This paper estimates the effect of cultural gender norms on the gender gap in math, and explores whether this effect is mitigated by municipality gender equality. To separate the effect of cultural norms from formal institutions, I estimate the effect of maternal source-country gender norms on the gender gap in math test performance for second-generation immigrants. By comparing the outcomes of opposite-sex siblings, I am able to control for everything that correlates with source-country but that is unrelated to gender. By construction, the remaining variation is the aspect of culture that affects opposite-sex siblings in different ways, i.e. gender norms. I show that cultural gender norms have a negative and sizeable impact on girls' relative performance, such that the sibling gender gap in math increases with the gender-traditional nature of the norms adhered to by the mother. Furthermore, I find similar effects for the gender gaps in final marks in math, Swedish, and the total ninth

[30]I expect this non-random placement to lead to an upward bias in the baseline effect of municipality gender equality on the gender gap in math, as fathers' gender norms become more equal when municipality gender equality increases. For the same reason, I expect a downward bias in the mitigation effect; this is because an increase in municipality gender equality implies an automatic decrease in fathers' gender-traditional norms, which will bias the three-way interaction estimate downwards. The results in Appendix Table A6 are in line with this expected bias.

grade mark. This result implies that the effect is not math-specific, and thus that cultural gender norms have an effect on girls' relative school performance more generally.

To investigate the mitigation effect of municipality gender equality, I exploit a refugee placement policy to obtain random variation in municipality characteristics. Again, I compare the outcomes of opposite-sex siblings, allowing me to control for any differential treatment by municipalities that is unrelated to the gender of the child. I show that municipality gender equality can almost completely mitigate the negative effect of cultural gender norms, which means that even though the sibling gender gap in math increases as mothers' gender norms become more traditional, this increase is smaller for siblings whose mothers were placed in gender-equal municipalities.

I contribute to the literature by providing causal evidence regarding the link between cultural gender norms and the gender gap in math, as well as by being the first to investigate how this effect interacts with neighbourhood gender equality. Accordingly, one novel contribution of this paper is that it merges the literature on cultural norms and neighbourhood exposure. I show that while cultural gender norms play an important role for the gender gap in math, they are not immune to the influence of surrounding characteristics. This result is important from a policy perspective. It would most likely be difficult to influence the norms transmitted to children by their parents; however, the understanding that these norms are affected by surrounding characteristics, which can be influenced, provides policy-makers with opportunities to affect change.

References

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Akerlof, G. A. and Kranton, R. E. (2002). Identity and schooling: Some lessons for the economics of education. *Journal of economic literature*, 40(4):1167–1201.

Aldén, L. and Neuman, E. (2019). Culture and the gender gap in gajor choice: an analysis using sibling comparisons. Working Paper 2019:3, Linnaeus University Centre for Discrimination and Integration Studies.

Alesina, A., Giuliano, P., and Nunn, N. (2013). On the origins of gender roles: Women and the plough. *The Quarterly Journal of Economics*, 128(2):469–530.

Angrist, J. and Pischke, J.-S. (2009). Mostly harmless econometrics: an empiricists guide.

Åslund, O., Edin, P.-A., Fredriksson, P., and Grönqvist, H. (2011). Peers, neighborhoods, and immigrant student achievement: Evidence from a placement policy. *American Economic Journal: Applied Economics*, 3(2):67–95.

Åslund, O. and Rooth, D.-O. (2007). Do when and where matter? initial labour market conditions and immigrant earnings. *The Economic Journal*, 117(518):422–448.

Bedard, K. and Cho, I. (2010). Early gender test score gaps across oecd countries. *Economics of Education Review*, 29(3):348–363.

Belot, M. V. and Hatton, T. J. (2012). Immigrant selection in the oecd. *The Scandinavian Journal of Economics*, 114(4):1105–1128.

Bisin, A. and Verdier, T. (2011). The economics of cultural transmission and socialization. In *Handbook of social economics*, volume 1, pages 339–416. Elsevier.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility i: Childhood exposure effects. *The Quarterly Journal of Economics*, 133(3):1107–1162.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.

Dahl, G. B., Felfe, C., Frijters, P., and Rainer, H. (2020). Caught between cultures: Unintended consequences of improving opportunity for immigrant girls. Working Paper 26674, National Bureau of Economic Research.

Dahl, G. B., Løken, K. V., and Mogstad, M. (2014). Peer effects in program participation. *American Economic Review*, 104(7):2049–74.

Dahl, G. B. and Moretti, E. (2008). The demand for sons. *The Review of Economic Studies*, 75(4):1085–1120.

Damm, A. P. and Dustmann, C. (2014). Does growing up in a high crime neighborhood affect youth criminal behavior? *American Economic Review*, 104(6):1806–32.

DiPrete, T. A. and Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. Russell Sage Foundation.

Dossi, G., Figlio, D. N., Giuliano, P., and Sapienza, P. (2019). Born in the family: Preferences for boys and the gender gap in math. Working Paper 25535, National Bureau of Economic Research.

Edin, P.-A., Fredriksson, P., and Åslund, O. (2003). Ethnic enclaves and the economic success of immigrants—evidence from a natural experiment. *The quarterly journal of economics*, 118(1):329–357.

Fernandez, R. (2007). Women, work, and culture. *Journal of the European Economic Association*, 5(2-3):305–332.

Fernández, R. (2011). Does culture matter? In *Handbook of social economics*, volume 1, pages 481–510. Elsevier.

Fernandez, R. and Fogli, A. (2009). Culture: An empirical investigation of beliefs, work, and fertility. *American economic journal: Macroeconomics*, 1(1):146–77.

Figlio, D., Karbownik, K., Roth, J., Wasserman, M., et al. (2019). Family disadvantage and the gender gap in behavioral and educational outcomes. *American Economic Journal: Applied Economics*, 11(3):338–81.

Finseraas, H. and Kotsadam, A. (2017). Ancestry culture and female employment—an analysis using second-generation siblings. *European sociological review*, 33(3):382–392.

Fryer Jr, R. G. and Levitt, S. D. (2010). An empirical analysis of the gender gap in mathematics. *American Economic Journal: Applied Economics*, 2(2):210–40.

Guiso, L., Monte, F., Sapienza, P., and Zingales, L. (2008). Culture, gender, and math. *Science*, 320(5880):1164–1165.

International Labour Organization (2019). Labour force participation rate by sex and age (annual). Data retrieved from the ILOSTAT Database, `https://www.ilo.org/shinyapps/bulkexplorer5/?lang=en&segment=indicator&id=EAP_DWAP_SEX_AGE_RT_A` Accessed: 2019-11-01.

Kranton, R. E. (2016). Identity economics 2016: Where do social distinctions and norms come from? *American Economic Review*, 106(5):405–09.

Maurin, E. and Moschion, J. (2009). The social multiplier and labor market participation of mothers. *American Economic Journal: Applied Economics*, 1(1):251–72.

Migrationsverket (1999). Asylsökande till sverige 1984–1999. Data retrieved from The Swedish Migrations Sgency's Statistics Database, `https://www.migrationsverket.se/Om-Migrationsverket/Statistik/Asyl.html` Accessed: 2019-11-01.

Nekby, L. and Pettersson-Lidbom, P. (2017). Revisiting the relationship between ethnic diversity and preferences for redistribution: Comment. *The Scandinavian Journal of Economics*, 119(2):268–287.

Nollenberger, N., Rodríguez-Planas, N., and Sevilla, A. (2016). The math gender gap: The role of culture. *American Economic Review*, 106(5):257–61.

Olivetti, C., Patacchini, E., and Zenou, Y. (2018). Mothers, peers, and gender-role identity. *Journal of the European Economic Association*.

Pope, D. G. and Sydnor, J. R. (2010). Geographic variation in the gender differences in test scores. *Journal of Economic Perspectives*, 24(2):95–108.

Rodríguez-Planas, N. and Nollenberger, N. (2018). Let the girls learn! it is not only about math… it's about gender social norms. *Economics of Education Review*, 62:230–253.

Rodríguez-Planas, N. and Sanz-de Galdeano, A. (2019). Intergenerational transmission of gender social norms and teenage smoking. *Social Science & Medicine*, 222:122–132.

Statistics Sweden (2019). Population and gainfully employed by year, sex and region of residence. Data retrieved from Statistics Sweden Labour Statistics Database (RAMS), `https://www.scb.se/en/finding-statistics/statistics-by-subject-area/labour-market/employment-and-working-hours/labour-statistics-based-on-administrative-sources/` Accessed: 2019-11-01.

Schüller, S. (2015). Parental ethnic identity and educational attainment of second-generation immigrants. *Journal of Population Economics*, 28(4):965–1004.

World Economic Forum (2011). The global gender gap report. World Economic Forum Genebra.

The World Bank (2019). Ratio of female to male labour force participation rate (modelled ilo estimate). Data retrieved from World Development Indicators, `https://data.worldbank.org/indicator/SL.TLF.CACT.FM.ZS` Accessed: 2019-11-01.

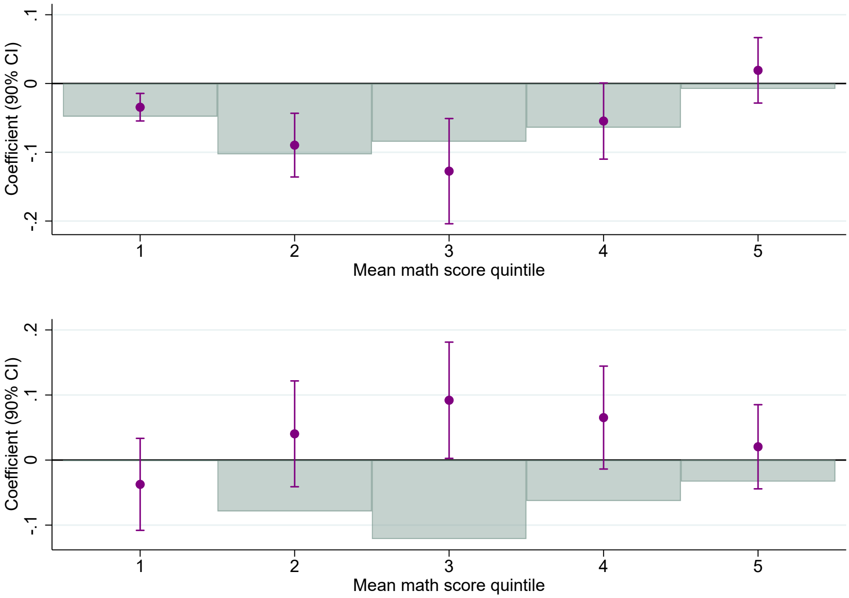
Appendix

Table A1: Effect of mothers' gender norms on the gender gap in birth weight

	(1)	(2)	(3)	(4)
Girl	-136.462***	-136.284***	-135.764***	-134.678***
	(7.696)	(7.615)	(7.506)	(4.729)
Mother's norms	-40.294***			
	(7.045)			
Girl × mother's norms	-8.203	-8.095	-8.107	-8.576*
	(8.165)	(8.157)	(8.037)	(4.901)
R-squared	0.03	0.06	0.07	0.75
Observations	24,135	24,135	24,135	24,135
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Source country & mun FE	No	Yes	Yes	No
Mother characteristics	No	No	Yes	No
Sibling FE	No	No	No	Yes

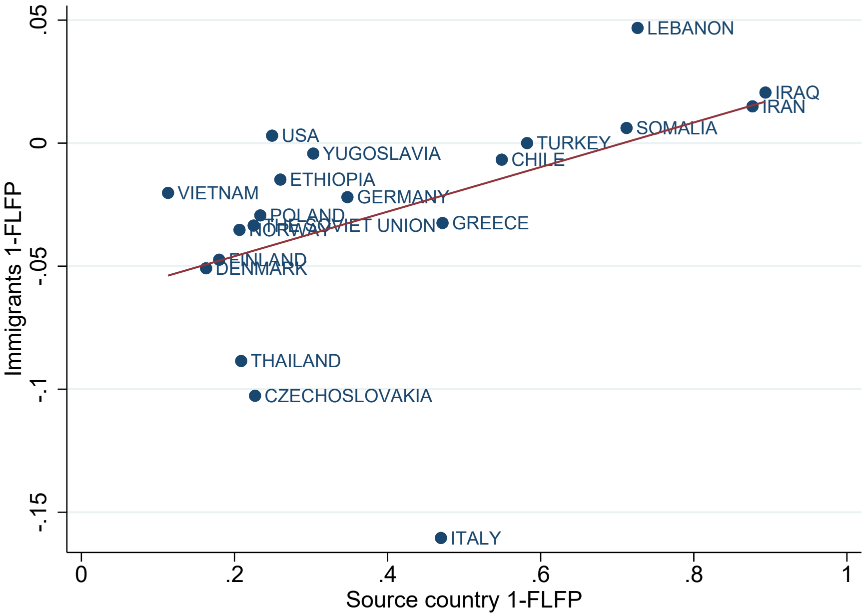
Notes: The dependent variable is birth weight, measured in grams. The effects are estimated for a one-standard-deviation increase in mothers' cultural gender norms (0.24). Mothers' characteristics are cohort, education level, household income and family size. Standard errors in parentheses are clustered at the source-country × cohort level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Figure A1: Heterogeneity: gender norm effect by math performance



Notes: The top panel plots the effect of mothers' cultural gender norms on the gender gap in math, when the effect is allowed to differ by the average math performance quintile of the siblings. The bottom panel plots the municipality mitigation effect over the same average math performance quintiles. The figure plots 90% confidence intervals. The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. The bars indicate the average gender gap in math scores per quintile.

Figure A2: Correlation: source-country and immigrant 1-FLFP rates



Notes: The figure plots the correlation between source-country 1-FLFP rates and the average current 1-FLFP rate of the first generation. The correlation coefficient is 0.57.

Table A2: Effect of parents' average gender norms on the math gender gap

	(1)	(2)	(3)	(4)
Girl	-0.068***	-0.068***	-0.068***	-0.068***
	(0.013)	(0.013)	(0.013)	(0.012)
Avg. norms	-0.157***	-0.066*	-0.029	
	(0.021)	(0.036)	(0.034)	
Girl × avg. norms	-0.044***	-0.043***	-0.044***	-0.044***
	(0.013)	(0.013)	(0.013)	(0.011)
R-squared	0.04	0.12	0.18	0.71
Observations	23,418	23,418	23,418	23,418
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Source country & mun FE	No	Yes	Yes	No
Parent characteristics	No	No	Yes	No
Sibling FE	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. The sample consists of second-generation immigrants defined by a foreign-born mother. The effects are estimated for a one-standard-deviation increase in parents' average cultural gender norms ((mothers' 1-FLFP + fathers' 1-FLFP)/2). Parents' characteristics are cohort, education level, household income and family size. Standard errors in parentheses are clustered at the mother's source country × father's source country level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A3: Parents' average gender norms: municipality mitigation effect

	(1)	(2)	(3)	(4)	(5)
Girl	-0.078***	-0.073***	-0.073***	-0.073***	-0.077***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.027)
Girl × avg. norms	-0.063**	-0.060**	-0.061**	-0.061**	-0.060**
	(0.029)	(0.030)	(0.029)	(0.030)	(0.029)
Girl × mun FLFP		-0.038	-0.041	-0.041	-0.040
		(0.026)	(0.026)	(0.026)	(0.026)
Girl × avg. norms × mun FLFP		0.045*	0.043	0.044*	0.044*
		(0.027)	(0.027)	(0.027)	(0.026)
R-squared	0.70	0.02	0.33	0.34	0.70
Observations	3,696	3,696	3,696	3,696	3,696
			Indicators		
Birth order	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
Source country FE	No	No	Yes	Yes	No
Mun. × Im.Year FE	No	No	Yes	Yes	No
Parent characteristics	No	No	No	Yes	No
Sibling FE	Yes	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. Municipality FLFP is measured at the assigned municipality, in the year in which the mother immigrated to Sweden. The sample consists of second generation immigrants defined by an asylum-seeking mother affected by the refugee placement policy. The effects are estimated for a one-standard-deviation increase in parents' average cultural gender norms ((mothers' 1-FLFP + fathers' 1-FLFP)/2), and a one-standard-deviation increase in municipality FLFP (0.07). Parents' characteristics are cohort, education level, family size and cultural density of assigned municipality, all of which are measured at time of immigration. Standard errors are clustered at assigned municipality × immigration year level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table A4: Effect of fathers' gender norms on the math gender gap

	(1)	(2)	(3)	(4)
Girl	-0.063***	-0.063***	-0.064***	-0.063***
	(0.013)	(0.012)	(0.012)	(0.010)
Father's norms	-0.166***			
	(0.022)			
Girl × father's norms	-0.041***	-0.040***	-0.040***	-0.042***
	(0.012)	(0.012)	(0.011)	(0.009)
R-squared	0.04	0.12	0.18	0.72
Observations	27,740	27,740	27,740	27,740
	Indicators			
Birth order	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes
Source country & mun FE	No	Yes	Yes	No
Father characteristics	No	No	Yes	No
Sibling FE	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. The effects are estimated for a one-standard-deviation increase in fathers' cultural gender norms (0.25). Fathers' characteristics are cohort, education level, household income and family size. Standard errors in parentheses are clustered at the source-country × cohort level. * p<0.1, ** p<0.05, *** p<0.01.

Table A5: Balance test for refugee placement policy: fathers

	(1)	(2)	(3)	(4)
	Outcome: assigned mun FLFP			
Father's norms	-0.090**	-0.063*	-0.062*	-0.062*
	(0.037)	(0.037)	(0.036)	(0.037)
Cultural density		0.189***	0.193***	0.193***
		(0.036)	(0.036)	(0.035)
Father's education		0.000	0.004	0.004
		(0.014)	(0.014)	(0.014)
Family size		-0.056**	-0.053**	-0.055*
		(0.025)	(0.026)	(0.029)
Father's cohort				-0.001
				(0.006)
R-squared	0.01	0.05	0.06	0.06
Observations	3,368	3,368	3,368	3,368
	Indicators			
Immigration year FE	No	No	Yes	Yes

Notes: The table reports the correlation between municipality FLFP and fathers' source-country gender norms and individual characteristics. The dependent variable is the FLFP of the assigned municipality. All fathers' characteristics are measured at the time of immigration. Standard errors are clustered at the assigned municipality × immigration year level. * p<0.1, ** p<0.05, *** p<0.01.

Table A6: Municipality mitigation effect: fathers

	(1)	(2)	(3)	(4)	(5)
Girl	-0.082***	-0.085***	-0.085***	-0.083***	-0.085***
	(0.028)	(0.028)	(0.028)	(0.028)	(0.028)
Girl × father's norms	-0.023	-0.025	-0.025	-0.023	-0.024
	(0.029)	(0.030)	(0.029)	(0.029)	(0.029)
Girl × mun FLFP		0.016	0.014	0.014	0.009
		(0.026)	(0.026)	(0.026)	(0.026)
Girl × father's norms × mun FLFP		-0.027	-0.027	-0.026	-0.026
		(0.029)	(0.029)	(0.029)	(0.029)
R-squared	0.71	0.01	0.33	0.34	0.71
Observations	3,368	3,368	3,368	3,368	3,368
	Indicators				
Birth order	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes
Source country FE	No	No	Yes	Yes	No
Mun. × Im.Year FE	No	No	Yes	Yes	No
Father characteristics	No	No	No	Yes	No
Sibling FE	Yes	No	No	No	Yes

Notes: The dependent variable is the student's test score on the national standardised test in math, standardised to have a mean of 0 and standard deviation of 1. Municipality FLFP is measured at the assigned municipality, in the year in which the father immigrated to Sweden. The effects are estimated for a one-standard-deviation increase in fathers' cultural gender norms (0.25), and a one-standard-deviation increase in municipality FLFP (0.07). Fathers' characteristics are cohort, education level, family size and cultural density of assigned municipality, all of which are measured at time of immigration. Standard errors are clustered at the assigned municipality × immigration year level. * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Chapter III

Backlash: Female Economic Empowerment and Domestic Violence

Abstract

This paper estimates the effect of female economic empowerment on domestic violence. I use individual level data from high-quality Swedish administrative registers on women's earnings and hospital visits relating to assault. With this third-party reported violence measure I overcome the issue of selective under-reporting of violence. I proxy female economic empowerment with a measure of women's potential earnings, caused by local changes in female-specific labour demand. This measure reflects the outside option of the marriage, and captures earnings variation that is not endogenous to domestic violence. I show that, even while keeping the earnings of husbands constant, the causal effect of increasing women's potential earnings on domestic violence is positive and substantial. In addition, I show that increasing women's potential earnings increase the husbands' risk of destructive behaviour, such as stress, anxiety, substance abuse and assault. Taken together, these results indicate that improving women's financial independence triggers a male backlash response, even in a gender-equal country like Sweden.

Keywords: domestic violence, potential earnings, household bargaining, male backlash, local labour demand
JEL Classifications: D13, I12, J12, J16, Z13

1 Introduction

Domestic violence is a major issue for public health, productivity and gender equality. Globally, one in three women will experience violence from a partner at some point during their lifetime (García-Moreno et al., 2013), and the global cost of total intimate partner violence is estimated to be 5.2% of world GDP (Hoeffler, 2017). Female economic empowerment is often cited as one of the most effective ways to combat domestic violence, but the theoretical predictions diverge. Models of household bargaining predict that a better relative economic position of the wife improves the outside option of the marriage, and, as a result, reduces violence (Farmer and Tiefenthaler, 1997; Aizer, 2010).[1] In contrast, models of male backlash predict that an improved relative economic position of the wife increases violence, as it violates traditional gender norms and redefines the power relationship between the spouses, which could trigger a violent backlash response from the husband (Hornung et al., 1981; Macmillan and Gartner, 1999).

However, investigating the relationship between female economic empowerment and domestic violence offers several empirical challenges. First, domestic violence is a sensitive topic that is prone to selective under-reporting (Ellsberg et al., 2001). Self-reported measures might not be representative of actual violence, but rather, the selection in who reports a violent incident. As the probability of reporting a violent incident is likely to increase with empowerment (Iyer et al., 2012), it is important to distinguish between changes in violence and changes in reporting behaviour.[2]

Second, if we measure female empowerment using earnings, an important threat to identification is earnings endogeneity. Realised earnings likely reflect unobserved individual characteristics, which could be an outcome of, or correlate with, violence. Assortative matching will create selective marriages that are functions of earnings, and of the underlying propensity for both perpetrating violence and for staying in a violent relationship (Pollak, 2004), which makes relative earnings a problematic measure of empowerment within households. Furthermore, the

[1] An improved economic position raises the threat point of the wife by improving her outside option of the marriage. The outside option is the situation she would face in case of a marriage dissolution. A higher threat point has a negative effect on violence, both indirectly, through more women leaving abusive spouses, and directly, through the deterrent effect of the threat of leaving.

[2] Furthermore, analysing aggregate measures carries a risk of ecological fallacy conclusions. An ecological fallacy occurs when we make inference about individuals based on inference about the groups to which the individuals belong. This issue becomes especially important with a relatively rare outcome such as domestic violence, for which data sparsity can lead to aggregate measures that might not properly represent the underlying distribution of violence.

outside option of a marriage is not determined by a woman's realised earnings, but rather by the earnings potential she would face in case of a marriage dissolution (Aizer, 2010).

In this paper I estimate the causal effect of female economic empowerment on domestic violence, and I overcome the empirical challenges noted above. I use high-quality Swedish administrative data, enabling me to observe both earnings and violence on an individual level. I measure domestic violence using hospital visits for assault, which I derive from third-party reported hospital records. The benefit of using hospital data is that my measure of violence suffers from very little selective reporting bias. In addition, information on hospital visits for accidents allows me to examine possible misreporting of injury causes at the hospital, and enables me to conclude that this does not pose a threat to my study.

To overcome the empirical challenges of earnings endogeneity, I proxy female economic empowerment with a measure of women's potential earnings. I exploit the fact that women and men tend to sort into different industries and create a measure of prevailing local female earnings potential, which captures exogenous variation in female economic empowerment caused by only local demand changes for female labour. Thus, this measure captures earnings variation that is not endogenous to domestic violence, and it provides a more accurate representation of the outside option of a marriage.

I show that increasing women's potential earnings, while keeping the earnings of their husbands constant, increases the risk of assault. The effect is substantial in magnitude, and does not depend on which spouse earns more than the other. Thus, my results are in line with the predictions of male backlash theory, as they show that an improved relative economic position for the wife increases the risk that she experiences assault. As further support for the backlash mechanism, I show that increasing women's potential earnings, while keeping the earnings of their husbands constant, increase the risk of husbands' destructive behaviour, such as visiting a hospital for reasons related to depression, anxiety, substance abuse and assault.

The richness of my data allows me to conduct a detailed heterogeneity analysis, where I show that the effect of increased potential earnings differs depending on the sub-group of the population. For the youngest women the effect of increased potential earnings is negative, but after the age of 40 the effect is consistently positive. Likewise, for the women with no more than high school education potential earnings reduce the risk of assault, but for women of higher education levels potential earnings increase the risk. Speculatively, the difference in effects could depend on te credibility of the threat of leaving an abusive spouse. A young woman's threat of leaving may be more credible as it has yet to be tested. Similarly, for the least educated women the outside option may be binding, such that increased potential earnings may significantly affect their ability to leave an abusive spouse. Thus, my results indicate that the women for whom a change in potential earnings actually affects the credibility of their threat of leaving, the results are in line with the predictions of bargaining power theory. But for the older women, and the women who may have the economic possibility to leave their spouse, but still do not, the effects are in line with male backlash theory. In line with this reasoning, I show that the backlash effect also increases with the duration of the marriage.

My paper relates to the growing and diverse literature on female empowerment and domestic violence. Several well-identified empirical studies find support for the bargaining power hypothesis, i.e. that female empowerment reduces domestic violence. Brassiolo (2016) and Stevenson and Wolfers (2006) show that the introduction of unilateral divorce laws in Spain and the US led to large reductions in domestic violence, and Anderberg et al. (2016) find that spousal abuse varies negatively with male unemployment but positively with female unemployment.[3]

In contrast, several studies also find support for the male backlash theory, i.e. that female empowerment increases violence. Recent studies show that female employment varies positively with spousal abuse in many developing countries (Heath, 2014; Cools and Kotsadam, 2017; Bhalotra et al., 2018). Chin (2012) and Guarnieri et al. (2018) show causal support for a backlash response, by using rainfall shocks or historic institutional differences to identify exogenous variation in women's employment opportunities.

My paper also relates to a strand of the literature that connects domestic violence prevalence to the concept of gender identity. The gender identity model in economics was introduced by Akerlof and Kranton (2000). The connection to relative earnings and domestic violence was first established by Atkinson et al. (2005). They construct a measure of "traditionalism", using couples level of agreement with various statements,[4] and find that the relative earnings of wives is only positively correlated with violence if the wives are married to "traditional" husbands. More recently,

[3]In a developing country context, La Mattina (2017) shows that Rwandan women who married after the genocide in 1994 experience more domestic violence, consistent with the hypothesis that a shortage of men led to reductions in women's bargaining power in the marriage market.

[4]In the economics literature traditional gender norms are usually measured by how much couples agree with statements such as "A mother can work full-time when she has a child under the age of 5", "It is much better for everyone if the man earns the main living and the woman takes care of the home and family", or "If a woman earns more money than her husband, it is almost certain to cause problems" (e.g. Atkinson et al., 2005; Bertrand et al., 2015).

Svec and Andic (2018) show that women who have higher earnings than their partners are more at risk of experiencing domestic violence in Peru, and Alonso-Borrego and Carrasco (2017) find that a woman's employment only reduces violence when her partner is also employed. Tur-Prats (2017) finds that male backlash responses to female relative employment only exists for couples who live in areas that historically contained families with more traditional gender norms.[5] Finally, a recent strand of the literature finds that domestic violence can be triggered by negative emotional cues or psychological stress (Card and Dahl, 2011; Cesur and Sabia, 2016; Beland and Brent, 2018).

The paper closest to mine is Aizer (2010), who studies the effect of changes in the gender wage gap on aggregate measures of female hospitalisations for assault in Californian counties. She exploits demand-driven exogenous variation in the gender wage gap and finds that narrowing the wage gap leads to reduced levels of female hospital admissions for assault, a relationship that is consistent with the bargaining power hypothesis. My main results stand in contrast to those of Aizer (2010), as I find a positive effect of women's potential earnings on domestic violence. However, my heterogeneity analysis reveal a negative relationship between potential earnings and assault for the lowest educated women, which are the only ones included in the study of Aizer (2010). Furthermore, I replicate Aizer's study on Swedish data and, consistent with her results, I find a negative relationship between the aggregate gender wage gap and municipality-level of assault. I find a positive effect, consistent with my main results, using an individual-level outcome measure that is as close as possible to Aizer's in my main model. Taken together, these results indicate that an aggregate study may mask what is happening on the individual level, as it cannot identify heterogeneous effects for different sub-groups of the population.

My contributions to the literature are threefold. First, my study is the first to use a (close to) objective measure of violence from individual level data when studying the effect of female empowerment on domestic violence. My study is also the first to investigate possible misreporting at the hospital, which allows me to conclude that my results are not suffering from selective reporting bias.

Second, my results demonstrates that the effect of increased potential earnings on domestic violence differs sharply for different subgroups of the population. This way, I show that both effects in line with the bargaining power hypothesis and effects in line with the theory of male backlash can co-exist, depending on the subgroup of the population and, speculatively, on how credible their threat of leaving an abusive

[5]In addition, cross-country evidence shows that domestic violence prevalence is higher in countries with more traditional gender norms (Heise and Kotsadam, 2015; González et al., 2018)

spouse is.

Third, I investigate the mechanisms behind the positive effect of women's potential earnings on domestic violence. By estimating the effect of women's potential earnings on various measures of husbands' destructive behaviour, I show that the mechanisms are in line with a male backlash response. This result is especially interesting to find in a gender-equal country like Sweden.

The remainder of the paper is organised as follows. In Section 2, I introduce the data and the dependent and independent variables; Section 3 outlines the empirical strategy; Section 4 presents the results; Section 5 ensures that they are robust; and Section 6 concludes.

2 Data

My data come from the Swedish Interdisciplinary Panel (SIP). It is a two-generational dataset covering all individuals born in Sweden between 1973 and 1995 and their parents, both whose outcomes I observe during 2001–2011. I acquire the indicator of violence exposure from individual level hospital records, which cover both in-patient and out-patient visits. The in-patient records refer to all hospital visits that last for at least one over-night stay at the facility. The out-patient records do not cover primary care, but contain all other contacts with specialised care providers. Most importantly for my study, they contain all contact with medical doctors that takes place in emergency rooms. Compared to the data used by Aizer (2010), my data contain a wider array of hospital visits, including less severe cases that do not require overnight hospitalisation.[6]

My dependent variable is a binary indicator of visiting a hospital for an injury caused by assault during the current year.[7] I identify cases of assault using ICD-10 diagnosis codes, reported by medical personnel at the moment of the visit.[8] In order to isolate assaults that are likely of a domestic nature, I use only assaults that took

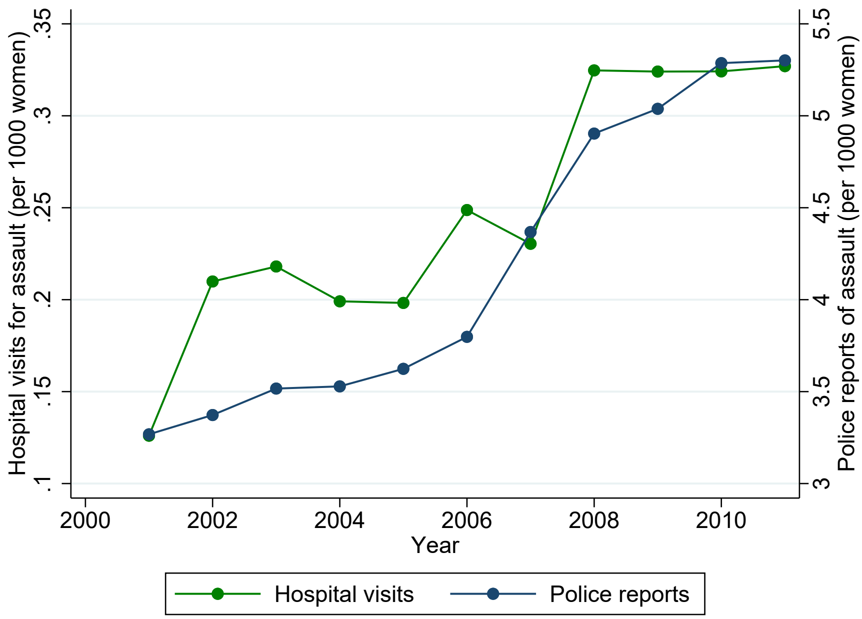
[6] As a robustness check I perform my analysis using the outcome measure closest to Aizer (2010), see column (4) of Table 7, and the results do not change.

[7] Some women visit the hospital multiple times a year for assault-related injuries, but these incidences are rare. However, my results are robust to using a count variable instead of a binary indicator.

[8] Each ICD-10 code contains a capital letter denoting the broad category of the diagnosis, followed by a sequence of digits. If the diagnosis was externally caused it contains a similar code denoting the circumstances of the external cause; this is where I can isolate cases of assault. Codes X85–Y09 denote assault as the cause of injury, both physical and sexual. The fourth letter of the code denotes where it took place.

place at home or in unspecified locations.[9] Measurement error due to violence committed by someone other than an intimate partner is unlikely, as the most common perpetrator of female assault is someone close to the victim (Frenzel, 2014).[10] Figure 1 shows the average number of hospital visits for assault that occurred at home or in unspecified locations, and the average number of filed police reports for domestic assaults, for each year between 2001–2011. Both measures depict a positive time trend, indicating that domestic violence has increased during the first decade of the 21st century.

Figure 1: Mean assaults over time



Notes: The figure plots the average number of hospital visits for assault that occurred at home or in unspecified locations, and the average number of filed police reports for domestic assaults, per 1000 women.

[9] Assaults taking place, e.g. in workplaces, bars, or public events are, therefore, not included in my main sample. Unspecified location is kept as discussions with medical personnel at Swedish emergency rooms revealed that this notation is often used to save time during stressful situations. However, my results are robust to including only the assaults that took place at home, see column (2) of Table 6.

[10] In 72% of physical assaults against women the perpetrator is someone close to the victim, a number which is likely understated as the probability of making a police report is lower when the perpetrator is someone close.

My final dataset is an unbalanced panel consisting of married women of working age, i.e. 20–65 years old.[11] The final dataset contains 7,965,166 observations spanning over 1,046,867 individuals. Table 1 shows the summary statistics. The women are on average 46 years old, and 8% of them will get a divorce within the time frame of my study. As I restrict the sample based on marriage, a large fraction of my sample consists of the parental generation, who are sampled simply because they are parents to someone born between 1973 and 1995, therefore, the fraction of women who have children is unusually high. Each year, on average 0.39 women per 1,000 visit a hospital due to assault.[12] Most women in my sample only visits a hospital for assault once during my time frame, which contradicts the notion that domestic violence is often a recurring event.[13] However, my data contain mainly aggravated assault and, therefore, only the tip of the iceberg of true violence numbers, as any minor assaults that do not result in a hospital visit are not captured in my study.

[11] I exclude same-sex married couples.

[12] Average assault without the marriage restriction 0.76 per 1,000 women. Compared to Aizer (2010) my average number of assaults, including every hospital visit, is about four times as common as what she finds in California, which is reasonable as Swedish hospital records contain a wider array of visit types. My numbers for in-patient hospitalisations only are similar to Aizer (2010).

[13] Over 20% of the women who reported to have been subjected to aggravated assault in 2012 stated that violence occurred several times a week; another 20% experienced violence several times a month; and the remaining women stated that that violence occurred between one and a few times a year (Frenzel, 2014).

Table 1: Summary statistics

	Socio-demographic variables
Age	46.08
	(9.98)
Wife older	0.17
	(0.38)
Have children	1.00
	(0.04)
Have university degree	0.38
	(0.49)
Ever get divorce	0.08
	(0.27)
	Hospital visits per year and 1000 women
Pr(hospital visit for assault at home or unspec.)	0.25
	(15.77)
Pr(hospital visit for similar accidents)	0.65
	(25.54)
Pr(hospital visit for appendix complications)	0.88
	(29.67)
Total hospital visit for any assault	0.39
	(32.17)
Total in-patient hospital visit for any assault	0.05
	(7.55)
	Earnings and income, in 1000
Employed	0.95
	(0.23)
Wives' yearly earnings	190.65
	(133.95)
Husbands' yearly earnings	290.68
	(258.80)
Household yearly income	549.29
	(304.05)
Relative earnings	-0.15
	(0.53)
	Potential earnings, in 1000
Women's potential earnings	198.99
	(46.38)
Husbands' potential earnings	291.34
	(72.99)
Observations	7,965,166
No. couples	1,046,867

Notes: The table reports means and standard deviations (in parentheses) of key variables. All earnings and income measures are inflation-adjusted and reported in 2000 levels.

2.1 Selection

Selection in the probability of visiting a hospital, given an assault, is unlikely to be a large issue for my results. Frenzel (2014) show that the number of assaulted women who visit a hospital is substantially higher than that of those who file a police report, which implies that women seek medical care out of need and severity of injury rather than by choice.[14] Selection due to income constraints is unlikely as visiting a hospital or health care unit in Sweden is free, except for a small fee of, at most, 1150 SEK (about 110 Euro) per year.[15]

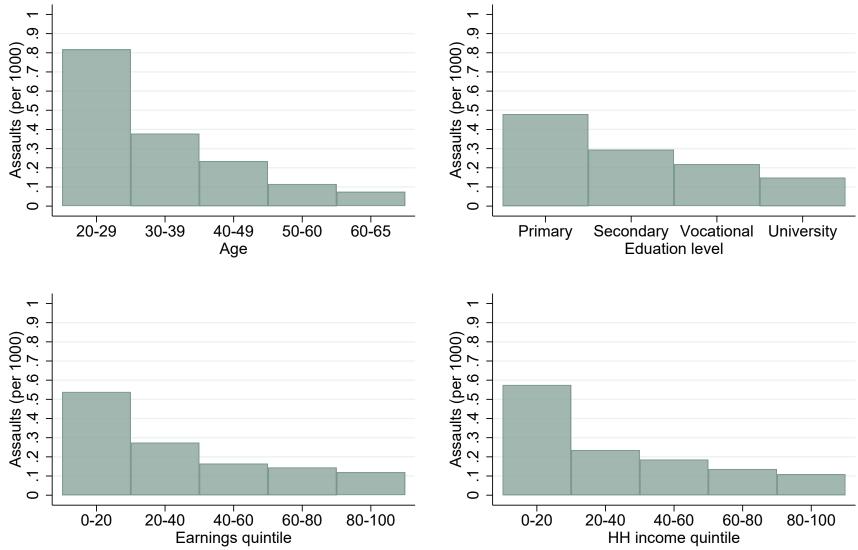
However, the risk of experiencing an assault is not randomly distributed among the population. Rather, it correlates with various socio-demographic measures. Figure 2 shows that the risk of assault significantly decreases with age, education level, earnings and household income. These correlations highlight the importance of adjusting flexibly for different socio-demographic characteristics when investigating the determinants of domestic violence.

Another important, but unobserved, factor is selective marriage matches. Pollak (2004) argues that growing up in a violent home makes men more prone to violence and women more prone to staying in a violent marriage. Moreover, men and women who grew up in violent homes are more likely to marry each other, which makes assortative matching in the marriage market a crucial determinant of domestic violence. As men and women are likely to marry assortatively based on earnings as well, this selection could bias my results if it is not accounted for.

[14]The number of women who suffered from physical intimate partner violence in 2012, who later filed a police report, was only 4%. In contrast, 12% of the assaulted women, and 29% of those who suffered from aggravated assault, contacted health care services.

[15]Visits are confidential and everyone working in a hospital is bound to professional secrecy by law. However, there are some exceptions. If a child is believed to be in danger, hospital staff are obligated to report this to the appropriate authority. If a crime with minimum punishment of one year in prison has been committed, staff have the option of reporting this to the police. Thus, according to Swedish law, hospital staff have the possibility to report cases of aggravated assault, but it is unlikely that they would do so against a patient's will.

Figure 2: Mean hospital visits for assault by age, education and income



Notes: The figure plots average hospital visits for assault, that took place at home or in unspecified locations, for different subgroups of the population.

3 Empirical strategy

3.1 Descriptive relationship: relative earnings

As a descriptive exercise, I first define female economic empowerment as relative earnings. I measure relative earnings as wives' share of half of the household labour earnings, as follows:

$$RelativeEarnings_{it} = \frac{WifeEarn_{it}}{(WifeEarn_{it} + HusbandEarn_{it})/2} - 1 \quad (1)$$

The measure spans between -1 and 1, where -1 means that the husband earns 100% of the household earnings, 0 implies perfect earnings equality and 1 means that the wife earns 100% of the household earnings. I measure earnings as labour income.

I model the relationship between spouses' relative earnings and domestic violence non-parametrically, without imposing any assumptions on its functional form.[16]

3.2 Identifying the effect of female economic empowerment

An important threat to identification in the descriptive exercise above is the endogeneity of earnings. Realised earnings likely reflect underlying characteristics of the individual, which could be a function of underlying violence (abused women are less productive) or unobservables that might correlate with violence. Furthermore, the outside option of a marriage is not determined by a woman's realised earnings, but rather by the earnings potential she would face in case of a marriage dissolution (Aizer, 2010).[17] The relative level of economic empowerment within a marriage depends on the outside option, and, therefore, on spouses' earnings potential rather than realised earnings.

To account for earnings endogenity, and for the fact that theory predicts that potential, rather than actual, earnings determine the bargaining power of a woman, I construct a measure of prevailing female earnings that reflects only the exogenous demand for female labour. I exploit the fact that men and women tend to sort differentially across industries (for example women are overrepresented in the health and service sector, whereas men sort into manufacturing and construction) and that these industries experience different wage growth over time. This approach builds on previous work by Bartik (1991), Aizer (2010), Bertrand et al. (2015) and Lindo et al. (2018), and isolates gender-specific variation in earnings, driven only by changes in local labour demand. Thus, it is a measure of women's local earnings potential. I allow the measure to vary by age and education, to take into account that wages are usually set depending on education level and that they tend to increase with experience. I construct the measure of women's potential earnings as:

$$PotentialEarnings_{maet} = \sum_{j} \gamma^{f}_{maej,2000} \times w^{f}_{eajt,-m} \qquad (2)$$

Where f denotes female, m municipality of employment, a age group, e education group, j industry, and t year. $\gamma^{f}_{maej,2000}$ is the share of women, in a given age and education group, who work in industry j in municipality m in the base year 2000.

[16] I use kernel-weighted local linear scatterplot smoothing for the estimations.

[17] Women's earnings at the threat point determine the bargaining power, and earnings at the bargaining equilibrium do not necessarily equal earnings at the threat point. Pollak (2005) provides an example of a married woman who does not work (zero earnings) at the cooperative equilibrium but who would work in the event of the dissolution of the marriage.

This proportion is fixed, so changes in earnings do not reflect selective sorting across industries. Variable $w^f_{eajt,-m}$ is the average national annual earnings in year t in industry j for women of a given age and education group, excluding municipality m. Consequently, the potential earnings of a woman, of a given age and education level, is higher the larger the share of similar women in her municipality who are employed in industries with high national wage growth is. I construct a similar measure for potential earnings of the husbands of my study.

I test the validity of the potential earnings proxy in two ways. First, one advantage of my data is that I can ensure that the potential earnings measure is correlated with realised, individual and relative, earnings. The first six columns of Table 2 show the estimated correlations between wives' and husbands' potential earnings and realised wives', husbands' and relative earnings. Reassuringly, the estimated correlation between wives' potential and realised earnings and is positive, sizeable and precise, which shows that potential earnings is an appropriate proxy for capturing earnings variation of women.

Second, a correlation between wives' (or husbands') potential earnings and realised earnings of the other spouse implies that the potential earnings measure is a more general measure of labour market shifts, and do not only capture gender-specific earnings shifts as it should. Column (2) of Table 2 shows that wives' potential earnings are, as they should, uncorrelated with the realised earnings of husbands. However, column (4) shows that husbands' potential earnings are positively correlated with wives' realised earnings, which implies that husbands' potential earnings is a less appropriate measure of earnings shifts for men only. Taken together, these validity tests show that women's potential earnings is an appropriate measure to capture female earnings variation that is not endogenous to domestic violence, but that husbands' potential earnings (and therefore also potential relative earnings) is more problematic.[18]

[18] Finally, following Goldsmith-Pinkham et al. (2018), I verify that no single industry contributes the majority of the identifying variation in potential earnings. Appendix Figure A1 shows the distribution of Rotemberg weights. Each weight corresponds to the misspecification elasticity for each industry-period pair, and measures how sensitive the parameter estimate is to each instrument. Although this test is mainly required for using industry shares as instruments, it is still reassuring that Appendix Figure A1 shows that the identifying variation is dispersed among several industries. The industry contributing the largest share of the identifying variation is the telecommunication sector, followed by the financial sector.

Table 2: Relationship between potential earnings and realised earnings

	Correlation with realised						Cross effects?
	Wives'	Husbands'	Relative	Wives'	Husbands'	Relative	Relative
Women's potential earnings	0.163***	0.003	0.069***				0.116***
	(0.007)	(0.005)	(0.005)				(0.005)
Husbands' potential earnings				0.018***	0.112***	-0.091***	-0.108***
				(0.003)	(0.005)	(0.003)	(0.003)
	Indicators						
Basic FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Detailed FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.35	0.37	0.26	0.35	0.37	0.38	0.07
Observations	7,965,166	7,965,166	7,965,166	7,965,166	7,965,166	7,965,166	7,965,166

Notes: The table reports correlations between potential earnings and realised earnings. The effect is measured as a change of one standard deviation. Basic FE: Municipality, year, age group, education group and cohort. Detailed FE: basic FE plus household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age span and husbands' earnings. Standard errors in parentheses are clustered at the $mun * edugroup * agegroup$ level. * p<0.05, ** p<0.01, *** p<0.001.

To capture the causal effect of women's potential earnings, as a proxy for female economic empowerment, I estimate the following two models:

$$\begin{aligned} Assault_{it} = \alpha + \beta PotentialEarnings_{it} + \delta HusbandsEarnings_{it} \\ + Z'_{it}\theta + X_i + \epsilon_{it} \end{aligned} \quad (3)$$

and

$$\begin{aligned} Assault_{it} = \alpha + \beta PotentialEarnings_{it} \\ + \gamma PotentialEarnings_{it} \times H \leq W_{it} + H \leq W_{it} \\ + \delta HusbandsEarnings_{it} + Z'_{it}\theta + X_i + \epsilon_{it} \end{aligned} \quad (4)$$

$Assault_{it}$ is a binary variable that takes the value one if the woman has visited a hospital for injuries related to assault. X_i contains indicators for age group, household income quintiles, municipality of residence, year, cohort, and education level of both spouses. Z_{it} contains linear controls for spousal age difference and municipality-level annual filed police reports for assaults against women. I keep the earnings of the husbands constant, to identify variation in relative female economic empowerment that is driven solemnly by changes in local labour demand for women. The coefficient of interest in equation (3) is β, which captures the effect of women's potential earnings on the probability that a the woman visits a hospital for assault. In equation (4) I allow the effect to differ depending on which spouse earns more than the other, and $H \leq W_{it}$ is a binary variable that takes the value of one if wives earn an equal wage to, or more than, their husbands. The coefficients of interest in equation (4) are the respective sums of β and γ, which capture the differential effect of women's potential earnings along the realised relative earnings distribution.

The identifying assumption is that women's potential earnings is as good as random, conditional on observables. Assuming that unobservable characteristics relating to domestic violence do not deviate from the $municipality * agegroup * educationgroup$ trend when its economic conditions deviate from the trend, this approach will uncover the causal effect of women's earnings potentials.

The estimated effect of women's potential earnings will consist of two mechanisms; first, as column (3) of Table 2 shows, women's potential earnings varies positively with realised relative earnings, which means the measure is partly capturing the effect of increasing relative earnings. Second, regardless of realised earnings, an increase in potential earnings implies an improved outside option of the marriage, which may also have an effect on domestic violence. I cannot distinguish between these two mechanisms, but as they both imply increased relative economic empowerment for wives there is no need to. However, this means that using potential earnings as an instrument to capture exogenous variation in relative earnings only is not appropriate.[19]

4 Results

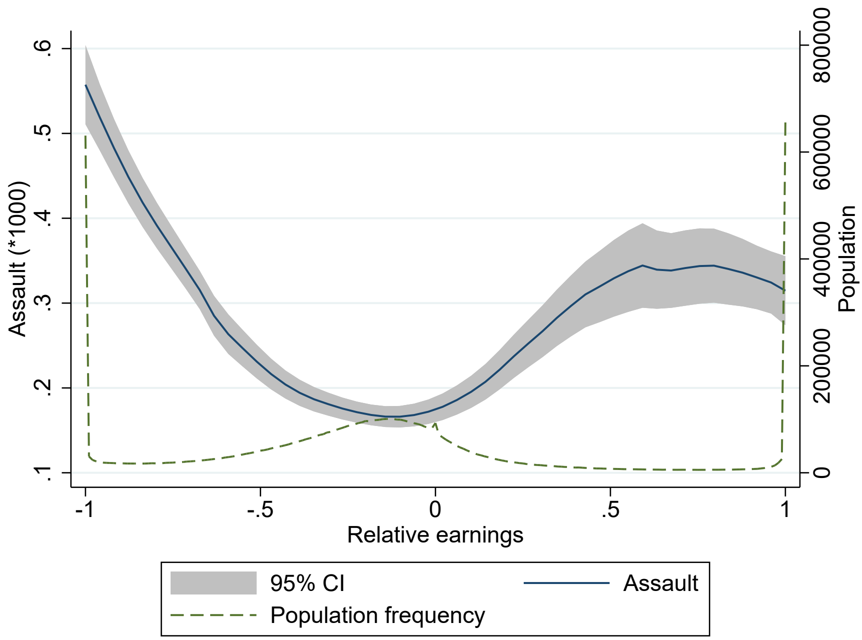
4.1 Descriptive relationship: relative earnings

Figure 3 shows the non-parametric relationship between relative earnings and hospital visits for assault. Figure 4 shows the same relationship, but excluding those with zero earnings. Both figures show that the relationship is U-shaped; such that higher resource inequality is associated with higher levels of assault. Consequently, the effect of a change in relative earnings differs depending on whether the husband or wife earns more than the other. An increase in the wife's relative earnings when the husband earns more than her is associated with lower levels of violence. However, increases in relative earnings are positively associated with violence when the wife earns more than the husband.[20]

[19] A full IV analysis requires strong additional identifying assumptions, such as an exclusion restriction, which is an inherently untestable assumption that I believe is hard to argue for in this setting. For this reason, I focus only on the reduced form model using potential earnings as a proxy rather than an instrument.

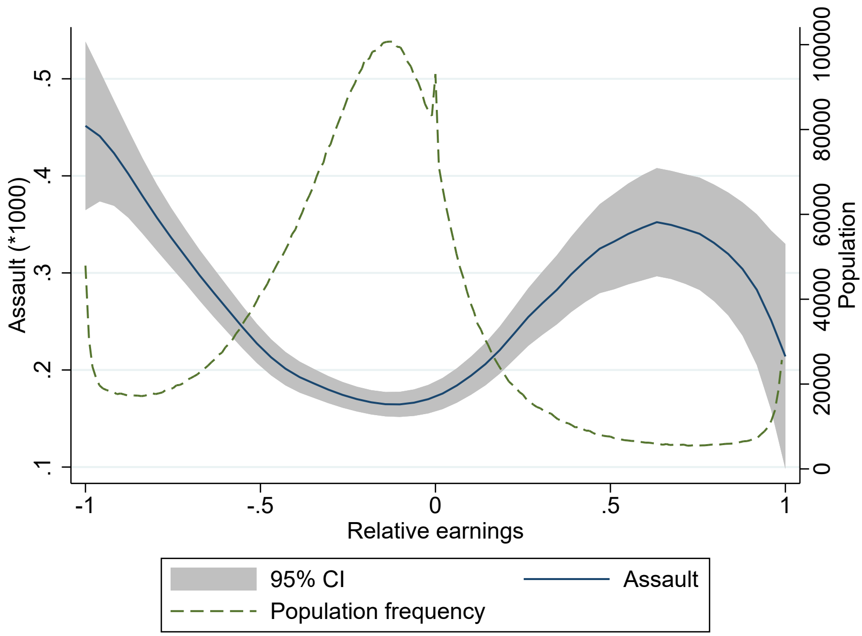
[20] Appendix Table A1 shows that this descriptive relationship holds also in parametric estimations. Columns (1)–(3) estimates a linear probability model. Column (4) accounts for unobserved heterogeneity not captured by the linear model (e.g. selective marriage matches), by exploiting the panel dimension of the data and introducing individual fixed effects to the model.

Figure 3: Descriptive relationship between relative earnings and assault



Notes: Figure 3 plots the descriptive relationship between female hospital visits for assault and relative earnings of spouses. The dashed line show the frequency distribution of relative earnings.

Figure 4: Descriptive relationship: dual-earners only



Notes: Figure 4 plots the descriptive relationship between female hospital visits for assault and relative earnings of spouses, for only dual-earner couples only. The dashed line show the frequency distribution of relative earnings.

4.2 Effect of potential earnings on domestic violence

Table 3 shows the effect of women's potential earnings on the risk of visiting a hospital for assault. All specifications are keeping the earnings of the husband constant, to identify variation in relative economic empowerment that is driven only by exogenous changes in local demand for female labour. Column (3) shows the preferred specification. The coefficients are positive, statistically significant and large in magnitude. For ease of interpretation, the coefficients are scaled to show the effect per 1000 women. A one standard deviation increase in women's potential earnings (46' SEK) increases the risk of assault by 0.246, which corresponds to an effect size of almost 100% of the mean.[21] Furthermore, the effect is positive on both intervals of

[21] Equivalent, an effect size of 0.246 corresponds to about 1.7% of the standard deviation in hospital visits for assault.

relative earnings, which shows that the causal effect of increased potential earnings on domestic violence does not depend on which spouse earns the most.

Table 3: Effect of women's potential earnings on assault

	(1)	(2)	(3)
Women's pot. earnings	0.244***	0.245***	0.246***
	(0.029)	(0.029)	(0.029)
	With intervals		
Husband > Wife	0.251***	0.247***	0.248***
	(0.031)	(0.030)	(0.030)
Husband ≤ Wife	0.233***	0.242***	0.244***
	(0.029)	(0.029)	(0.029)
Observations	7,965,166	7,965,166	7,965,166
No. couples	1,046,867	1,046,867	1,046,867
Mean	0.25	0.25	0.25
	Indicators		
Basic FE	Yes	Yes	Yes
Detailed FE	No	Yes	Yes
Controls	No	No	Yes

Notes: The dependent variable is an indicator of a hospital visit for assault in the current year, scaled by 1000. The top panel reports baseline estimates of the effect of a one-standard-deviation increase women's potential earnings (46' SEK) on assault. The bottom panel allows the effect to differ along two intervals of relative earnings: [-1, 0) and [0, 1]. Basic FE: Municipality, year, age group, education group and cohort. Detailed FE: basic FE plus household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age span and husbands' earnings. Standard errors in parentheses are clustered at the $mun * edugroup * agegroup$ level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

My results are in line with the predictions of male backlash theory, as I show that increases in women's potential earnings, as a proxy for female relative economic empowerment, almost doubles the risk of a hospital visit for assault compared to the mean value. According to the theory of male backlash, the main driver of the backlash response is the violation of traditional gender norms and the stress and anxiety this causes for the husband (Macmillan and Gartner, 1999). Unfortunately administrative data do not allow me to investigate the gender norm channel directly, but in order to investigate the mechanisms behind my results further, I test whether there is a relationship between women's potential earnings and various measures of husbands' destructive behaviour that could indicate increased stress and anxiety. Table 4 shows that increasing women's potential earnings, while keeping the earnings of their husbands constant, increases the risk of the husband visiting a hospital for reasons related to depression, anxiety and stress (denoted "mental instability"), substance abuse, and for himself having been assaulted. Interestingly, I find a very small and only marginally significant, effect for husbands' assault that took place at home or in unspecified locations (my measure of domestic violence). Thus, the increased assault risk of husbands is for assaults taking place in other locations such as bars, sports arenas, public places etc. I also show that women's potential earnings increase the disk of divorce, although, this effect is likely to be endogenous to domestic violence, as assault is a strong predictor of divorce.

Table 4: Mechanisms of a backlash effect: women's potential earnings and husbands' destructive behaviour

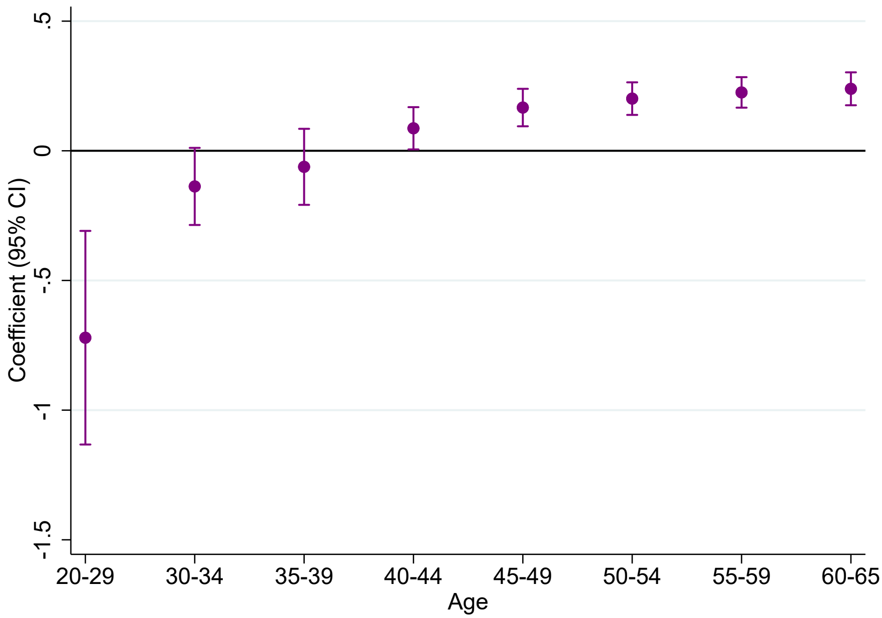
	Mental inst.	Substance abuse	Any assault	HU assault	Divorce
Women's pot. earn.	1.914***	0.321***	0.443***	0.025*	7.446***
	(0.166)	(0.063)	(0.050)	(0.012)	(0.347)
Observations	7,965,166	7,965,166	7,965,166	7,965,166	7,965,166
No. couples	1,046,867	1,046,867	1,046,867	1,046,867	1,046,867
Mean	5.89	1.22	0.50	0.43	17.41
			Indicators		
Basic FE	Yes	Yes	Yes	Yes	Yes
Detailed FE	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is an indicator of husbands' hospital visit for stress, depression or anxiety (1), substance abuse (2), any assault (3), assaults that took place at home or in unspecified locations (4), and couples risk of divorce (5), all scaled by 1000. The table reports estimates of the effect of a one-standard-deviation increase women's potential earnings (46' SEK) on assault. Basic FE: Municipality, year, age group, education group and cohort. Detailed FE: basic FE plus household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age span and husbands' earnings. Standard errors in parentheses are clustered at the $mun * edugroup * agegroup$ level. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

4.3 Heterogeneity analysis

The response to an improved relative economic position may differ depending on the socio-demographic characteristics of the woman. For example, a threat of leaving an abusive spouse may be more credible for younger women who are in a newer relationship, than for those who have already stayed in a possibly toxic relationship for a long time. Figure 5 shows how the effect of increasing women's potential earnings differs by the woman's age. The effect is negative for the younger women, but it grows with age and stabilises at a positive level from around age 40 and onwards. Although the youngest age group is a quite small (about 6% of the total sample) and selective sample of those women who married when they were young, the negative estimate indicates that the effect of improving the outside option of marriage differs with the age of the wife.[22]

Figure 5: Heterogeneity by age

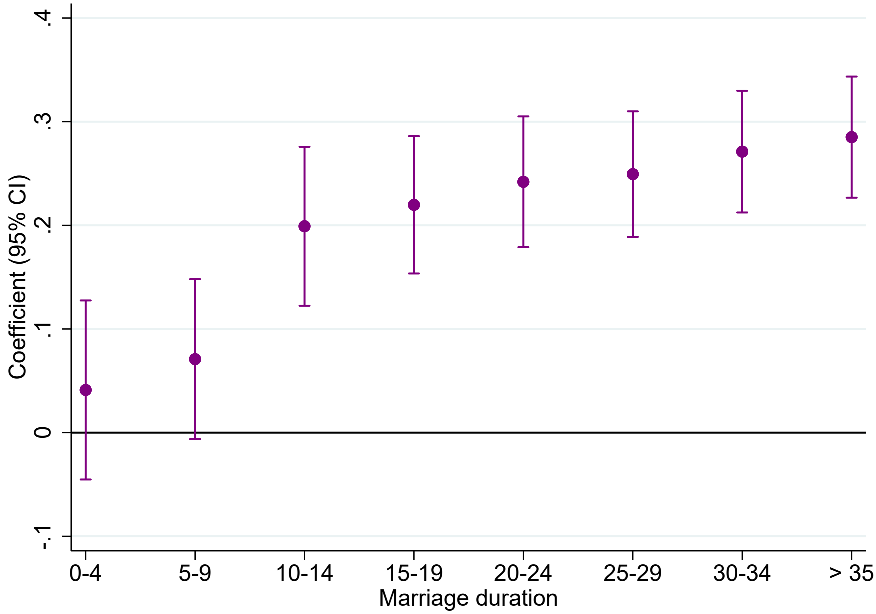


Notes: The figure plots the effect of a one-standard-deviation increase women's potential earnings (46' SEK) on assault, when allowing the effect to differ by age group of the woman. The model controls for indicators for municipality, year, age group, education group, cohort, household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. The figure plots 95% confidence intervals.

[22]Furthermore, Appendix Figure A3 show that these heterogeneous effects exists also for a sample that contains all women, including the non-married, indicating that a selective group of those who marry young is not driving the heterogeneity.

Speculatively, the mechanism behind these differences could be the credibility of a threat of leaving, which is likely to decline with the time a woman chooses to remain in the marriage. In support of this argument, Figure 6 allows the effect to differ by length of the marriage.[23] The figure demonstrates that potential earnings has no effect on the risk of assault for the youngest marriages, but that the backlash effect grows steadily with marriage duration.

Figure 6: Heterogeneity by length of marriage



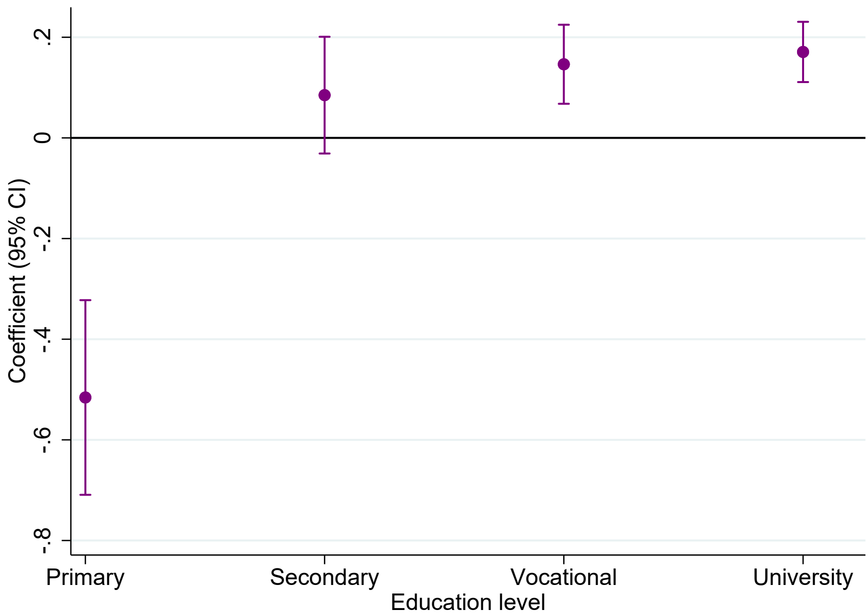
Notes: The figure plots the effect of a one-standard-deviation increase women's potential earnings (46' SEK) on assault, when allowing the effect to differ by how long the couple have been married. The model controls for indicators for municipality, year, age group, education group, cohort, household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. The figure plots 95% confidence intervals.

Education and income level may also influence spouses' possible responses to increased potential earnings. Figure 7 shows that the women with no more than compulsory schooling are less likely to experience assault when their potential earnings increase, whereas the the assault risk is increasing with potential earnings for women

[23]Unfortunately, I can only observe marriage dates after 1968. Everyone who got married earlier than 1968 receives 1968 as their year of marriage.

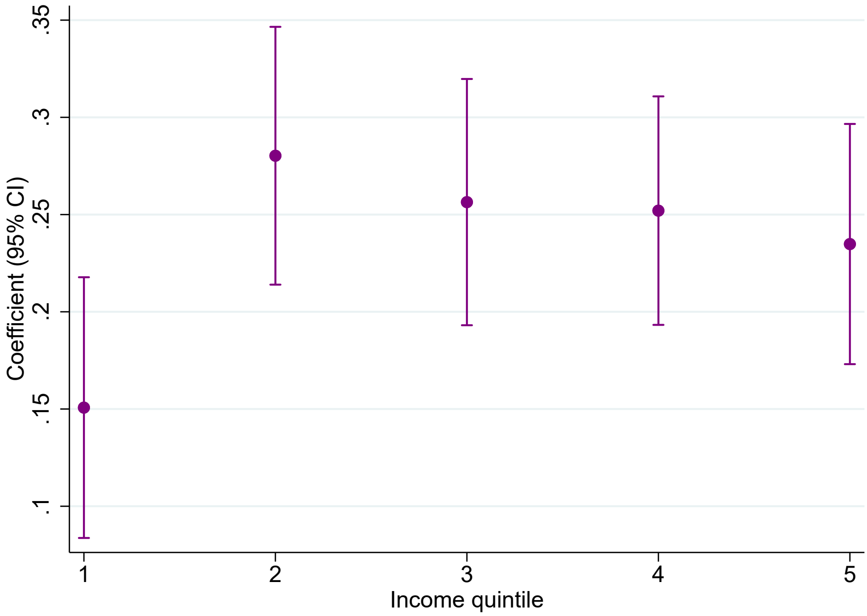
with higher education levels. One possible explanation is that the outside option is binding for the lowest educated, as their labour market opportunities are fewer. Thus, once the outside option improves, their threat of leaving abusive spouses becomes more credible as they now may have the economic possibility to break up the marriage. In contrast, women with higher education levels most likely have the possibility to support themselves economically, hence, staying with an abusive spouse is more of a choice than a necessity. Figure 8 and 9 show that the effect does not differ by realised income or household income, which supports the argument that earnings potential, rather than realised earnings, is what determines the outside option of a marriage.

Figure 7: Heterogeneity by education level



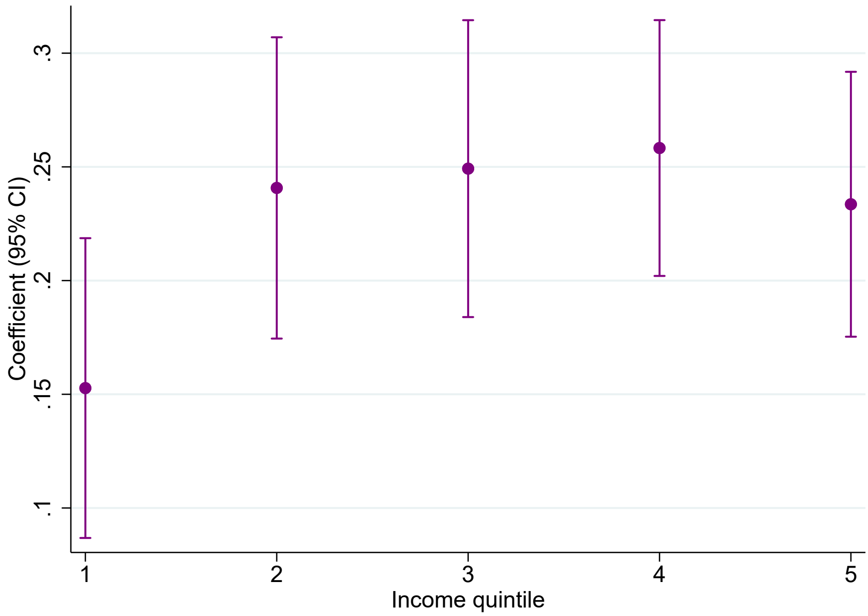
Notes: The figure plots the effect of a one-standard-deviation increase women's potential earnings (46' SEK) on assault, when allowing the effect to differ by education group of the woman. The model controls for indicators for municipality, year, age group, education group, cohort, household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. The figure plots 95% confidence intervals.

Figure 8: Heterogeneity by income level



Notes: The figure plots the effect of a one-standard-deviation increase women's potential earnings (46' SEK) on assault, when allowing the effect to differ by income quintile of the woman. Income includes labour earnings and social transfers. The model controls for indicators for municipality, year, age group, education group, cohort, household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. The figure plots 95% confidence intervals.

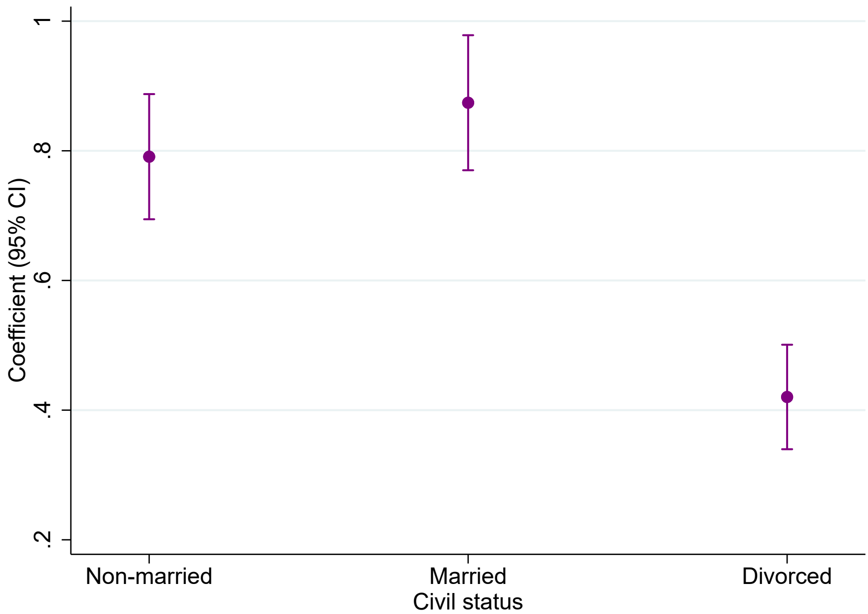
Figure 9: Heterogeneity by household income



Notes: The figure plots the effect of a one-standard-deviation increase women's potential earnings (46' SEK) on assault, when allowing the effect to differ by household income quintile. The model controls for indicators for municipality, year, age group, education group, cohort, household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. The figure plots 95% confidence intervals.

Finally, the effect of increasing women's potential earnings may differ depending on civil status. My main results contain only married women, as those are the only ones for which I can be sure that they have an intimate partner. However, being in a relationship without being married, either domestic or living apart, is very common in Sweden. To ensure that my results are not driven by a selective population of those who marry, I investigate whether the effect differs by civil status. For these estimations, I use a larger sample containing all women in my dataset, regardless of marriage status. That the perpetrator of an assault is an intimate partner becomes a stronger assumption to make for this population, as I cannot observe whether they have an intimate partner or not. However, as most female assault are done by an intimate partner, I believe measurement error is still small. Figure 10 shows that increasing women's potential earnings increase the risk of assault, regardless of whether the woman is non-married, married or divorced (although the effect is smaller in magnitude for the divorced).[24]

Figure 10: Heterogeneity by civil status



Notes: The figure plots the effect of a one-standard-deviation increase women's potential earnings on assault, when allowing the effect to differ by civil status of the woman. The model controls for indicators for municipality, year, age group, education group, cohort, household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. The figure plots 95% confidence intervals. The sample used for this model is larger than the main sample, as it also includes non-married and divorced women.

5 Robustness

Even though the hospital visits for assault are reported by medical personnel, there is still the concern that women at the hospital might not truthfully state the causes of their injuries. If this misreporting is non-random and related to potential earnings,

[24]Instead of controlling for husbands' earnings linearly, I include indicators for husbands earnings decile, or for not having a husband. Appendix Figure A2 show the risk of assault by husbands' earnings decile, and Appendix Table A1 show the effect of women's potential earnings for the larger sample that includes also those women who are not married. Appendix Figures A3 and A4 show that the same pattern of heterogeneous effects by age and education level exists also for this sample.

it would invalidate my results as I cannot distinguish between a reduction in assaults and a switch in reporting behaviour. I address this concern by investigating the relationship between potential earnings and hospital visits for accidents, as a misreported assault would likely be coded as an accident. To isolate those accidents that are most likely to be hidden assaults, I only use accidents that happened at home or in unspecified locations, with a similar main diagnosis to that of an assault. Column (1) in Table 5 shows that potential earnings has a positive, but small, effect on the risk of visiting a hospital for these types of accidents. This positive effect implies that some misreporting might take place, and, for this reason, I replicate my main analysis using an outcome measure that consists of both assaults and similar accidents (thereby capturing all assaults, including the misreported ones). Reassuringly, the estimates in column (2) depict the same relationship as the main results of Table 3, but are of larger magnitude. Thus, if anything, possible misreporting attenuates my results, and I can conclude that it does not pose a threat to my study. Finally, my results are not due to the low frequency of assaults or data sparsity at the tails of the potential earnings distribution. Column (3) shows the results of a placebo test in which I show that no relationship exists between appendix complications, which are about as common as assaults in my sample, and women's potential earnings.

Table 5: Placebo tests: accidents and appendix complications

	(1)	(2)	(3)
	Accidents	Accidents + assaults	Appendix
Women's pot. earnings	0.113***	0.359***	-0.044
	(0.032)	(0.040)	(0.040)
Observations	7,965,166	7,965,166	7,965,166
No. couples	1,046,867	1,046,867	1,046,867
Mean	0.65	0.90	0.88
		Indicators	
Detailed FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes

Notes: The dependent variable in column (1) an indicator for visiting a hospital for an accident with similar characteristics as the assaults. The dependent variable of column (2) is these accidents and assaults combined. Column (3) show the effect of hospital visits for appendix complications, as a placebo test. All coefficients are scaled by 1000. The table reports estimates of the effect of a one-standard-deviation increase women's potential earnings (46' SEK) on assault. Basic FE: Municipality, year, age group, education group and cohort. Detailed FE: basic FE plus household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. Standard errors in parentheses are clustered at the $mun * edugroup * agegroup$ level. * p<0.05, ** p<0.01, *** p<0.001.

My results hold for measuring domestic violence as assault that took place at home only, and are not sensitive to excluding the unemployed. I control for non-random attrition via divorce by estimating the model on a balanced sample where I include only those couples who are observed every year of my study. The results for this subgroup are slightly weaker than the main results, as would be expected since some of the "low-quality" marriages opt out of my full dataset over time. My results are also robust to controlling even more flexibly for household income using decile fixed effects and a linear trend, and to excluding the top 1% and zero-earners from my sample. Table 6 shows a summary of all these robustness checks.

Table 6: Robustness checks

	Main effects	At home	Excl. unempl.	Balanced	Finer HH FE	Excl. zero & top 1%
Women's pot. earn.	0.246***	0.040***	0.219***	0.062**	0.243***	0.221***
	(0.029)	(0.011)	(0.031)	(0.020)	(0.028)	(0.031)
Observations	7,965,166	7,965,166	6,604,002	4,589,827	7,965,166	6,597,080
No. couples	1,046,867	1,046,867	989,798	417,257	1,046,867	960,078
Mean	0.25	0.06	0.20	0.11	0.25	0.21
	Indicators					
Detailed FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The dependent variable is female hospital visits for assault, scaled by 1000. Column (1) show the main effects. Column (2) show the effects for hospital visits for assaults that took place only at home. Column(3) excludes all couples where either the wife or the husband received any unemployment benefits during the current year. Column (4) reports the effect on a balanced sample of couples who I can observe full the full 11 years. Column (5) includes decile fixed effects and linear trends of household income. Column (7) excludes zero-earners and the top 1% of the earnings distribution. The table reports estimates of the effect of a one standard deviation increase women's potential earnings (46' SEK) on assault. Basic FE: Municipality, year, age group, education group and cohort. Detailed FE: basic FE plus household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. Standard errors in parentheses are clustered at the $mun * edugroup * agegroup$ level. * p<0.05, ** p<0.01, *** p<0.001.

5.1 Replication of Aizer (2010) and using police reports

My findings differ from those of Aizer (2010), who finds a significant negative effect of a reduction in the gender wage gap on aggregate levels of hospitalisations for assault. One contributing factor to the difference between my results and those of Aizer (2010) is that she restricts her dataset to contain only hospital visits by women with no more than high school education. My heterogeneity analysis reveals that increasing potential earnings for the lowest educated women reduces their assault risk. Thus, my results do not contradict those of Aizer (2010) if I restrict my sample to contain a similar study population as hers. In contrast, my study contributes to Aizer (2010), as I show that the effect of an increased relative economic position differs depending on the sub-group of the population of women.

Furthermore, I replicate the findings of Aizer (2010) using a dataset and model specification that mimics hers as closely as possible.[25] Table 7 shows the results of this replication exercise, where column (1) shows the results from using her model with the outcome measure most similar to hers (all overnight hospitalisations for assault), column (2) contains her model with the outcome measure of my study (hospital visits for assaults that took place at home or in unspecified locations), column (3) contains an aggregate version of my model using police reports for domestic violence as the dependent variable, and column (4) show the results using my individual level model with the outcome variable of Aizer (2010). The results show that I can replicate Aizer's findings on aggregate Swedish data (containing women of all education levels), but that our results differ if the analysis is done on an individual level dataset. The difference in our results are not due to any difference in our preferred outcome; the estimates in column (2) are still negative and the estimates in column (4) are positive. The results for police reports are positive, but small and insignificant. Thus, using the model and aggregate data set-up of Aizer (2010) yields negative results for both hers and my outcome measure, whereas my model and individual level data yields positive results for both my outcome and that of Aizer (2010). These results indicate that an effect measured on an aggregate level can mask what is really happening on the individual level, as it fails to account for heterogeneous effects in the population.

[25] I aggregate my data to municipality level, and construct a measure of the potential gender wage gap that is as close to hers as possible.

Table 7: Replication of Aizer (2010)

	Hospitalisation	Hosp. visits	Police reports	Hospitalisation
Pot. relative earn.	-0.162**	-0.086		
	(0.050)	(0.170)		
Women's pot. earn.			0.134	0.031**
			(0.962)	(0.011)
Observations	790	790	2,880	7,965,166
Mean	0.11	0.96	3.76	0.05
	Indicators			
Municipality FE	Yes	Yes	Yes	See T3
Year FE	Yes	Yes	Yes	See T3
Controls	Yes	Yes	Yes	See T3

Notes: The dependent variable in column (1) is the natural log of female in-patient hospitalizations for assault, column (2) shows the natural log of female hospital visits for assaults that took place at home or in unspecified location, and column (3) the number of police reports filed for domestic abuse, per municipality and year. The dependent variable in column (4) is an indicator of overnight hospitalizations for assault in the current year. Potential relative earnings are calculated following Aizer (2010). The effect is estimated for a one-standard-deviation change. All regressions include fixed effects for municipality and year. Controls for column (1)-(3) are natural log of municipality mean income, natural log of municipality population, natural log of female homicides and lagged natural log of female hospitalizations or lagged police reports. Municipalities with a female population below 10,000 are excluded from the analysis in column (1) and (2), following Aizer (2010). Standard errors in parentheses are clustered at municipality level. Controls for column (4) are the same as in column (3) of Table 3. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

6 Conclusion

This paper estimates the effect of female economic empowerment on domestic violence, using high-quality data from Swedish administrative registers. I measure domestic violence using individual hospital visits for assault. Distinguishing violence changes from reporting behaviour is of utter importance with an outcome as sensitive as domestic violence, especially as the propensity to report a violent incident is likely increasing with empowerment. Hospital data is reported by medical personnel, hence, it suffers from less selective reporting bias compared to other measures. Furthermore, information on hospital visits for accidents allows me to investigate possible misreporting at the hospital, which enables me to conclude that non-random misreporting does not pose a threat to the validity of my results.

Descriptively, I show that the relationship between spouses' realised relative earnings and risk of assault is U-shaped. Violence is increasing with earnings inequalities in both directions, and the sign of the association differs depending on which spouse is the main breadwinner. The U-shaped relationship is an interesting finding, but it does not depict a causal relationship. As spouses select into marriage in an assortative way, which may be a function of earnings as well as both underlying propensity for violence and for staying in a violent relationship, the characteristics of who marries whom may be partly driving the relationship between relative earnings and domestic violence. Furthermore, the outside option of a marriage is determined by earnings potentials rather than realised earnings, and realised earnings may be endogenous with respect to violence and unobserved characteristics that correlates with violence.

To account for these empirical challenges, I proxy women's relative economic empowerment with a demand-driven measure of women's potential earnings, which reflects the outside option of the marriage and captures earnings variation that is not endogenous to domestic violence. In all estimations I keep the earnings of the husbands constant, in order to identify variation in the wives' relative economic empowerment that is driven only by changes in female labour demand.

I show that increasing women's potential earnings has a positive causal effect on the risk of domestic violence, regardless of which spouse earns more than the other. In addition, I show that women's potential earnings increase the husbands' risk of visiting a hospital for stress, anxiety, substance abuse and assault. Thus, the mechanisms are in line with a male backlash response to improved female economic independence. This result is perhaps in contrast to a priori expectations of what we should find in a gender-equal country like Sweden, and it indicates that traditional gender norms may play an important role in determining the relationship between female economic empowerment and domestic violence.

Heterogeneity analysis show that the backlash response to women's potential earnings is increasing with age of the woman and duration of the marriage. Likewise, for the women with no more than high school education increases in potential earnings reduce the risk of assault, but for women of higher education levels potential earnings increase the risk. These results indicate that the women for whom a change in potential earnings actually affects the credibility of their threat of leaving, the results are in line with the predictions of bargaining power theory. But for the women who may have the economic possibility to leave, but still do not, the effects are in line with male backlash theory. An important implication of these results is that women from typically less vulnerable groups still experience a high risk of a backlash response to an improved relative economic position within the marriage.

More research is needed in this area as domestic violence is a significant issue for public health. My effect sizes are substantial, and suggest a high risk of a backlash response to women' improved financial independence within a marriage. Furthermore, I show that the risk of backlash is higher for older women and highly educated women, which are demographic groups that are usually overlooked in the discussion on how to best combat domestic violence. Policy makers should consider how to best reach these women, who may already have the economic opportunity to leave their abusive spouse but who still choose to remain in the relationship. Furthermore, along with providing shelter and legal support for assaulted women, policy makers should consider policies targeted to promote less traditional gender norms in boys and girls, preferably early in their lives.

References

Aizer, A. (2010). The gender wage gap and domestic violence. *American Economic Review*, 100(4):1847–59.

Akerlof, G. A. and Kranton, R. E. (2000). Economics and identity. *The Quarterly Journal of Economics*, 115(3):715–753.

Alonso-Borrego, C. and Carrasco, R. (2017). Employment and the risk of domestic violence: does the breadwinner's gender matter? *Applied Economics*, 49(50):5074–5091.

Anderberg, D., Rainer, H., Wadsworth, J., and Wilson, T. (2016). Unemployment and domestic violence: Theory and evidence. *The Economic Journal*, 126(597):1947–1979.

Atkinson, M. P., Greenstein, T. N., and Lang, M. M. (2005). For women, breadwinning can be dangerous: Gendered resource theory and wife abuse. *Journal of Marriage and Family*, 67(5):1137–1148.

Bartik, T. J. (1991). *Who benefits from state and local economic development policies?* WE Upjohn Institute for Employment Research.

Beland, L.-P. and Brent, D. A. (2018). Traffic and crime. *Journal of Public Economics*, 160:96–116.

Bertrand, M., Kamenica, E., and Pan, J. (2015). Gender identity and relative income within households. *The Quarterly Journal of Economics*, 130(2):571–614.

Bhalotra, S. R., Kambhampati, U. S., Rawlings, S., and Siddique, Z. (2018). Intimate partner violence and the business cycle. Working Paper 11274.

Brassiolo, P. (2016). Domestic violence and divorce law: When divorce threats become credible. *Journal of Labor Economics*, 34(2):443–477.

Card, D. and Dahl, G. B. (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics*, 126(1):103–143.

Cesur, R. and Sabia, J. J. (2016). When war comes home: The effect of combat service on domestic violence. *Review of Economics and Statistics*, 98(2):209–225.

Chin, Y.-M. (2012). Male backlash, bargaining, or exposure reduction?: women's working status and physical spousal violence in india. *Journal of population Economics*, 25(1):175–200.

Cools, S. and Kotsadam, A. (2017). Resources and intimate partner violence in subsaharan africa. *World Development*, 95:211–230.

Ellsberg, M., Heise, L., Pena, R., Agurto, S., and Winkvist, A. (2001). Researching domestic violence against women: methodological and ethical considerations. *Studies in family planning*, 32(1):1–16.

Farmer, A. and Tiefenthaler, J. (1997). An economic analysis of domestic violence. *Review of social Economy*, 55(3):337–358.

Frenzel, A. (2014). *Brott i nära relationer: en nationell kartläggning*. Brottsförebyggande rådet (BRÅ).

García-Moreno, C., Pallitto, C., Devries, K., Stöckl, H., Watts, C., and Abrahams, N. (2013). *Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence*. World Health Organization.

Goldsmith-Pinkham, P., Sorkin, I., and Swift, H. (2018). Bartik instruments: What, when, why, and how. Working Paper 24408, National Bureau of Economic Research.

González, L., Rodríguez-Planas, N., et al. (2018). Gender norms and intimate partner violence. Working Paper 1620, Economics Working Paper Series.

Guarnieri, E., Rainer, H., et al. (2018). Female empowerment and male backlash. Working Paper 7009, CESifo Group Munich.

Heath, R. (2014). Women's access to labor market opportunities, control of household resources, and domestic violence: Evidence from bangladesh. *World Development*, 57:32–46.

Heise, L. L. and Kotsadam, A. (2015). Cross-national and multilevel correlates of partner violence: an analysis of data from population-based surveys. *The Lancet Global Health*, 3(6):e332–e340.

Hoeffler, A. (2017). What are the costs of violence? *Politics, Philosophy & Economics*, 16(4):422–445.

Hornung, C. A., McCullough, B. C., and Sugimoto, T. (1981). Status relationships in marriage: Risk factors in spouse abuse. *Journal of Marriage and the Family*, pages 675–692.

Iyer, L., Mani, A., Mishra, P., and Topalova, P. (2012). The power of political voice: women's political representation and crime in india. *American Economic Journal: Applied Economics*, 4(4):165–93.

La Mattina, G. (2017). Civil conflict, domestic violence and intra-household bargaining in post-genocide rwanda. *Journal of Development Economics*, 124:168–198.

Lindo, J. M., Schaller, J., and Hansen, B. (2018). Caution! men not at work: Gender-specific labor market conditions and child maltreatment. *Journal of Public Economics*, 163:77–98.

Macmillan, R. and Gartner, R. (1999). When she brings home the bacon: Labor-force participation and the risk of spousal violence against women. *Journal of Marriage and the Family*, pages 947–958.

Pollak, R. A. (2004). An intergenerational model of domestic violence. *Journal of Population Economics*, 17(2):311–329.

Pollak, R. A. (2005). Bargaining power in marriage: Earnings, wage rates and household production. Working Paper 11239, National Bureau of Economic Research.

Stevenson, B. and Wolfers, J. (2006). Bargaining in the shadow of the law: Divorce laws and family distress. *The Quarterly Journal of Economics*, 121(1):267–288.

Svec, J. and Andic, T. (2018). Cooperative decision-making and intimate partner violence in peru. *Population and development review*, 44(1):63–85.

Tur-Prats, A. (2017). Unemployment and intimate-partner violence: A gender-identity approach. Working Paper 963, Barcelona Graduate School of Economics.

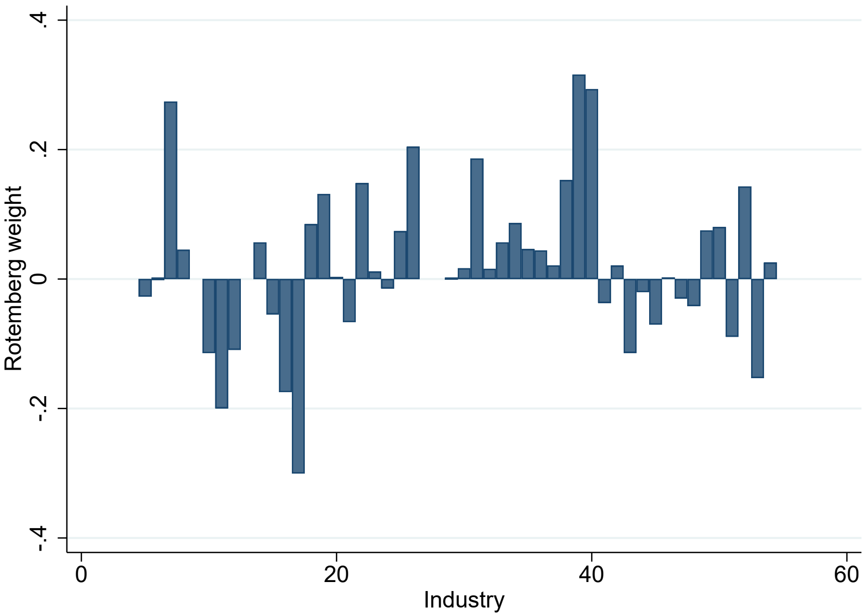
Appendix

Table A1: Descriptive relationship between relative earnings and assault

	(1)	(2)	(3)	(4)
Relative earnings	-0.003	-0.022**	-0.022**	0.016
	(0.007)	(0.008)	(0.008)	(0.014)
	With intervals			
Husband > Wife	-0.152***	-0.111***	-0.111***	0.029
	(0.014)	(0.014)	(0.014)	(0.023)
Husband $\leq$ Wife	0.104***	0.045***	0.046***	0.006
	(0.010)	(0.011)	(0.011)	(0.019)
Observations	7,965,166	7,965,166	7,965,166	7,894,612
No. couples	1,046,867	1,046,867	1,046,867	976,313
Mean	0.25	0.25	0.25	0.24
	Indicators			
Basic FE	Yes	Yes	Yes	No
Detailed FE	No	Yes	Yes	No
ID FE	No	No	No	Yes
Controls	No	No	Yes	Yes

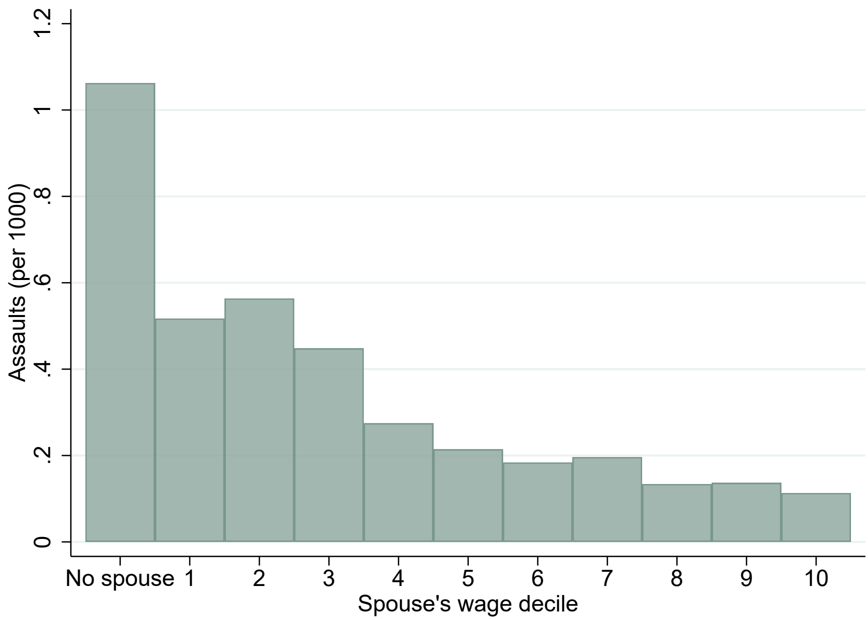
Notes: The dependent variable is an indicator of a hospital visit for assault in the current year, scaled by 1000. The associations are estimated for a one-standard-deviation increase in relative earnings (0.53). The top panel reports baseline estimates of the effect of a change in relative earnings on assault. The bottom panel allows the effect to differ along two intervals: [-1, 0) and [0, 1]. Basic FE: Municipality, year, age group, education group and cohort. Detailed FE: basic FE plus household income quintile and spouse's education group. Controls are municipality level of police reports and spousal age difference. Standard errors in parentheses are clustered at the $mun * edugroup * agegroup$ level. * p<0.05, ** p<0.01, *** p<0.001.

Figure A1: Rotemberg weights for each industry



Notes: The figure plots the Rotemberg weights for each industry, which quantifies the contribution of each industry to the identification. These are constructed using the Stata -bartik weight- command, as outlined in the supplemental material of Goldsmith-Pinkham et al (2018). The endogenous variable is relative earnings, and the outcome is hospital visits for assault. By construction, the weights sum to one.

Figure A2: Assault risk by spouse's wage decile: all women



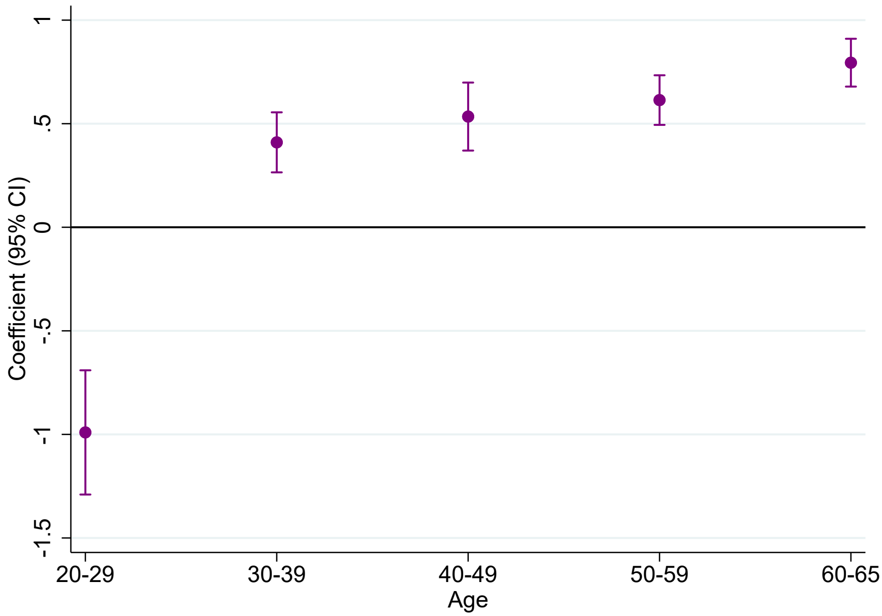
Notes: The figure plots mean risk of visiting a hospital for assault (that took place at home or in unspecified locations) by husbands' income decile, and for unmarried women.

Table A2: Effect of women's potential earnings on assault: all women

	(1)	(2)	(3)
Women's pot. earnings	0.705***	0.781***	0.782***
	(0.047)	(0.047)	(0.047)
Observations	21491632	21491157	21491157
No. couples	2,422,648	2,422,648	2,422,648
Mean	0.76	0.76	0.76
		Indicators	
Basic FE	Yes	Yes	Yes
Detailed FE	No	Yes	Yes
Controls	No	No	Yes

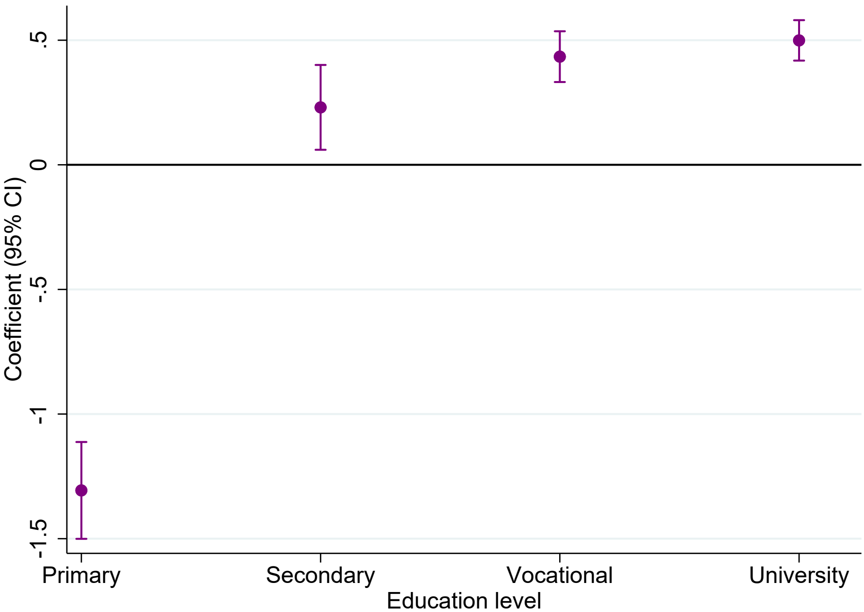
Notes: The dependent variable is an indicator of a hospital visit for assault in the current year, scaled by 1000. The top panel reports baseline estimates of the effect of a one-standard-deviation increase women's potential earnings on assault. The bottom panel allows the effect to differ along two intervals of relative earnings: [-1, 0) and [0, 1]. Basic FE: Municipality, year, age group, education group and cohort. Detailed FE: basic FE plus household income quintile and spouse's education group. Controls are municipality level of police reports and spousal age difference and indicators for husbands' income decile or for being unmarried. Standard errors in parentheses are clustered at the $mun*edugroup*agegroup$ level. * p<0.05, ** p<0.01, *** p<0.001.

Figure A3: Heterogeneity by age: all women



Notes: The figure plots the effect of a one-standard-deviationon increase women's potential earnings on assault, when allowing the effect to differ by age group of the woman. The model controls for indicators for municipality, year, age group, education group, cohort, household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. The figure plots 95% confidence intervals. The sample used for this model is larger than the main sample, as it also includes non-married and divorced women.

Figure A4: Heterogeneity by education level: all women



Notes: The figure plots the effect of a one-standard-deviation increase women's potential earnings on assault, when allowing the effect to differ by education group of the woman. The model controls for indicators for municipality, year, age group, education group, cohort, household income quintile and spouse's education group. Controls are municipality level of police reports, spousal age difference and husbands' earnings. The figure plots 95% confidence intervals. The sample used for this model is larger than the main sample, as it also includes non-married and divorced women.

Lund Economic Studies

1. Guy Arvidsson	Bidrag till teorin för verkningarna av räntevariationer, 1962
2. Björn Thalberg	A Trade Cycle Analysis. Extensions of the Goodwin Model, 1966
3. Bengt Höglund	Modell och observationer. En studie av empirisk anknytning och aggregation för en linjär produktionsmodell, 1968
4. Alf Carling	Industrins struktur och konkurrensförhållanden, 1968
5. Tony Hagström	Kreditmarknadens struktur och funktionssätt, 1968
6. Göran Skogh	Straffrätt och samhällsekonomi, 1973
7. Ulf Jakobsson och Göran Norman	Inkomstbeskattningen i den ekonomiska politiken. En kvantitativ analys av systemet för personlig inkomstbeskattning 1952-71, 1974
8. Eskil Wadensjö	Immigration och samhällsekonomi. Immigrationens ekonomiska orsaker och effekter, 1973
9. Rögnvaldur Hannesson	Economics of Fisheries. Some Problems of Efficiency, 1974
10. Charles Stuart	Search and the Organization of Marketplaces, 1975
11. S Enone Metuge	An Input-Output Study of the Structure and Resource Use in the Cameroon Economy, 1976
12. Bengt Jönsson	Cost-Benefit Analysis in Public Health and Medical Care, 1976
13. Agneta Kruse och Ann-Charlotte Ståhlberg	Effekter av ATP - en samhällsekonomisk studie, 1977
14. Krister Hjalte	Sjörestaureringens ekonomi, 1977
15. Lars-Gunnar Svensson	Social Justice and Fair Distributions, 1977
16. Curt Wells	Optimal Fiscal and Monetary Policy - Experiments with an Econometric Model of Sweden, 1978
17. Karl Lidgren	Dryckesförpackningar och miljöpolitik - En studie av styrmedel, 1978
18. Mats Lundahl	Peasants and Poverty. A Study of Haiti, London, 1979
19. Inga Persson-Tanimura	Studier kring arbetsmarknad och information, 1980
20. Bengt Turner	Hyressättning på bostadsmarknaden - Från hyresreglering till bruksvärdesprövning, Stockholm 1979
21. Ingemar Hansson	Market Adjustment and Investment Determination. A Theoretical Analysis of the Firm and the Industry, Stockholm 1981
22. Daniel Boda Ndlela	Dualism in the Rhodesian Colonial Economy, 1981
23. Tom Alberts	Agrarian Reform and Rural Poverty: A Case Study of Peru, 1981
24. Björn Lindgren	Costs of Illness in Sweden 1964-75, 1981

25. Göte Hansson	Social Clauses and International Trade. An Economic Analysis of Labour Standards in Trade Policy, 1981
26. Noman Kanafani	Oil and Development. A Case Study of Iraq, 1982
27. Jan Ekberg	Inkomsteffekter av invandring, 1983
28. Stefan Hedlund	Crisis in Soviet Agriculture?, 1983
29. Ann-Marie Pålsson	Hushållen och kreditpolitiken. En studie av kreditrestriktioners effekt på hushållens konsumtion, sparande och konsumtionsmönster, 1983
30. Lennart Petersson	Svensk utrikeshandel, 1871-1980. En studie i den intraindustriella handelns framväxt, 1984
31. Bengt Assarsson	Inflation and Relative Prices in an Open Economy, 1984
32. Claudio Vedovato	Politics, Foreign Trade and Economic Development in the Dominican Republic, 1985
33. Knut Ödegaard	Cash Crop versus Food Crop Production in Tanzania: An Assessment of the Major Post-Colonial Trends, 1985
34. Vassilios Vlachos	Temporära lönesubventioner. En studie av ett arbetsmarknadspolitiskt medel, 1985
35. Stig Tegle	Part-Time Employment. An Economic Analysis of Weekly Working Hours in Sweden 1963-1982, 1985
36. Peter Stenkula	Tre studier över resursanvändningen i högskolan, 1985
37. Carl Hampus Lyttkens	Swedish Work Environment Policy. An Economic Analysis, 1985
38. Per-Olof Bjuggren	A Transaction Cost Approach to Vertical Integration: The Case of Swedish Pulp and Paper Industry, 1985
39. Jan Petersson	Erik Lindahl och Stockholmsskolans dynamiska metod, 1987
40. Yves Bourdet	International Integration, Market Structure and Prices. A Case Study of the West-European Passenger Car Industry, 1987
41. Krister Andersson and Erik Norrman	Capital Taxation and Neutrality. A study of tax wedges with special reference to Sweden, 1987
42. Tohmas Karlsson	A Macroeconomic Disequilibrium Model. An Econometric Study of the Swedish Business Sector 1970-84, 1987
43. Rosemary Vargas-Lundius	Peasants in Distress. Poverty and Unemployment in the Dominican Republic, 1989
44. Lena Ekelund Axelson	Structural Changes in the Swedish Marketing of Vegetables, 1991
45. Elias Kazarian	Finance and Economic Development: Islamic Banking in Egypt, 1991
46. Anders Danielson	Public Sector Expansion and Economic Development. The Sources and Consequences of Development Finance in Jamaica 1962-84, 1991

47. Johan Torstensson	Factor Endowments, Product Differentiation, and International Trade, 1992
48. Tarmo Haavisto	Money and Economic Activity in Finland, 1866-1985, 1992
49. Ulf Grönkvist	Economic Methodology. Patterns of Reasoning and the Structure of Theories, 1992
50. Evelyne Hangali Maje	Monetization, Financial Development and the Demand for Money, 1992
51. Michael Bergman	Essays on Economic Fluctuations, 1992
52. Flora Mndeme Musonda	Development Strategy and Manufactured Exports in Tanzania, 1992
53. Håkan J. Holm	Complexity in Economic Theory. An Automata Theoretical Approach, 1993
54. Klas Fregert	Wage Contracts, Policy Regimes and Business Cycles. A Contractual History of Sweden 1908-90, 1994
55. Per Frennberg	Essays on Stock Price Behaviour in Sweden, 1994
56. Lisbeth Hellvin	Trade and Specialization in Asia, 1994
57. Sören Höjgård	Long-term Unemployment in a Full Employment Economy, 1994
58. Karolina Ekholm	Multinational Production and Trade in Technological Knowledge, 1995
59. Fredrik Andersson	Essays in the Economics of Asymmetric Information, 1995
60. Rikard Althin	Essays on the Measurement of Producer Performance, 1995
61. Lars Nordén	Empirical Studies of the Market Microstructure on the Swedish Stock Exchange, 1996
62. Kristian Bolin	An Economic Analysis of Marriage and Divorce, 1996
63. Fredrik Sjöholm	R&D, International Spillovers and Productivity Growth, 1997
64. Hossein Asgharian	Essays on Capital Structure, 1997
65. Hans Falck	Aid and Economic Performance - The Case of Tanzania, 1997
66. Bengt Liljas	The Demand for Health and the Contingent Valuation Method, 1997
67. Lars Pålsson Syll	Utility Theory and Structural Analysis, 1997
68. Richard Henricsson	Time Varying Parameters in Exchange Rate Models, 1997
69. Peter Hördahl	Financial Volatility and Time-Varying Risk Premia, 1997
70. Lars Nilsson	Essays on North-South Trade, 1997
71. Fredrik Berggren	Essays on the Demand for Alcohol in Sweden - Review and Applied Demand Studies, 1998
72. Henrik Braconier	Essays on R&D, Technology and Growth, 1998
73. Jerker Lundbäck	Essays on Trade, Growth and Exchange Rates, 1998
74. Dan Anderberg	Essays on Pensions and Information, 1998

75. P. Göran T. Hägg	An Institutional Analysis of Insurance Regulation – The Case of Sweden, 1998
76. Hans-Peter Bermin	Essays on Lookback and Barrier Options - A Malliavin Calculus Approach, 1998
77. Kristian Nilsson	Essays on Exchange Rates, Exports and Growth in Developing Countries, 1998
78. Peter Jochumzen	Essays on Econometric Theory, 1998
79. Lars Behrenz	Essays on the Employment Service and Employers' Recruitment Behaviour, 1998
80. Paul Nystedt	Economic Aspects of Ageing, 1998
81. Rasha M. Torstensson	Empirical Studies in Trade, Integration and Growth, 1999
82. Mattias Ganslandt	Games and Markets - Essays on Communication, Coordination and Multi-Market Competition, 1999
83. Carl-Johan Belfrage	Essays on Interest Groups and Trade Policy, 1999
84. Dan-Olof Rooth	Refugee Immigrants in Sweden - Educational Investments and Labour Market Integration, 1999
85. Karin Olofsdotter	Market Structure and Integration: Essays on Trade, Specialisation and Foreign Direct Investment, 1999
86. Katarina Steen Carlsson	Equality of Access in Health Care, 1999
87. Peter Martinsson	Stated preference methods and empirical analyses of equity in health, 2000
88. Klas Bergenheim	Essays on Pharmaceutical R&D, 2000
89. Hanna Norberg	Empirical Essays on Regional Specialization and Trade in Sweden, 2000
90. Åsa Hansson	Limits of Tax Policy, 2000
91. Hans Byström	Essays on Financial Markets, 2000
92. Henrik Amilon	Essays on Financial Models, 2000
93. Mattias Lundbäck	Asymmetric Information and The Production of Health, 2000
94. Jesper Hansson	Macroeconometric Studies of Private Consumption, Government Debt and Real Exchange Rates, 2001
95. Jonas Månsson	Essays on: Application of Cross Sectional Efficiency Analysis, 2001
96. Mattias Persson	Portfolio Selection and the Analysis of Risk and Time Diversification, 2001
97. Pontus Hansson	Economic Growth and Fiscal Policy, 2002
98. Joakim Gullstrand	Splitting and Measuring Intra-Industry Trade, 2002
99. Birger Nilsson	International Asset Pricing, Diversification and Links between National Stock Markets, 2002
100. Andreas Graflund	Financial Applications of Markov Chain Monte Carlo Methods, 2002

101. Therése Hindman Persson	Economic Analyses of Drinking Water and Sanitation in Developing Countries, 2002
102. Göran Hjelm	Macroeconomic Studies on Fiscal Policy and Real Exchange Rates, 2002
103. Klas Rikner	Sickness Insurance: Design and Behavior, 2002
104. Thomas Ericson	Essays on the Acquisition of Skills in Teams, 2002
105. Thomas Elger	Empirical Studies on the Demand for Monetary Services in the UK, 2002
106. Helena Johansson	International Competition, Productivity and Regional Spillovers, 2003
107. Fredrik Gallo	Explorations in the New Economic Geography, 2003
108. Susanna Thede	Essays on Endogenous Trade Policies, 2003
109. Fredrik CA Andersson	Interest Groups and Government Policy, A Political Economy Analysis, 2003
110. Petter Lundborg	Risky Health Behaviour among Adolescents, 2003
111. Martin W Johansson	Essays on Empirical Macroeconomics, 2003
112. Joakim Ekstrand	Currency Markets - Equilibrium and Expectations, 2003
113. Ingemar Bengtsson	Central bank power: a matter of coordination rather than money supply, 2003
114. Lars Pira	Staples, Institutions and Growth: Competitiveness of Guatemalan Exports 1524-1945, 2003
115. Andreas Bergh	Distributive Justice and the Welfare State, 2003
116. Staffan Waldo	Efficiency in Education - A Multilevel Analysis, 2003
117. Mikael Stenkula	Essays on Network Effects and Money, 2004
118. Catharina Hjortsberg	Health care utilisation in a developing country -the case of Zambia, 2004
119. Henrik Degrér	Empirical Essays on Financial Economics, 2004
120. Mårten Wallette	Temporary Jobs in Sweden: Incidence, Exit, and On-the-Job Training, 2004
121. Tommy Andersson	Essays on Nonlinear Pricing and Welfare, 2004
122. Kristian Sundström	Moral Hazard and Insurance: Optimality, Risk and Preferences, 2004
123. Pär Torstensson	Essays on Bargaining and Social Choice, 2004
124. Frederik Lundtofte	Essays on Incomplete Information in Financial Markets, 2005
125. Kristian Jönsson	Essays on Fiscal Policy, Private Consumption and Non-Stationary Panel Data, 2005
126. Henrik Andersson	Willingness to Pay for a Reduction in Road Mortality Risk: Evidence from Sweden, 2005

127. Björn Ekman	Essays on International Health Economics: The Role of Health Insurance in Health Care Financing in Low- and Middle-Income Countries, 2005
128. Ulf G Erlandsson	Markov Regime Switching in Economic Time Series, 2005
129. Joakim Westerlund	Essays on Panel Cointegration, 2005
130. Lena Hiselius	External costs of transports imposed on neighbours and fellow road users, 2005
131. Ludvig Söderling	Essays on African Growth, Productivity, and Trade, 2005
132. Åsa Eriksson	Testing and Applying Cointegration Analysis in Macroeconomics, 2005
133. Fredrik Hansen	Explorations in Behavioral Economics: Realism, Ontology and Experiments, 2006
134. Fadi Zaher	Evaluating Asset-Pricing Models in International Financial Markets, 2006
135. Christoffer Bengtsson	Applications of Bayesian Econometrics to Financial Economics, 2006
136. Alfredo Schclarek Curutchet	Essays on Fiscal Policy, Public Debt and Financial Development, 2006
137. Fredrik Wilhelmsson	Trade, Competition and Productivity, 2006
138. Ola Jönsson	Option Pricing and Bayesian Learning, 2007
139. Ola Larsson	Essays on Risk in International Financial Markets, 2007
140. Anna Meyer	Studies on the Swedish Parental Insurance, 2007
141. Martin Nordin	Studies in Human Capital, Ability and Migration, 2007
142. Bolor Naranhuu	Studies on Poverty in Mongolia, 2007
143. Margareta Ekbladh	Essays on Sickness Insurance, Absence Certification and Social Norms, 2007
144. Erik Wengström	Communication in Games and Decision Making under Risk, 2007
145. Robin Rander	Essays on Auctions, 2008
146. Ola Andersson	Bargaining and Communication in Games, 2008
147. Marcus Larson	Essays on Realized Volatility and Jumps, 2008
148. Per Hjertstrand	Testing for Rationality, Separability and Efficiency, 2008
149. Fredrik NG Andersson	Wavelet Analysis of Economic Time Series, 2008
150. Sonnie Karlsson	Empirical studies of financial asset returns, 2009
151. Maria Persson	From Trade Preferences to Trade Facilitation, 2009
152. Eric Rehn	Social Insurance, Organization and Hospital Care, 2009
153. Peter Karpestam	Economics of Migration, 2009
154. Marcus Nossman	Essays on Stochastic Volatility, 2009
155. Erik Jonasson	Labor Markets in Transformation: Case Studies of Latin America, 2009

156. Karl Larsson	Analytical Approximation of Contingent Claims, 2009
157. Therese Nilsson	Inequality, Globalization and Health, 2009
158. Rikard Green	Essays on Financial Risks and Derivatives with Applications to Electricity Markets and Credit Markets, 2009
159. Christian Jörgensen	Deepening Integration in the Food Industry – Prices, Productivity and Export, 2010
160. Wolfgang Hess	The Analysis of Duration and Panel Data in Economics, 2010
161. Pernilla Johansson	From debt crisis to debt relief: A study of debt determinants, aid composition and debt relief effectiveness, 2010
162. Nils Janlöv	Measuring Efficiency in the Swedish Health Care Sector, 2010
163. Ai Jun Hou	Essays on Financial Markets Volatility, 2011
164. Alexander Reffgen	Essays on Strategy-proof Social Choice, 2011
165. Johan Blomquist	Testing homogeneity and unit root restrictions in panels, 2012
166. Karin Bergman	The Organization of R&D - Sourcing Strategy, Financing and Relation to Trade, 2012
167. Lu Liu	Essays on Financial Market Interdependence, 2012
168. Bujar Huskaj	Essays on VIX Futures and Options, 2012
169. Åsa Ljungvall	Economic perspectives on the obesity epidemic, 2012
170. Emma Svensson	Experimenting with Focal Points and Monetary Policy, 2012
171. Jens Dietrichson	Designing Public Organizations and Institutions: Essays on Coordination and Incentives, 2013
172. Thomas Eriksson	Empirical Essays on Health and Human Capital, 2013
173. Lina Maria Ellegård	Political Conflicts over Public Policy in Local Governments, 2013
174. Andreas Hatzigeorgiou	Information, Networks and Trust in the Global Economy - Essays on International Trade and Migration, 2013
175. Gustav Kjellsson	Inequality, Health, and Smoking, 2014
176. Richard Desjardins	Rewards to skill supply, skill demand and skill match-mismatch: Studies using the Adult Literacy and Lifeskills survey, 2014
177. Viroj Jienwatcharamongkhol	What Drives Exports? Empirical Evidence at the Firm Level, 2014
178. Anton Nilsson	Health, Skills and Labor Market Success, 2014
179. Albin Erlanson	Essays on Mechanism Design, 2014
180. Daniel Ekeblom	Essays in Empirical Expectations, 2014

181. Sofie Gustafsson	Essays on Human Capital Investments: Pharmaceuticals and Education, 2014
182. Katarzyna Burzynska	Essays on Finance, Networks and Institutions, 2015
183. Mingfa Ding	Corporate Ownership and Liquidity in China’s Stock Markets, 2015
184. Anna Andersson	Vertical Trade, 2015
185. Cecilia Hammarlund	Fish and Trips in the Baltic Sea - Prices, Management and Labor Supply, 2015
186. Hilda Ralsmark	Family, Friend, or Foe? Essays in Empirical Microeconomics, 2015
187. Jens Gudmundsson	Making Pairs, 2015
188. Emanuel Alfranseder	Essays on Financial Risks and the Subprime Crisis, 2015
189. Ida Lovén	Education, Health, and Earnings – Type 1 Diabetes in Children and Young Adults, 2015
190. Caren Yinxia Nielsen	Essays on Credit Risk, 2015
191. Usman Khalid	Essays on Institutions and Institutional change, 2016
192. Ross Wilson	Essays in Empirical Institutional Analysis, 2016
193. Milda Norkute	A Factor Analytical Approach to Dynamic Panel Data Models, 2016
194. Valeriia Dzhamalova	Essays on Firms’ Financing and Investment Decisions, 2016
195. Claes Ek	Behavioral Spillovers across Prosocial Alternatives, 2016
196. Graeme Cokayne	Networks, Information and Economic Volatility, 2016
197. Björn Thor Arnarson	Exports and Externalities, 2016
198. Veronika Lunina	Multivariate Modelling of Energy Markets, 2017
199. Patrik Karlsson	Essays in Quantitative Finance, 2017
200. Hassan Sabzevari	Essays on systemic risk in European banking, 2017
201. Margaret Samahita	Self-Image and Economic Behavior, 2017
202. Aron Berg	Essays on informational asymmetries in mergers and acquisitions, 2017
203. Simon Reese	Estimation and Testing in Panel Data with Cross-Section Dependence, 2017
204. Karl McShane	Essays on Social Norms and Economic Change, 2017
205. Elvira Andersson	From Cradle to Grave: Empirical Essays on Health and Economic Outcomes, 2017
206. Yana Pryymachenko	Heavy Metal Exposure in Early Life - Health and Labour Market Perspectives, 2017
207. Alemu Tulu Chala	Essays on Banking and Corporate Finance, 2017
208. Jim Ingebretsen Carlson	Essays on economic behavior, focusing and auctions, 2018
209. Jörgen Kratz	Essays on Matching, 2018

210. Anna Welander Tärneberg	Essays on Health in Developing countries, 2018
211. Osmis Areda Habte	Essays on competition and consumer choice, 2018
212. Thomas Hofmarcher	Essays in Empirical Labor Economics, 2019
213. Hjördis Hardardottir	Time and inequality – A study of individual preferences, 2019
214. Erik Grenestam	Essays in Applied Microeconomics, 2019
215. Sara Moricz	Institutions, Inequality and Societal Transformations, 2019
216. John Källström	Mobility in Science, 2019
217. Mehmet Caglar Kaya	Essays on Corporate Growth and Corporate Credit Risk, 2020
218. Dinh-Vinh Vo	Essays on risk spillover and information transmission in the financial markets, 2020
219. Kristoffer Persson	Essays on Expectations - Information, Formation and Outcomes, 2020
220. Polina Knutsson	Empirical Studies on Firm and Labor Market Dynamics, 2020
221. Sanna Ericsson	Reaching For Equality: Essays in Education and Gender Economics, 2020

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

SCHOOL OF ECONOMICS AND MANAGEMENT

Lund University
Department of Economics
ISBN 978-91-7895-498-8
ISSN 0460-0029