Essays on the Economics of Development, Education and Violence

María Paula Gerardino Gutiérrez

TESI DOCTORAL UPF / ANY 2014

DIRECTOR DE LA TESI

Libertad González Luna, Departament of Economics and Business

to Da

Acknowledgements

First, I want to thank my advisor Libertad González for her support, guidance and encouragement during my PhD. Her patience, counsel and knowledge have been truly invaluable. I also owe a lot to Alessandro Tarozzi, Giacomo De Giorgi and Stephan Litschig. They have been great mentors during different stages of my PhD. Without their support, things would have been a lot more difficult. I admire all of them not only as economists but also as human beings. I would like to thank Adriana Camacho for being a great co-author and friend. I am indebted to Paula Bustos, Gabrielle Fack, Marta Reynal-Querol, Antonio Ciccone, Gianmarco León, Maia Güell, Larbi Alaoui and Christian Fons-Rosen for their help and suggestions during the realization of this thesis. Special thanks go to Marta Araque and Laura Agusti, who have always helped me with everything and have made things easier during these years. Thanks for having a place for me in your office.

Thanks enormously to Mona, Ignasi, Lina, Petri and Ponti for being my family in Barcelona. Pompeu gave me not just a PhD degree but also three brothers. Thanks to Oriol, Kiz and Cris for all your help and support and for the good times we had in the office 20.143. Michael you are always in my thoughts. Thanks to Tom and Felipe for always being there. Thanks loads to Paula for being a wonderful friend. During this time, I have been fortunate to meet several people who have also become friends: Gene, Felipe, Miguel, Jose, Ciccio, Fer, Pablo, Flo, Bruno, Ambro, Giulia, Laura, Tanya, Elisa, Marc, Emma, Andrea T, Thomaz, Feray and Diego. Thank you all for the good and bad times we had together inside and outside UPF. I also want to thank all my fellow PhD students at UPF.

Last but not least, I want to thank my family. Thanks Mom, Dad, Libar, Fati, Marti, Mati and Nani for their unconditional love. For them and because of them I am able to present this work. Mom and Dad, thanks for being such as great example. Finally, the greatest thanks go to my husband, Da. Tesoro, no words can describe how thankful I am to have you in my life.

Abstract

This thesis investigates various aspects of economic development. In the first chapter, I study the impact of violence on human capital investments, focusing mainly on the relationship between male-biased violence and the educational gender gap in Colombia. I find that boys are less likely to be enrolled at secondary school age relative to girls when male-biased violence is high. In the second chapter, I explore the impact of violence on fertility choices. In particular, I study how gender-biased violence in Colombia affects the number and the gender of actual and desired children demanded by women. I find that women living in areas that experience faster increases in male-biased violence have and want fewer children. Furthermore, they show a lower preference for boys. Finally, in the third chapter I explore the impact of public procurement audits on compliance with procurement regulation, cost and performance in Chile. The results show that the audits affect the public entities' subsequent procurement practices by a temporary shift toward less transparent modalities of procurement.

Resum

Aquesta tesis investiga varis temes de desenvolupament econòmic. En el primer capítol, estudio l'impacte que té la violència sobre les inversions en capital humà, fent èmfasis en la relació entre la violència esbiaixada cap al gènere masculí i la diferència en els nivells d'educació entre nois i noies a Colòmbia. Trobo que els nois tenen una probabilitat menor que les noies d'estar inscrits en l'educació secundària quan la violència que afecta més al gènere masculí és més alta. En el segon capítol, exploro l'impacte que té la violència en les decisions de fertilitat. En concret, estudio com, a Colòmbia, la violència esbiaixada cap al gènere masculí afecta al nombre i al gènere de descendència real i desitjada per les dones. Trobo que les dones que viuen en àrees que experimenten un increment ràpid de la violència que afecta més al gènere masculí que al femení tenen i volen menys descendència. A més a més, aquestes dones mostren una preferència alta per tenir filles en lloc de fills. Finalment, en el tercer capítol exploro l'impacte que tenen les auditories fetes en els processos de contractació pública a Xile, en el compliment de les regulacions existents, en els seus costos i en el seu rendiment. Els resultats mostren que les auditories afecten les subseqüents pràctiques que fan les ens públiques tot realitzant temporalment modalitats de contractació menys transparents.

Foreword

My doctoral thesis is a collection of three essays that study various aspects of economic development, with a special emphasis on Latin America. The first two chapters analyze some of the determinants of human capital investments and other family decisions in developing countries. In the first chapter, I study the impact of violence on the educational gender gap. In the last few decades, Latin America has experienced a substantial increase in violence related to gang/organized crime, fueled by the expansion of narcotraffic. This paper analyzes the impact of this type of violence on human capital investment decisions. I focus on the relationship between the male versus female homicide rate differential and the gender gap in education. Using data from Colombia and exploiting the temporal and spatial variation in violence between 1985 and 2005, I find that boys are less likely to be enrolled at secondary school age relative to girls when male-biased violence is high. An increase of one standard deviation in violence leads to a 1.1 percentage point enlargement of the gender gap in enrolment, in disfavor of boys. This effect is important since the gender gap in enrolment in secondary school in Colombia is estimated to be 8 percentage points, in favor of girls. I find a similar effect on years of school completed. Estimates are larger in families with lower levels of education and robust to the inclusion of municipality-year fixed effects and household fixed effects. In addition, results are not driven by migration or coca production. The evidence in this paper suggests that violence has an impact on investments in education through two main chan-nels: changes in the opportunity cost of schooling, and changes in life expectancy and perceived safety.

The second chapter, coauthored by Adriana Camacho, focuses on the relationship between male-biased violence and fertility decisions. We exploit the variation in homicide rates across time and over municipalities in Colombia in order to explore the effect of gender-biased violence on fertility choices. We investigate how violence affects the number and the gender of actual and desired children demanded by households. Since males in Colombia are more likely to participate in war and in criminal activities, we use the number of homicides per municipality, as a measure of male-biased violence. The demand for children is affected by their expected survival rates. We consider violence to be a factor that affects the expected survival rates of children conditional on gender and that therefore might affect the fertility decisions of women. Results suggest that women that experience higher levels of violence during their fertile years have and want fewer children and show a lower preference for boys.

The third chapter, coauthored with Stephan Litschig and Dina Pomeranz, studies the impact of public procurement audits on compliance with procurement regulation in Chile. The government is the biggest buyer in the economy of most countries. At the same time, the public procurement process if often thought to be fraught with corruption and malpractice. However, there is little evidence regarding the impact of audits aimed at reducing such malpractice. This paper investigates the effect of being audited on public entities' subsequent procurement practices in Chile. For identification, we exploit a scoring rule of the national auditing agency, which allows for regression discontinuity analysis. Our results show that the audits lead to a temporary shift toward less transparent modalities of procurement. The share of the amount of total purchases through direct negotiations increases by around 20 percentage points, at the expense of the use of public auctions. The effect is most pronounced during the months when the audit is taking place and disappears completely by the subsequent fiscal year. Since audits in Chile rarely happen in consecutive years, and since the audit typically only covers the most recently completed fiscal year, this time pattern of effects is consistent with public agents responding to a temporary drop in audit risk during the year of the audit.

Table of Contents

List of Figures

List of Tables

Chapter 1

The Effect of Violence on the Educational Gender Gap

1.1 Introduction

Human capital investments have been shown to promote economic growth and development (Galor and Moav, 2004). On the other hand, there is a consensus among economists and policymakers that violence and crime negatively affect the quality of life of individuals (Heinemann and Verner, 2006; United Nations, 2011). As a result, studying how violence affects investments in education has compelling policy relevance. In this paper, I investigate the impact of violence on education, focusing on the relationship between the gender bias of violence and the education gap between boys and girls.

Violence and crime are a major concern in both developed and developing countries. In particular, Latin America has the highest homicide rate in the world for ages 15–24 (United Nations, 2011), four times higher than the world average. The homicide rate for Latin American youth ages 10–29 is 36.4 per 100,000 inhabitants, compared to 17.6 in Africa, 11.0 in the United States, and 0.9 in the high-income countries of Europe. Moreover, the gender structure of violence is very unbalanced. Male homicides account for 90 percent of all homicide victims, compared to 81 percent in Africa and 73 percent in Europe. This large share of male homicide victims has been linked to narcotraffic, organized crime and the proliferation of violent youth gangs (United Nations, 2011).

This paper studies the effect of violence on the educational gender gap. In contrast to many other developing countries where investments in the education of girls continue to lag behind those of boys, most Latin American countries have a "reverse" gender gap in education. Women are achieving nowadays higher average years of schooling than men (Duryea et al., 2007). To the best of my knowledge, this is the first paper to look at the extent to which male-biased violence has contributed to this gap. I consider a period of intense violence in Colombia that caused a large increase in the male versus female homicide rate differential, and analyze its consequences on schooling.[1] Using data on violence at the municipality level between 1983 and 2005, I find that municipalities that faced faster increases in violence experienced an enlargement of the gender gap in school enrolment, in disfavor of boys. This is particularly the case for children of secondary school age.[2]

Research has found that the direct and indirect costs associated with investments in schooling might be exacerbated in a violent context. For example, in areas with high levels of violence, disruption of access routes and, more importantly, destruction of schools will increase the time and cost of going to school (Akbulut-Yuksel, 2009). Violence and crime may also affect the opportunity cost of going to school and the returns to education. Following Becker (1968), crime may be considered as an occupational choice or an investment opportunity. Individuals compare the payoffs and costs of crime versus the returns to investments in education and work (see Lochner, 2004).[3] In addition, recent research has shown that unexpected shocks to life expectancy have a negative impact on an individual's educational attainment by reducing expected returns to school (Jayachandran and Lleras-Muney, 2009; Oster et al., 2013). Looking at changes in life expectancy and in job opportunities, Evan et al. (2012) argue that some of the enlargement of the white/black gap in education in the United States is due to the decrease in the returns to schooling of black males after the arrival of crack markets and the subsequent large increase in murder rates.

In this paper, I focus on the heterogenous effects of violence on education between boys and girls. Although there are different mechanisms through which violence can affect investments in education, such as school availability and school quality, many of these factors are likely to affect boys and girls similarly. I argue that male-biased

[1]During the 1990s, Colombia experienced a severe increase in homicides, with rates rising 159 percent between 1985 and 1994, from 36.7 to 95 per 100,000 inhabitants on average (Krug et al., 2002).

[2]In the empirical analysis, I mainly focus on children aged 11–17 (secondary school age). In Colombia, primary school enrolment rates are very high and most of the school investment variation and changes in the opportunity cost of schooling happens at the secondary school age.

[3]Most of the literature on crime and education studies the effects of education on adult crime (Lochner and Moretti, 2004). See Hjalmarsson and Lochner (2012) for a detailed literature review on this topic.

violence might affect enrolment decisions of boys relative to girls through two main channels. First, it increases the opportunity cost of attending school, since there is a higher demand for (male) workers in the violence sector. Second, it reduces the returns of attending school as life expectancy and perceived safety fall. I combine these two channels in a simple conceptual framework (based on Becker, 1962 and Ben-Porath, 1967) in which agents compare their expected returns to schooling with the direct and indirect costs associated with its acquisition. A family makes the decision of whether to enrol its child in school taking into consideration the returns to violence and schooling, and discounting the expected returns by the child's life expectancy/perceived safety. As a result, families may decide to educate only some of their children (Behrman et al, 1982, 1990; Dahan and Gaviria, 2003). In this framework, when more men are targeted by violence (relative to women), fewer boys are predicted to go to school (relative to girls).

To empirically identify the effect of gender-biased violence on the gender gap in education, I use multiple years of census data with information on school enrolment and family characteristics. I match this dataset with data on exposure to male-biased violence, proxied by the difference between male and female homicide rates in the municipality of residence. I exploit the time-municipality variation of male-biased violence. To address the potential endogeneity of violence arising from municipality and time characteristics, I include a full set of municipality fixed effects (that are allowed to vary by gender), time fixed effects, along with region-specific time trends. Furthermore, I control for time-varying municipality characteristics that may also affect enrolment decisions such as the population density, the income tax per capita, and the unemployment rate. To account for additional time-varying omitted variables across municipalities and unobserved heterogeneity across families in dimensions such as job opportunities, income and preferences for education, I evaluate the robustness of the results to the inclusion of municipality-year fixed effects and household fixed effects. To this end, I compare changes in the schooling gender gap within municipalities and also within families that are affected by different levels of violence over time.

My empirical results are in line with the conceptual framework. I find that when the risk of homicide is higher for males, families are more likely to favor girls' education. The magnitude of the estimate is substantial: at the upper-secondary school age, an increase of one standard deviation in the level of male-biased violence accounts for 13.8 percent (1.1 percentage points) of the gender gap in enrolment in secondary school.[4] This impact is larger in families with lower levels of education. I find that the negative effect of violence on the enrolment gender gap is roughly 1.5 times larger in families in which either the head of household or the mother has less than a primary school education. The differential effect of violence on education is significant not only for enrolment decisions but also for years of school completed. This implies that male-biased violence also has a long-term effect. Furthermore, the impact of violence on the gender gap is driven by the changes in the difference between male and female homicides and not by changes in other causes of mortality. When I use a measure of male-biased mortality that includes all causes of death other than homicides, the estimates suggest that differential mortality between genders by itself does not have an impact on schooling decisions conditional on gender.

Since violence may induce migration, positive or negative selection into migration could bias the results. To account for this possibility, I first compare migrant and non-migrant children. Second, I consider the potential migration of boys, which in turn may affect the proportion of boys versus girls in a municipality. I look at the effect of violence on the ratio of boys relative to girls in school age. In both cases, I find that selective migration is not biasing my estimates.

My dataset includes the two periods within the last four decades when the rate of homicides in Colombia increased the most. The first is associated with the consolidation of the drug cartels due to the growth of the cocaine trade in the late 1980s and early 1990s. The second coincides with the consolidation of the guerrilla groups in the late 1990s. I also consider the potential effect of the increase in coca cultivation in Colombia during the late 1990s. Coca harvesting played an important role not only in fueling violent conflict, but also in bringing new job opportunities to the municipalities where coca is produced (Angrist and Kugler, 2008). I find that although coca production is associated with higher violence in Colombia, it is not a determinant of the broadening of the educational gender gap in violent regions.

This paper primarily relates to three strands of the literature: research looking at the determinants of human capital investments (since Becker, 1962; Berhrman et al., 1982, 1990), the effects of crime on education (see Hjalmarsson and Lochner, 2012; Evan et al., 2012), and the impact of civil war on schooling decisions. Although civil war has been identified as a deterrent for growth and development, more research is needed before conclusions can be drawn about the specific effects of war on individual and family welfare (Blattman and Miguel, 2010). Empirical

[4] The gender gap in enrolment in secondary school in Colombia is 8 percentage points, in disfavor of boys (Hausmann, Tyson, and Zahidi, 2012)

evidence at the micro-level suggests that violence has a negative impact on educational attainment (Justino, 2011). For example, in Zimbabwe, children affected by civil war completed fewer years of school than children that were not affected (Alderman, Hoddinott and Kinsey, 2006). Similar results are found by Swee (2009) for Bosnia and Herzegovina and by León (2012) for Perú. Rodríguez and Sánchez (2012) find that armed attacks in Colombia made children more likely to drop out of school. These papers, however, do not consider the differential role played by boys and girls as victims and actors in violent conflicts.

Recent empirical research on the heterogeneous effects of conflict suggests that the effects of war-related destruction and civil war depend on gender and on the particular characteristics of each conflict. Shemyakina (2011), and Chamarbagwala and Moran (2011) examine the impact of the civil war on education in Tajikistan and in Guatemala, respectively. These papers use as a measure of civil conflict household damage dwelling reports in Tajikistan and the number of human rights violations in Guatemala. In contrast to the results I find in my paper, the authors find that girls are more adversely affected by conflict. They argue that parents may keep their girls at home in order to protect them from sexual assaults or harassment on their way to school. Furthermore, if the conflict causes a negative income shock and the returns to education for boys are larger than for girls, families that are income constrained may invest just in their boys' education. In contrast, Valente (2013) finds that in Nepal, an increase in the intensity of Maoist activity is associated with an increase in the school attainment of girls. This impact is mainly attributed to the strong Maoist ideological agenda of reducing inequality. Looking at children as participants in conflict, two studies explore the effect of random abductions of children in Uganda on educational outcomes and find that boys are the ones affected (Blattman and Annan (2010) and Annan et al. (2009)). This finding coincides with Akresh and De Walque's (2008) study on Rwanda, in which they show that boys from non-poor families were negatively affected by the conflict. They attribute this finding to wealthy families being targeted during the genocide and to the low enrolment of girls and poor children prior the conflict.

The empirical findings of my paper contribute to this literature by studying the effects of violence in an ongoing conflict and quantifying the exposure to a certain type of violence using a gender-specific measure. One of the drawbacks of the data used in some of the previous studies is that the information on violence comes from individuals' self-reported measures. By using male and female homicide records collected by hospitals and forensic offices across the Colombian territory in different years, I am able to consider changes in the intensity of violence without worrying about the bias if particular unobserved individual characteristics lead to different levels of reporting. Furthermore, I use the measure of male-biased violence as a proxy for gender-specific changes in the opportunity costs and in the returns of going to school. This allows me to pin down the potential mechanisms that explain the heterogeneous effects of violence on education.

The rest of the paper is organized as follows. Section 1.2 provides an overview of the educational system and chronicles the history of crime and violence in Colombia. Section 1.3 describes the data. In Section 1.4, I present the conceptual framework. Section 1.5 specifies the empirical methodology and presents the main results. Section 1.6 and Section 1.7 report robustness checks and additional specifications, respectively. Finally, Section 1.8 concludes.

1.2 Historical and Institutional Context

1.2.1 Violence in Colombia

Over the past two decades, Colombia's evolution of violent crimes has been unique. In the early 1970s national homicide rates were not very different from its neighboring countries. Beginning in the late 1970s, however, homicide rates increased dramatically, growing more than three-fold by 1991 (Gaviria, 2000). With a homicide rate three times higher than those of other violent countries such as Mexico and Brazil, and 36 times greater than Canada, the magnitude of violent crimes in Colombia is overwhelming (see Figure 1.1). These numbers are astonishing when considering that Colombia has had a s democratic government for nearly 50 years and has sustained improvements in social and economic indicators during the past several decades (Moser and Mcllwaine, 2000). The increase and expansion of violence and crime in Colombia in the last two decades has been attributed mainly to the emergence of drug cartels and the economy of illicit drugs, the increase in coca production, and the consolidation of guerrilla groups.

Drug trafficking has played an important role in the intensification and expansion of violence in Colombia. The country became the largest cocaine exporter in the world during the 1980s, fueling the growth and strengthening of the main drug cartels (United Nations, 2009). In 1989, the murder of presidential candidate Luis Carlos Galan triggered a war against the drug cartels (Levitt and Rubio, 2000). Furthermore, drug and terrorism policy in the United States and Europe also affected crime and violence during this period. For example, the enforcement of the extradition laws in regard to drug lords and the increase in antinarcotics aid destabilized the Colombian drug market, causing more violence between organized drug cartels trying to gain control over local areas. This contributed to the increase in the number of homicides during the late 1980s and the early 1990s. Homicides peaked in 1991, when nearly one in every 1,000 Colombians was murdered (see Figure 1.2). This surge in violence disproportionately affected young males due to the proliferation of youth gangs related to drug trafficking. During this period, Colombian men between the ages of 15 and 35 were 15 times more likely to be homicide victims than women in the same cohort (Heinemann and Verner, 2006).

Before 1993, Peru and Bolivia were the main countries responsible for harvesting coca leaf, and the Colombia drug cartels produced and trafficked the cocaine around the world. In the early 1990s, the drug industry had to adapt to changes in United States antinarcotics policies, which made transportation between countries that harvest coca and those that produce and traffic cocaine costly and dangerous (Serafino, 2002). Coca cultivation shifted to Colombia, going from 20,000 cultivated hectares in 1990 to 160,000 in 2000 (United Nations, 2001). These changes also prompted the development of linkages between the illegal drug industry and the most important guerrilla groups (FARC and ELN) and illegal self-defense groups (AUC). These groups derive substantial income by taxing coca production and recruiting people in the lower ranks of the drug business (Rabasa and Chalk, 2001).

Colombia's first guerrilla group emerged in 1949, when the communist party began to organize self-defense groups for the peasants to fight the Conservative Party in power. The 1980s marked a drastic turn in the growth and consolidation of FARC, when it expanded and brought the war to the urban centers and used kidnapping to finance its activity. During the 1990s and with the surge of coca production in Colombia, FARC proliferated in the coca-growing regions. This expansion allowed it to grow from 350 fighters at its founding in 1966 to approximately 3,600 in 32 fronts in 1986, 7,000 in 60 fronts in 1995, and 15,000–20,000 in more than 70 regiments in 2000 (Rangel, 1998). The second most important guerrilla group in Colombia is the National Liberation Army (ELN), founded in 1962. This group also grew dramatically during the 90s going from around 16 people in 1962 to 4,500 people in 2000 (Formisano et al., 2005). While allegedly the main objective of these groups is taking over political power, their actions have increasingly relied on terrorism (Vargas, 2009). Other important actors in the Colombian conflict are the paramilitary or self-defense groups. These groups emerged as a counter-insurgency organization known as the United Self-Defense of Colombia (AUC). They were small groups sponsored and financed by land owners to fight guerrillas and protect them. In the 1990s, these groups also grew exponentially, reaching 10,000 people. Just like guerrilla groups, they also received financing from drug cartels. Between 1988 and 2005, Colombian guerrillas killed 1,200 civilians in about 200 massacres, and more than 6,100 died in just under 1,000 massacres perpetrated by right-wing militias (Vargas, 2009).

This paper looks at the period between 1983 and 2005, which includes the 1991 and 2002 peaks in violence. The first period (1983–1993) corresponds to the time when coca exports were rapidly expanding and drug cartels were fighting for control of the drug markets (Pecaut and González, 1997). The second period (1993–2005) covers the increase in the cultivation of coca across the Colombian territory (see Appendix 1) and the strengthening of the guerrilla and paramilitary groups (Rabasa and Chalk, 2001).

1.2.2 Education in Colombia

The education system is one of the largest components of the public sector in Colombia, both in terms of size and share of public expenditures. In addition, public school teachers and educational staff comprise the largest share of public sector employees within the economy. From the early 1960s, the Colombian government began to improve public access to education. Government funding for education increased five-fold between 1966 and 1986, causing primary school enrolment to more than double by 1987 (Hanratty and Meditz, 1988). Literacy rates similarly rose from 54 percent in 1973 to 88 percent by 1987. Since the 1970s, the gender gap in schooling has closed, but it not was until the late 1980s that this gap reversed. As a result, more girls are now enrolled in high school than boys. It is notable that the education gap between girls and boys got wider around the same time that Colombia started experiencing high levels of violence (Appendix A1- Figure A1.1). The reversal of the gender gap in education is not specific to Colombia; research has shown than in most Latin American and Caribbean countries, women achieve higher levels of education than their male counterparts (Duryea et al., 2007).

The formal education system in Colombia is composed of three levels: preschool, primary, and secondary (see Table 1.1). Secondary school is divided into two levels: basic secondary and upper-secondary. Since 1991, education has been compulsory between the ages of 5 and 14. Enrolment is free for children in primary school, but only families in the lowest income groups do not pay for secondary school.[5]

1.3 Description of the Data

My data are derived from two main sources. First, children's school outcomes and household and family characteristics come from a 10-percent sample of the Colombian Population Census (CPC) for 1985, 1993, and 2005. Second, the data on homicides by municipality are from the Departamento Administrativo Nacional de Estadística (DANE) for the periods 1983–1984, 1991–1992, and 2003–2004.

1.3.1 Individual Data

The CPC microdata sample, consisting of individual records, is acquired from the Integrated Public Use Microdata Series (IPUMS)-International.[6] The CPC records are compiled through direct dwelling questionnaires in which the head of the household provides information about the dwelling and household, and any children below the age of 10. Residents 10 years and older individually provide personal information. This study's main sample only includes information on basic school-age children. The sample contains all CPC respondents who were between 6 and 17 years of age in the year of the census. Colombia consists of 33 regions and 1,123 municipalities that the CPC has grouped into 533 units.[7] I use only 515 of these units due to data availability and comparability of the municipalities across years. I exclude 17 units because several of the municipalities that form these groups were created after 1985. The sample thus comprises information on 2,428,282 school-age children and their families. Table 1.2 presents descriptive statistics. Around 51 percent of the sample is male, and 64 percent of the children live in urban areas. Children that had migrated in the previous five years account for 10 percent of the sample. In terms of family characteristics, around 78 percent of children live in a household headed by a man, and in 73 percent, the head of the household is employed. Furthermore, in 88 percent of the cases, the mother is present in the house and in 72 percent the father. In 83 percent of cases, the head of the household is the child's parent, but 11 percent are grandparents and 1 percent are another type of relative. Moreover, 68 percent of the children live in a family-owned house, whereas 21 percent and 11 percent live

[5] http://www.cna.gov.co/1741/article-187279.html

[6] The data set is formed by taking a systematic sample of every 10th private dwelling, excluding group quarters and the indigenous population.

[7] Municipalities that had fewer than 20,000 people in 2005 are grouped together.

in a rented dwelling or free of charge, respectively. Finally, around 84 percent of the dwellings have electricity, 71 percent have a flush toilet, and 20 percent do not have a toilet in the dwelling.

Table 1.3 provides additional descriptive statistics on school enrolment of children by level of school, gender, and census year. [8]Although enrolment rates for girls are higher than for boys in each year, in general the rates are increasing for all levels of school. I calculate the difference between the average enrolment rates of girls minus boys. In the case of primary education, the difference in means between genders is statistically significant and is around 2.2 percent in 1985, 2.1 percent in 1993, and 1.2 percent in 2005. In contrast, the gender gap in education grows in secondary school from 4.0 percent in 1985 to 5.2 percent in 1993 and 4.3 percent 2005. The mean upper-secondary school enrolment for girls varies from 62 percent in 1985 to 68 percent in 1993 to 76 percent in 2005. For boys, the enrolment rate changes from 57 percent to 59 percent and then to almost 70 percent, respectively. The difference in the average enrolment rate of girls versus boys at the upper-secondary school level is larger than for the other levels, ranging from 4.5 percent in 1985, to its peak in 1993 (8.4 percent), and decreasing to 6.2 percent in 2005.

1.3.2 Homicide Data

The Departamento Administrativo Nacional de Estadística (DANE) collects the number of homicides in each municipality using Vital Statistics Records, which include listings of individual deaths and basic demographic information on the deceased (gender, educational level, and marital status). The information on deaths comes from the forensics offices, hospitals, and police reports in each municipality and is available annually from 1979 to 2011. For this study, I aggregate by municipality and year all the deaths caused by homicide and guerrilla or governmental actions. Table 1.4 summarizes the descriptive statistics of homicides by year. The number of homicides increases over time: 10,256 in 1984, 30,573 in 1992, and 25,120 in 2004. Most of the victims were male (around 92 percent in 1984 and 1992 dropping to 90.7 percent in 2004) between the ages of 18 to 35 (more than 60 percent). Not every death record is complete, but among the subsample with full information, 32–59 percent of victims were single, depending on the year considered. Data on the education level of the victims are only available for the year 2004. Around 25

[8]I organize school levels into three different groups: Primary school (age 6–10), secondary school (age 11–17) and upper-secondary school (age 15–17).

percent of the victims had at most a preschool education; 43 percent had at least a primary school education (either incomplete or complete, 16 and 27 percent, respectively); and 23 percent had attended secondary school. Around 8 percent of the victims went to university.

To generate the main variable of interest—the difference between the male homicide rate minus the female homicide rate by municipality—I match the number of male homicides with the intercensal estimates and projections of population for each municipality each year.[9] Small municipalities are grouped together in the CPC sample, so I compute the weighted average of the number of male and female homicides by population size for each group and assign these values to all individuals in the municipality unit.

1.3.3 Coca Production and Household Member Characteristic

The increase of coca production in Colombia played an important role not just in fueling violent conflict but also in creating new job opportunities in the municipalities that cultivate this crop. I use data from the United Nations Office on Drugs and Crime (UNDOC) to account for coca harvesting. Coca cultivation figures are collected through the Illicit Crop Monitoring System (SIMCI-Sistema Integrado de Monitoreo de Cultivos Ilicitos) and are available at the municipality level from 1999 through 2008. I also use the SIMCI census data to examine the economic and demographic situation of farmers in areas where coca is cultivated. In this census, farmers (coca and non-coca growers) were asked about the social structure and the characteristics of their households.

1.4 Conceptual Framework

I next present a simple conceptual framework, based on Becker (1962) and Ben-Porath (1967), that attempts to shed light on the effects of violence on enrolment decisions in secondary education for boys versus girls. In this framework a family makes the decision whether to send the child to school taking into account the relative returns to education. These returns are affected by changes in the local level of violence, since these affect the returns that children can get from engaging in violence-related activities and may also impact life expectancy as well as perceived

[9]Censuses in Colombia were conducted in 1973, 1985, 1993, and 2005. These also comes from DANE.

safety of children.

In my model, there are two periods. In the first period, taking into consideration the returns to violence, the returns to school and discounting expected returns by the child's life expectancy/perceived safety, the family makes the decision whether to enrol the child in school. The returns to education are acquired in the second period whereas the returns to violence are obtained in the first and second period. In this framework, I focus on the effects of violence that are potentially different for girls and boys: returns to education, returns to violence, and life expectancy (and the perception of safety). I make three key assumptions: 1) Payoffs of violence are different for boys and girls; 2) life expectancy and perceived safety is different across genders; and 3) violence levels are exogenous. Gender differences are key for the results of my model. I therefore abstract from all the effects of violence on education which are not gender-specific. For example, in areas with high levels of violence, disruption of access routes and, more importantly, destruction of schools will increase the time and cost of going to school (Akbulut-Yuksel, 2009). In principle, these type of effects should not have different impacts between boys and girls.

Because the type of violence I consider primarily affects males, the enrolment decisions of boys and girls are affected differently. Male-biased violence creates more opportunities for boys to earn money in violence-related activities. As a result, the returns to violence are larger for boys than for girls. Also, when male-biased violence increases, the life expectancy and the perception of safety is likely to be lower for boys relative to girls.

These changes in payoffs affect children's enrolment decisions as families compare their children's expected returns to education and to violence. When violence is low, the payoffs to violence are low and life expectancy (and the perception of safety) is high. In this case, the family will decide to send the child to school. In contrast, if violence increases above a certain threshold, the family will not send the child to school. Because the returns to violence are higher for boys, the threshold of violence needed for leaving school is lower for boys. Therefore, an increase in male-biased violence is likely to increase the gender gap in education, as boys are more likely to drop out of school. Appendix B1 provides a formal statement of the problem, comparative statics, and implications.

1.5 Empirical Strategy and Main Results

Before discussing the empirical strategy and results of the paper, I describe the basic notation that I use throughout the remainder of the paper. First, the variable $HR_{m,t-1} = ((\frac{Number\,of\,Male\,Homicides_{m,t-1}}{MalePopulation_{m,t-1}}) * 100)$-$((\frac{Number\,of\,Female\,Homicides_{m,t-1}}{FemalePopulation_{m,t-1}}) * 100)$ denotes the difference between the average number of male homicides and female homicides in years $t-1$ and $t-2$, divided by the male and female population, respectively, in each municipality. I take the number of male and female homicides in the year prior to each census because enrolment decisions are made at the beginning of the year. I average the number of homicides in each municipality over the previous two years to account for the persistence of violence.[10] $Male_{i,m,t}$ is equal to 1 if the child is a boy and 0 otherwise. Finally, the main variable of interest —$HR_{m,t-1} * Male_{i,m,t}$— is the interaction between $Male_{i,m,t}$ and $HR_{m,t-1}$. I estimate all regressions using ordinary least squares (OLS), clustering standard errors at the municipality level.[11]

1.5.1 Baseline Specification

The empirical strategy exploits the time-municipality variation in male versus female homicides in Colombia in order to identify the effect of male-biased violence on male versus female education decisions.[12] I empirically test the relationship between the male homicide rate in a municipality and school enrolment, conditional on the gender of the child. For this purpose, I specify the following model:

$$\begin{aligned} Enrol_{i,m,t} = \beta_0 + \beta_1 HR_{m,t-1} + \beta_2 Male_{i,m,t} + \beta_3 \left(HR_{i,m,t-1} * Male_{i,m,t}\right) \\ + \gamma\chi_{i,m,t} + \delta Z_{m,t} + \mu_m + \lambda_t + \varphi_r(t) + \varepsilon_{i,m,t} \quad (1.1) \end{aligned}$$

where $Enrol_{i,m,t}$ is a binary variable that takes a value of 1 if individual i is enrolled in school in municipality m in year t, and 0 otherwise. $\chi_{i,m,t}$ includes several control variables.[13] First, it encompasses individual controls: age and a dummy indicating

[10] When I use different homicide lags up to the fourth lag, I find that homicides in the years *t-1* and *t-2* have a larger and significant impact on enrolment.

[11] I also use a probit model because the dependent variable is dichotomic. The marginal effects are very similar to the OLS estimates. I use the OLS results because of their easier interpretation.

[12] See Figure 1.4 for the variation of male minus female homicides across time and municipalities in Colombia.

[13] I consider the possible endogeneity of some control variables when describing the results.

if the child has been living in the municipality for the past five years. Second, I include family characteristics because they are important determinants of schooling choices. I include dummy variables for father present, mother present, gender of the head of household (HoH), employment status of the HoH, level of education of the HoH, level of education of the mother, and the relationship of the child with the HoH.[14] Additionally, dummy variables are included to control for children with missing characteristics. Third, I control for household characteristics. A household's income may influence a family's ability to invest in their children's education. Although the main data set does not record household income and savings, Burger et al. (2006) suggest that ownership of a number of non financial assets can proxy for accumulated wealth holdings. Therefore, I use three different proxies for income: a dummy indicating if the house is owned by the family, the type of toilet available in the dwelling, and a dummy for whether the dwelling has electricity.[15] In addition, I include a dummy variable that indicates if the child lives in an urban area. This variable controls for the possibility that individuals residing in rural regions have access to fewer schools and may face higher opportunity costs due to child labor needs at home or agricultural job opportunities. $z_{m,t}$ controls for other municipality characteristics that vary over time and could influence enrolment rates due to changes in economic development of the local areas. For this purpose, I construct three variables: population density[16], municipal income tax per capita[17], and the unemployment rate. Moreover, a dummy variable is added for the municipalities for which I do not have information on income. The term μ_m corresponds to municipality fixed effects, which control for all time invariant characteristics varying at the municipality level, and λ_t is year fixed effects, which account for any national time trends. Furthermore, to control for potential long-run differences in regional development, I include a region-specific linear time trend, $\varphi_r(t)$.[18]

To identify children's exposure to violence, I consider boys and girls of school age, and I match their municipality of residence with the data on male minus female homicides by municipality. I compare the enrolment decisions of children within municipalities with varying levels of violence. Additionally, I compare the differential enrolment of boys and girls for different levels of violence. Because I control

[14] Five categories are included: child, grandchild, child-in-law, sibling, or other relative.

[15] Quality of the toilet is divided into three categories: flush toilet, non-flush toilet, and no toilet.

[16] See Cornwell and Trumbull (1994).

[17] Municipal income tax is used as a proxy for the wealth of the municipality. This measure comes from Departamento de Planacion Nacional (DNP).

[18] Since the study includes three points in time, I include just a linear trend.

for municipality fixed effects, year fixed effects, and regional trends, I use changes in school enrollments of boys versus girls over time within municipalities (beyond those predicted by national trends) to identify the gender differential effect of male minus female homicides. As a result, estimates will measure the average effect of male-biased violence on school enrolment conditional on gender with reference to municipality averages and year averages.

The main objective of this paper is to determine if changes in violence have an impact on the education gender gap. Given that young males are the main victims of homicide in Colombia as well as the main participants in narcotraffic and gang activities, based on the conceptual framework I expect to find a differential effect between enrolment rates of girls versus boys. According to equation (1.1), the main variable of interest is the interaction term between the male homicide rate minus the female homicide rate and the male dummy (β_3). I denote the expected change in the gender gap in school enrolment due to an increase in the homicide rate of δ as ($\beta_3 * \delta$). Consequently, the expected change in enrolment for girls and boys would be captured by ($\beta_1 * \delta$) and ($(\beta_1 + \beta_3) * \delta$), respectively.

In light of the nature of violence in Colombia, parents may regard this violence as a proxy for factors such as new job opportunities, changes in life expectancy, and changes in perceived safety, which in turn will affect the expected returns to education. Since homicide victims are mostly male, more girls relative to boys are likely to be enrolled in school in violent areas, producing a negative coefficient for β_3. Thus, I expect the gender gap in education to be wider in municipalities where violence increases. β_1 can be either positive or negative for girls because enrolment in school could be affected positively or negatively in municipalities where violence increases. For example, if families face negative income shocks due to changes in male-biased violence, girls and boys may be affected negatively if financial resources are constrained and families cannot invest in the education of their children. In contrast, if families are receiving extra income because of the money boys are earning from getting involved in violent activities, these additional resources may be invested on girls' education and positively affect their enrolment. For boys ($\beta_1 + \beta_3$), an increase in violence is expected to have a negative effect on school enrolment. That is, the more male-biased violence there is in a municipality, the lower the probability that boys will go to school.

To consistently estimate the causal effect of violence on the gender gap in education, the main identifying assumption is that within municipalities, changes in violence are not correlated with unobserved changes in the determinants of enrolment for girls versus boys. There are two main threats to this identification. First, there could be other unobserved characteristics that affect education in places where violence is high. If this is the case, there may be time-varying omitted variables that will bias the estimates. For example, if municipalities with higher levels of violence also experience changes in job opportunities not related to violence, I may be wrongly attributing the negative effect on education to violence. By the same token, if the state presence in a municipality also changes over time affecting violence and investments in the education sector and this in turn is positively (negatively) correlated with enrolment decisions, the estimates of violence will be biased downward (upward). The second potential threat is the relationship between education and violence. The identifying assumption would be violated if municipalities affected by violence are also those with lower educational levels and this in turn is driving the changes in violence. I tackle the main potential threats to identification in Section 1.6.

1.5.2 Baseline Results

Table 1.5 shows the results of equation (1.1) and includes the three years of the census (1985, 1993, and 2005). I consider three different age groups based on levels of schooling: primary school age (columns 1–4), secondary school age (columns 5–8), and upper-secondary school age (columns 9–12).[19] Each column corresponds to a different set of control variables. In all specifications, I control for year fixed effects, municipality fixed effects, and a linear regional trend. Columns 1, 5, and 9 display the results with no additional controls. In columns 2, 6, and 10, I include a first set of individual controls: age, education level of the head of household, ownership of the dwelling, type of toilet in the household, and electricity. Some family and municipality characteristics may affect enrolment decisions and may also be, in turn, affected by violence. I include a second set of individual controls that incorporates some of these variables as a robustness check. In columns 3, 7, and 11, I control for migration status, the education level of the mother, the child's relationship with the head of household, father present, mother present, the employment status of the head of household, a dummy for urban location, and a dummy for whether the child lives in a metropolitan area. Similarly, to control for changing municipality characteristics, columns 4, 8, and 12 include population density, municipality income

[19]These specifications are also estimated based on the total number of homicides per municipality. Results are very similar since the correlation between the homicide rate and the male homicide rate is very high (0.9). Estimates are available from the author upon request.

tax per capita, and the unemployment rate.

In general, for all ages, the estimates of the main variable of interest (β_3)—the interaction term between the male minus female homicide rate and the dummy for male ($HR * Male$)—are very stable to the inclusion of all the controls. The size of the coefficient increases when I add the first set of individual controls (columns 2, 6, and 10), suggesting that selection on observables attenuates the estimated effect of male homicide on the gender gap in education.[20] The inclusion of the second set of individual controls (columns 3, 7, and 11) and the municipality characteristics (4, 8, and 12) that are potentially endogenous because they are affected by violence, do not change the estimates.

I do not find a statistically significant effect at the primary school level. This result is not surprising since primary education in Colombia is mandatory and free. Furthermore, the increase in the opportunity cost of school due to violence is expected to be more important for older children since at these ages they can engaged in violence-related activities. The effect of violence changes considerably for enrolment of children of secondary school age and even more if we look at those of upper-secondary school age. The coefficient becomes negative and strongly significant. Male-biased violence more negatively affects the school enrolment of boys relative to girls. There is a negative impact of violence on the enrolment in school of boys compared to girls. I focus mainly on the impact of violence on children aged 15–17 because this is the age group at which violence has a larger differential effect between genders. For these ages, an increase of one standard deviation in male-biased violence entails an increase in the gender gap in enrolment of 1.1 percentage points (columns 9–12).[21] This effect is quite large considering that the gender gap in enrolment in upper-secondary school in Colombia favors girls by around 8 percentage points (Hausmann et al., 2012).

The positive coefficient of HR (β_1) indicates that male-biased violence is positively associated with school enrolment of girls. For all ages, the estimates get smaller when I include the two sets of individual controls. The size of the coefficient for HR decreases, suggesting that observable characteristics of families play an important role in education decisions. At the secondary-school age level, the estimate is no longer statistically significant after the inclusion of municipality characteristics that vary over time (column 8), suggesting that omitted variables bias upwards the

[20]In particular, estimates suggest that the inclusion of the level of education of the head of household is very important. I conduct a further analysis in the robustness checks section.

[21]The standard deviations of male minus female homicide rate and enrolment in upper-secondary school are 0.12 and 0.4815, respectively.

effect of the violence (HR). The coefficient of the male dummy (β_2) is statistically significant and negative for all specifications. On average and for all age groups, fewer boys than girls go to school (around 2 to 4 percent). This difference gets larger with years of age. Furthermore, when looking at the total impact for boys, (β_1+β_3) is either negative or zero for all school ages. Estimates suggest that the differential effect of violence is more pronounced at the upper–secondary school age (columns 10–12). The sample used in this analysis does not include children who died or that are members of guerrilla and paramilitary forces because I only have information on children living in their family house. Since most of these children are boys, the effect of male-biased violence on the gender gap and on boys may be underestimated.

Not finding an effect at the primary school age and finding a negative one that increases with age at the secondary school level helps to shed light on the importance of the channels through which male-biased violence is affecting the gender gap in education. Although both mechanisms seem to play a role in investments in education, changes in the opportunity cost of schooling for boys due to new violence-related activities appears to be the most important one. Homicide can be considered as a proxy for gang activity/delinquency and therefore new job opportunities for older children. In Colombia, the use of young men as hired killers is a reality.[22] On the other hand, the male population between the ages of 18 and 35 face a higher risk of homicide. Hence, reduction in life expectancy and in perceived safety are also important to consider when making educational investment decisions because these factors reduce the returns to school. I find that male-biased violence has a negative effect on the gender gap in education that increases with age, suggesting that although changes in life expectancy and perceived safety are important, their effect is triggered by changes in the violence-related job opportunities.

Over the last two decades, Colombia has experienced two turning points in the history of its violence: the consolidation of the drug cartels due to the growth of the cocaine trade during the late 1980s, and the solidification of the guerrilla groups after the mid–1990s. I divide the sample into two subsamples, 1985–1993 and 1993–2005, to explore whether changes taking place in Colombia during these two periods affected school enrolment. Because evidence shows that the main impact of male-biased violence coalesces during secondary school age, Table A1.1 in Appendix A1 shows the results for these ages. Estimates on the effect of male-biased

[22]For example, Medellin during the 1990s had 400 gangs with approximatley 10,000 child members (http://watch list.org/reports/pdf/colombia.report.es.pdf).

violence on the gender gap in education are very similar for the two periods, suggesting that male violence has a negative effect on the education of boys relative to girls, irrespective of the source of violence. In comparison, the positive estimate of (β_1) in the first period is reversed, meaning that higher levels of male violence are associated with a negative or not statistically significant effect on the enrolment of girls at the secondary and upper-secondary school age. According to a United Nations Refugee Unit report (2002), a quarter or more of the Colombian guerrillas are women. Although the male/female homicide ratio remains high throughout the study period, women began to take a more active role in the conflict, which could explain the change in the coefficient for girls.

1.6 Robustness Checks

This section presents several robustness checks in order to account for the potential endogeneity of violence. First, I control for the potential presence of omitted variable biases at the municipality and household levels by introducing municipality–year fixed effects or municipality–gender fixed effects, and family fixed effect (Section 1.6.1 and 1.6.2, respectively). Second, I control for coca production as an additional source of bias in Section 1.6.3. Third, as an alternative measure of gender-biased violence, I explore the effect of relative changes in the homicide rate of women relative to men in Section 1.6.4. Fourth, I look at the effect of migration in Section 1.6.5.

1.6.1 Year–Municipality Fixed Effects

Municipality fixed effects account for all municipality-specific and time-invariant characteristics. On top of this, year fixed effects control for all time-specific, spatial-invariant conditions that could bias the estimates. One concern is that the male homicide variable is capturing the effect of municipality time-variant characteristics that have an impact on enrolment decisions and that cannot be controlled with the set of control variables used in equation (1.1). To overcome this potential omitted variable bias, I include year-municipality effects. Using this approach, I exploit the variation given by the interaction between the male dummy and the male versus female homicide rate. The model thus becomes

$$Enrol_{i,m,t} = \beta_0 + \beta_1 Male_{i,m,t} + \beta_2 \left(HR_{m,t-1} * Male\right)_{i,m,t-1} + \gamma\chi_i + \lambda_{m,t} + \varepsilon_{i,m,t}, \quad (1.2)$$

where $\lambda_{m,t}$ is the municipality-year fixed effects. In this estimation, the main variable of interest is β_2, which captures the effect of violence on the educational gender gap. One would expect that the larger the changes in male-biased violence in a municipality, the larger the negative effect on the gender gap in disfavor of men will be. One important consideration is that using all the sets of fixed effects, may be taking away some important variation in the male violence variable. Consequently, the estimation of the gender gap in school should be interpreted as conservative.

Table 1.6 shows the estimates for equation (1.2) for three age groups: primary school age (columns 1–3), secondary school age (columns 4–6), and upper-secondary school age (columns 7–9).[23] Estimates of the dummy for male *(Male)* and its interaction with the net male homicide rate *(HR)* are stable to the inclusion of the two sets of individual control variables. The male dummy (β_1) is negative. The coefficient goes from 2 to 5 percentage points, depending on the level of school.

The interaction term—Male*HR—indicates a negative and statistically significant impact of violence on the gender gap. The coefficient (β_2) is not statistically significant for primary school age (column 3), similar to the estimate in Table 1.5 (column 3), but it is significant at secondary and upper-secondary school age (columns 4 and 9, respectively). At these levels, violence has a larger, negative, and strongly significant effect on the gender gap in education. As in Table 1.5, at the upper-secondary school age, one standard deviation in violence leads to a 1.1 percentage point increase in the enrolment gap. Thus, in regions with high rates of violence, the gap in enrolment is even wider than in other parts of Colombia. This gap grows even more during the last three years of secondary school. Estimates of the effect of violence on the gender gap in education are consistent and stable after controlling for potential unobserved heterogeneity at the municipality-year level that could bias the results.[24] Furthermore, estimates of the violence-induced gender gap

[23] Each column corresponds to a different set of control variables, as in Table 1.5. The only control variables that are not included are population density, municipality income tax per capita, and unemployment rate.

[24] The main narcotraffic cartels were located in Medellin and Cali. As a robustness check, I remove these two potential outliers from the sample in additional specifications. Furthermore, municipalities with a homicide rate of zero are also dropped. The results hold consistent with previous findings and are available upon request from the author.

is persistent when dividing the sample by time period (1985–1993 and 1993–2005).[25]

Although the methodology does not allow me to estimate the correlation between the male versus female homicide rate in a certain municipality and the probability of being enrolled in school ($HR_{m,t-1}$), by exploiting the variation given by the interaction term, I am able to consistently estimate the differential impact that violence has on the gender gap in enrolment among certain age groups.

An additional potential concern is that there are persistent differences in outcomes across genders and municipalities that could bias the results. To control for this, I include municipality fixed effects, which I allow to vary over gender. Results are very similar to those presented in Table 1.5, although the point estimates are smaller (see Table A1.3 in Appendix A1). In general, at the secondary and upper-secondary school age, there is a negative effect of male-biased violence on the gender gap in education in disfavor of boys. Furthermore, it seems that this measure of violence is positively associated with girls' education.

1.6.2 Intra-Household Estimations

Families compare their children's expected returns to school with the direct and indirect costs associated with its acquisition. I consider violence to be an additional cost that may alter investment decisions in education. Families may reduce investments in education when confronted with violence because income decreases and the return to education may shrink due to outside job opportunities or/ and reductions in life expectancy and perceived safety. I use a family fixed effects model to estimate the differential effect of violence on boys and girls, thereby differencing out any family-specific or municipality characteristics that affect children within families. This fixed effect model looks at the enrolment differences within families residing in violent regions that have both a boy and a girl. The family fixed effects approach translates into the following specification:

$$Enrol_{i,m,j} = \beta_0 + \beta_1 Male_{i,m,j} + \beta_2 \left(HR_{m,t-1} * Male_{i,m,j}\right) + \gamma\chi_i + \lambda_j + \varepsilon_{i,m,j}. \quad (1.3)$$

χ_i controls for age, and λ_j is family-fixed effects. The parameter of interest is β_2, which captures the effect of violence on the enrolment of boys that have a sister. Based on the theoretical framework, I expect β_2 to be negative. Boys are

[25]Results are in Table A1.1 in Appendix A1.

more affected by crime in terms of life expectancy and job opportunities; therefore, families will adjust their investment decisions by favoring the education of the girls in the household.

I restrict the sample to families with just two school-age children. Table 1.7 displays the results of equation (1.3). Column 1, which looks at primary school age, shows no differential effect between genders due to violence is found *(HR*Male)*. The estimate is negative and small but not statistically significant. The coefficient for the male dummy is negative and statistically significant, even though it is very small (0.009 percent). Columns 2 and 3 show the impact of the male-biased homicide rate *(HR)* on secondary and upper-secondary school age. The estimate of the interaction term between the male dummy and *HR* increases in size and becomes statistically significant. The differential effect of violence by gender on school enrolment, (β_2), shows that as children get older a larger differential effect between boys and girls occurs in families that face higher levels of violence. At the upper-secondary school age, the size of the coefficient is double that of the secondary school age (going from 5 percent to 10 percent), suggesting that older boys are the most affected by violence. The size of the effect is very robust across estimations from equations (1.1), (1.2), and (1.3). At the secondary and upper-secondary school age, the male dummy remains negative and the size of the coefficient increases with age (columns 2 and 3). These results are consistent with the previous findings in Tables 1.5 and 1.6. The results indicate that families and children consider violence and crime opportunities when making their investment decisions in education.[26]

This methodology has some limitations. First, if parents have gender-based educational preferences unrelated to violence, their choices will be incorrectly attributed to the effect of violence and β_2 could be overestimated. Second, the restricted sample includes only children with one sibling, which is around 30 percent of the total sample and could limit representativeness.[27]

Finally, to make the results more comparable to the initial findings, I also estimate equation (1.1) with the current sample, including the same set of control variables as in the baseline estimation. Table A1.5 in Appendix A1 provides the results for families with two children grouped by the number of sons in the household.

[26]This sample only considers children who live in a household headed by a parent. The same model is estimated including children who live in a household headed by a relative who is not necessarily their parent. As an additional robustness check, I estimates equation (1.3) using the different samples (1985–2005, 1985–1993, and 1993–2005), and the results are consistent with previous estimates.

[27]See the distribution of the number of children per family in Appendix A1, Table A1.4.

Violence increases the probability of being enrolled in school for girls when there is a boy in the household, especially at the secondary and upper-secondary school age. Additionally, the effect on the gender gap *(HR*Males)* is always negative and statistically significant. The negative effect for boys increases with age. No significant effect is found in families with only girls or only boys, but the signs of the coefficients are as expected: negative for boys and positive for girls.

1.6.3 Coca Production

Coca leaf cultivation rose from 20,000 hectares in 1990 to 160,000 in 2000. Angrist and Kugler (2008) consider violence in urban versus rural regions that grow coca and find that violence increases more in the rural parts of the coca-growing regions that have experience prior conflict. Furthermore, male-biased job opportunities, such as harvesting coca, could affect school enrolment decisions. Thus, one concern is that the reduction in enrolment rates of boys is due to this new job opportunity (coca cultivation) but not to violence-related activities. Furthermore, coca production may also increase job opportunities for girls. Both women and men cultivate and harvest coca. Hence, women will also have a new labour outlet, which could increase their opportunity cost of going to school.

I take two additional approaches to account for the effect of coca production on schooling decisions. First, I create a dummy variable indicating if the municipality has ever produced coca, and I interact it with the homicide rate to measure the impact of male-biased violence and coca production on education. The following model aims at separating the possible effects of cultivating coca and residing in a violent area:

$$Enrol_{i,m,t} = \beta_0 + \beta_1 HR_{m,t-1} + \beta_2 Male_{i,m,t} + \beta_3 (HR_{m,t-1} * Male_{i,m,t}) + \beta_4 (Coca_m * HR_{m,t-1})$$
$$+ \beta_5 (Coca_m * HR_{m,t-1} * Male_{i,m,t}) + \gamma \chi_{i,m,t} + \delta Z_{m,t} + \mu_m + \lambda_t + \varphi_p(t) + \varepsilon_{i,m,t}, \quad (1.4)$$

where $Coca_m * HR_{m,t-1}$ is the interaction between a dummy variable that takes a value of 1 if the municipality has ever had coca cultivation and the average rate of male minus female homicides in years $t-1$ and $t-2$ in the same municipality. Additionally, $(Coca_m * HR_{m,t-1} * Male_{i,m,t})$ is the previous interaction term times the dummy variable that indicates whether the child is a boy. The other control variables are the same as in equation (1.1). I focus on the impact of residing in an area with coca cultivation and high levels of violence, which is captured by β_4. Additionally, β_5 measures the differential effect between boys and girls who reside in areas with coca cultivation and a high level of violence.

Table 1.8 displays the estimates of equation (1.4).[28] For every school age group, the interaction term between the dummy for coca and the male minus female homicide rate (β_4) is negative and not statistically significant independent of the set of control variables considered. These results suggest that having coca production in places with high levels of violence is not associated with school enrolment. Furthermore, when I include a triple interaction between the homicide measure, the dummy for male, and the dummy for coca production *(HR*Male*Coca)*, I get a negative and not statistically significant coefficient. These results indicate that coca production in addition to high levels of violence in a municipality does not differentially affect the school enrolment of girls versus boys. This empirical evidence is in line with Angrist and Kugler's (2008) finding that coca production seems to have little impact on school enrolment. The estimates for the main variables of interest—the male versus female homicide rate (β_1) and its interaction with the male dummy (β_3)— are robust to the inclusion of these additional interaction terms, and the coefficients resemble those reported in Table 1.5. Within municipalities, higher levels of male-biased violence lead to an enlargement of the gender gap in favor of women, independent of coca production. This approach has one drawback: I cannot measure the effect of the intensive margin of coca cultivation and its relation with violence and education. Since I do not have data on the amount of coca produced before 1998, I cannot measure the impact of the increase in coca harvesting and how these changes may affect the level of violence.

In the second approach, I exclude from the sample the 161 municipalities that have had coca production or that have been grouped with a municipality that have had coca production. I estimate equation (1.1) with the remaining sample (354 municipalities) to check if the effects of violence are driven by the specifics of these municipalities. Results are shown in Table A1.6 in Appendix A1. The gap in enrolment rates is negative and increasing as school level progresses. Although significance levels are similar to baseline results, the magnitudes of the estimates are smaller for secondary school age, perhaps because I exclude some municipalities with high levels of violence, since violence is positively correlated with coca production.

[28]Each column corresponds to a different set of control variables as in Table 1.5.

1.6.4 Male/Female Survival Ratio

Although most of the homicide victims in Colombia are men, since the 2000s there has been an increase in the number of female homicides, changing the male/female homicide ratio.[29] This trend coincides with guerrilla groups enlarging their forces by recruiting women in addition to men since the late 1990s.[30]

To investigate the potential effect of changes in the share of male homicides versus female homicides, I compute a male/female survival ratio:

$$MFRS = \frac{1 - (male\, homicide_{\,t-1}/male\, population_{\,t-1})}{1 - (female\, homicide_{\,t-1}/female\, population_{\,t-1})} \, ,$$

which determines the relative probability of survival. I estimate equation (1.1) using the male/female survival ratio instead of the difference between the male minus female homicide rate.[31]

This estimation will capture how changes in the number of male homicides relative female homicides affects school enrolment. The main coefficient of interest is the interaction between the survival ratio and the dummy for male (β_3). It captures how changes in the survival ratio impact the educational gender gap *(MFRS*Male)*. Furthermore, an increase in the survival ratio of α is expected to change the enrolment for girls by $(\beta_1 * \alpha)$ and for boys by $((\beta_1 + \beta_3) * \alpha)$.

Table 1.9 shows the results. An increase in the probability of survival among men relative to women *(MFRS)* is negatively associated with the probability of girls going to school (β_1). This implies that in places where fewer men are dying, fewer girls are going to school. Moreover, the effect of changes in the survival ratio on the gender gap captured by (β_3) is positive. In places where men die less frequently, the probability of being enrolled in school is larger for boys than for girls. The effects are larger among children age 15–17: a one standard deviation increase in the survival ratio (0.1039) leads to an increase in the enrolment gender gap of 0.09 percent in favor of boys. The net effect for boys ($\beta_1 + \beta_3$) is positive as well. Consistent with the previous analyses, the inclusion of individual characteristics increases the size of the estimate for the interaction term between homicide and the male dummy but decreases the value of the coefficients for homicide (columns 2 and 5). Furthermore, the results on the gender gap are robust to the inclusion of additional individual controls and municipality (columns 3 and 6).

[29]See Appendix A1, Figure A1.2 for the variation of the male/female homicide ratio across time.
[30]United Nations Refugee Unit (2002).
[31]Control variables are the same as in equation (1.1).

In conclusion, the difference in the probability of going to school between girls and boys is positive when the survival probability increases for boys.[32] Based on the conceptual framework, when families make decisions about education, they compare the returns to education for boys relative to girls; thus, relative changes in the probability of surviving will also affect these decisions.

1.6.5 Migration

Migration is another concern when estimating the impact of violence. Around 10 percent of the children in the sample had migrated in the five years prior the year of the census. Since violence may induce migration, positive or negative selection into migration could bias the results. Estimates may be driven by selective migration of individuals affected by male-biased violence. In Colombia, forced migration is an important concern, and there is evidence that families who migrate usually have lower incomes and wealth. Additionally, these families often face poor schooling and employment opportunities in the places of destination (Ibáñez and Moya, 2010). If this is the case, the effect of violence on the gender gap will be underestimated. On the other hand, if the families who migrate are among the wealthiest and those most affected by violence, the estimates of the effect of violence on the gender gap in enrolment would be overestimating the impact of male-biased violence since these families would have access to better schooling options. The census data do not include individuals' entire migration histories. I only have information about where the child lived for the five years prior to the census year, where the child was born, and information about the place of residence at the time of the census. As a result, I cannot assign with precision the level of violence to which children that migrated in the previous five years were exposed. Therefore, I first determine whether the probability of migration in the last five years is related to the average level of violence. Second, I restrict the sample to people that have not migrated or those who have been living in the municipality for at least five years prior the year of the census, and I compare them with the original sample (migrants and non-migrants).

In Table 1.10, the dependent variable is a dummy for whether the child had migrated in the five years prior the census. I want to determine whether changes in the homicide rate affect the probability of migration. Controls are the same as

[32] I also look at the two samples separately, and the results are consistent with the findings here. These estimates are available upon request from the author.

in equation (1.1), and the columns present estimates when I include different sets of controls for children, separated by age (columns 1–3 and 4–6, respectively). The effect of male minus female homicide rates *(HR)* and for the interaction of this variable with the male dummy *(HR*Male)* are not statistically significant in any case.[33] Thus, migration status of children does not appear to be associated with male versus female homicides. Although migration may be higher in the regions affected by violence, results suggest that migration of children does not respond to changes in violence over time.

In Table 1.11, I compare the estimates of the initial sample (columns 1 and 2) with those of a subsample of respondents who report living in the municipality for at least five years (columns 3 and 4). Columns 1 and 3 show estimates for secondary school age children, and columns 2 and 4 are for upper-secondary school age children. The coefficients on the gender differential effect of violence on education *(HR*Male)* remain relatively stable and significant for the subsample of non-migrants, suggesting that migration does not pose a significant threat to the validity of the results. On the other hand, the differential effect of male versus female violence on the education gender gap is higher among non-migrants than for the full sample, indicating that perhaps children who migrate out of violent districts have fewer educational opportunities. Additionally, the smaller and sometimes less statistically significant estimates in the complete sample may be the result of the potential error in assigning male violence exposure to those who moved, which in turn biases the estimate towards zero. Overall, the results in Tables 1.10 and 1.11 do not show evidence of a bias due to migration.

Another potential concern is that male-biased violence is causing selective migration among boys. Thus, talented boys may be leaving the municipalities where violence is high and looking for better educational opportunities in places with less male-biased violence. If this is the case, my estimates of violence may be capturing the effect of having less able boys in municipalities with higher levels of violence. To account for this, I generate a boys/girls ratio for different levels of school age in every municipality and I estimate the effect of my measure of male-biased violence on this ratio. If violence is causing sorting between talented and untalented boys, I would expect to find more girls relative to boys in regions where violence is high. I do not find a significant effect of male-biased violence on the boys/girls ratio. This suggests that potential migration of more talented boys is not affecting my results (Table A1.7 in Appendix A1).

[33] Results are also not statistically significant when I consider primary school.

1.7 Additional Specifications

This section presents additional specifications that consider the heterogenous effects of violence as well as the effect of violence in the long-term. First, I consider the heterogeneous effects of male-biased violence on families with different levels of education (Section 1.7.1). Second, Section 1.7.2 looks at the effect of violence on child labor. Third, I look at the effect of gender differences in mortality unrelated to homicides in Section 1.7.3. Finally, Section 1.7.4 examines whether violence also has a long-term effect by looking at years of school completed.

1.7.1 Heterogeneous Effects of Violence

This section looks at the heterogeneous effects of male-biased violence between children from families with different characteristics. Exploring whether violence has a differential impact on children allows me to better understand the channels through which male-biased violence affects the educational attainment of boys versus girls. I use the education level of the head of the household in which the child resides as a proxy for wealth. I generate a dummy variable equal to 1 if the head of household has less than a primary school education, and 0 otherwise. Around 51 percent of the children live in a house where the head of the household has less than a primary school education, 34 percent has primary education completed, 10 percent has a secondary school education, and 3 percent has a university education. I estimate a fully interacted model using equation (1.1) by interacting the main coefficients of interest with the dummy that differentiates between families with different levels of education. Table 1.12 displays the results. Columns 1–3 show the estimates for the secondary school age children and columns 4–6 for children ages 15–17. The set of control variables is the same as in Table 1.5. In general, HR is always positive, implying that male-biased violence is positively associated with girls' education. As before, the dummy for male is negative. The estimates for the dummy indicating that the family head has less than a primary education are always negative and increase with age. The probability of being enrolled in school is 10 percent lower when the head of household has less than a primary school education *(Less than primary)* for children ages 11–17 and 14 percent for children ages 15–17. This effect appears to be even larger for boys *(Male*Less than primary)*. There is no differential effect of violence between girls from different educational backgrounds *(HR*Less than primary)*. When looking at the effect on the gender gap, results are similar to those in previous sections. Estimates suggest that boys are always negatively affected relatively to girls when there is an increase in male-biased violence. When violence is high, boys are less likely to go to school *(HR*Male)*. This effect is even larger when the boy lives in a family with a low level of education *(HR*Male*Less than primary)*. The negative effect of violence on the enrolment gap is more than 1.5 times larger in families where the head of household has less than a primary education. This effect of male-biased violence is larger at the upper secondary school age; the point estimate changes from -0.05 to -0.08 (columns 3 and 6, respectively). In general, male-biased violence has a negative impact on the gender gap in education, affecting more boys in comparison to girls. In addition, in families in which the head of household has completed fewer years of education, the effect is much larger. This suggests that changes in opportunities in crime-related activities have a stronger effect on the education gender gap than changes in life expectancy and perceived safety, since one would expect the changes in life expectancy to affect both groups in a similar way. These estimates also suggest that while boys whose families are less educated are already at a disadvantage when it comes to school enrolment, the children of less educated families residing in areas with higher levels of violence are even more affected. As a result, boys in less educated families residing in more violent areas face greater challenges in terms of school enrolment relative to girls. I use the education level of the mother as an additional robustness check (Table A1.8 in Appendix A1). I restrict the sample to children living with the mother in the household (approximately 88 percent of the full sample).[34] In general, boys are less likely to be enrolled in school. There is always a negative and statistically significant coefficient for the interaction term between the male dummy, the dummy that indicates if the mother has less than primary school, and the measure of male-biased violence.

1.7.2 Employment Outcomes

I have shown that within municipalities, gender is a factor in determining how an increase in the rate of male-biased violence affects school enrolment. This effect might also lead to changes in employment outcomes of children. For example, Rodriguez and Sánchez (2012) find that Colombian households in municipalities with armed attacks also experience an increase in child labor and school drop-outs. In this section, I explore the effect of violence on child labor.

[34] Although using this restricted sample could introduce some selection problems if having a mother present in the house is also related to violence, I use these results as a robustness check for the estimates in Table 1.12.

My sample includes only children ages 10 to 17 because the 1985 census only asked this question to children 10 years old or more. Table 1.13 shows the estimates of equation (1.1) using child labour (a dummy variable if the child is working) as an outcome. The first set of results considers children ages 10 to 17 for whom I have information on employment (columns 1–3). In columns 4–6, I further restrict the sample to children ages 15–17 because earlier analyses indicate that violence has the largest impact on education among children in this age range. Although I find no statistically significant effect for boys versus girls *(HR*Male)*, the point estimate is positive as expected. This implies that in violent areas, more boys are working (and not enrolled in school) in comparison to girls. Furthermore, the size of the coefficients is larger when considering older children. The same occurs when looking at the coefficient of girls. No statistically significant effect is found, but the coefficient is negative as one would expect if more girls are going to school.

The coefficient for the male dummy is always positive and significant, meaning that on average more boys than girls are working. This also matches the estimates in Table 1.5, which indicate that on average fewer boys go to school relative to girls, independent of the changes in male-biased violence.

1.7.3 Differences in Mortality Unrelated to Homicides

In this section, I consider the effect of the difference between non-homicide mortality of males and females on the gender gap in school enrolment. I generate a variable equal to male minus female mortality rate (taking out all deaths were the cause of death was homicide). This variable *(GM)* is used to explore whether differences in homicides rates are driving my results and not just differences in mortality unrelated to violence. Furthermore, this measure allows me to capture the importance of changes in life expectancy between males and females in a long-term horizon. I estimate equation (1.1) using gender differences in mortality rates (unrelated to homicide) instead of gender differences in homicide rates. Results are displayed in Table 1.14. Columns 1–3 show the results for children ages 11 to 17, and Columns 4–6 for children ages 15 to 17.[35] I do not find a statistically significant effect of non-violent mortality on the gender gap *(GM*Male)*. The sign of the coefficient is negative and increasing with age, and the size of the coefficients is almost half the size of the effect when using the homicide rate measure (-0.051 and -0.098, respectively). The estimates for GM are also not statistically significant. The coefficients

[35]The set of control variables is the same as in Table 1.5.

change from positive to negative when adding municipality controls at the secondary school age and are always positive at the upper-secondary school age. At these ages, the point estimates are also different and are half the size of the estimates in Table 1.5. As in the previous results, the male dummy is negative and statistically significant. Estimates based on this gender-specific mortality measure suggest that changes in life expectancy by itself does not have an impact on schooling decisions conditional on gender. Since point estimates are smaller when using GM instead of HR, one could think that although differences in life expectancy may play a role when making investments decision in education, its relative importance is triggered when it is accompanied by an outside option (i.e., violence-related opportunities). Furthermore, it may be also the case that since this measure of general mortality is also capturing the probability of dying during late adult life, individuals and families do not strongly weight changes in this measure when making investment decisions in education. On the contrary, since mortality caused by homicides in Colombia occurs mainly during youth and early adulthood life, this would be translated into a large life expectancy loss.

1.7.4 Other Educational Outcomes: Years of School Completed

To study the long-term effect of gender-biased violence on education, I next look at the violence exposure of an individual during his/her school period. By law in Colombia, students should enter school at age 6 and finish at age 17. I use the 2005 census to examine all individuals from age 18 to 32. All of these individuals should have completed school, and they have been exposed to different levels of violence depending on the year they entered school. Using this approach, I am able to consider individuals affected by different levels of violence in different periods of their life. For instance, I consider individuals entering the educational system in 1979 (age 6) and finishing in 1996 (age 17), and individuals entering school in 1994 (age 6) and finishing in 2005 (age 17). This allows me to capture the effect of being exposed to different levels of violence on school completion.

I estimate the following equation based on the methodology proposed by León

(2012):

$$School - Years_{i,m,t} = \beta_0 + \beta_2 Male_{i,m} + \gamma_\tau HR_{m,\tau} + (\psi_\tau HR_{m,\tau} * Male_{i,m}) + \chi_{i,m} + \mu_m + \lambda_t + \varphi_d(t) + \varepsilon_{i,m,t}\,, \quad (1.5)$$

where $School - Years_{i,m,t}$ is the number of years of schooling achieved by individual i born in municipality m in year t. $HR_{m,\tau}$ are three measures of exposure to violence. I aggregate the estimation of γ_τ and ψ_τ to three different periods of exposure to violence $\tau\epsilon\{6–10, 11–14, 15–17\}$: when children were 6–10 (primary school age), 11–14 (basic secondary school age), or 15–17 (upper-secondary school age). Different to León (2012), I use as a measure of exposure the average of each municipality's homicide rate during the aforementioned age ranges and not the average years of exposure to violence. This methodology allows me to examine the school levels at which changes in violence have a larger effect on the investments in education. $\chi_{i,m}$ is a vector of individual time-invariant characteristics, such as race.[36] μ_m is municipality fixed effects, which controls for any specific characteristic of the municipality that might affect the children born there. Year-of-birth fixed effects (λ_t) incorporates time shocks. To account for potential long-run differences in regional development, I include a region–specific time trend $\varphi_d(t)$. Finally, to control for any potential intraclass correlation across individuals born in the same municipality, $\varepsilon_{i,m,t}$ standard errors are clustered at the municipality-of-birth level. My independent variable is years of school completed. I restrict the sample to people living in municipalities founded after 1973 (year of birth of the oldest child considered in the sample), and I truncate the variable $School - Years_{i,m,t}$ to 11 years of education completed (upper-secondary completed).

I focus mainly on the estimates of γ_τ and ψ_κ. The three measures of exposure to violence γ_τ capture the effects of being exposed to male-biased violence during different periods of life on the years of school completed. Furthermore, the interaction of these measures of exposure with the male dummy (ψ_κ) captures the differential effect for boys versus girls. Based on the theoretical framework, I expect to find a negative coefficient for ψ_κ, because males are expected to be affected more negatively than females. In contrast, γ_τ should be positive if families are investing more in the education of the girls or negative if violence affects the income resources of

[36]The 2005 sample contains information on race, which allows me to control for this variable. The importance of education and gender preferences may differ among races. I include dummies for black (11 percent) and indigenous people (6 percent) to control for this.

the families and they cannot invest more in education. Thus, girls should complete more years of education than boys when violence is higher.

Table 1.15 displays the results for equation (1.5). Column 1 shows the effect of violence based on three different periods of exposure when children are of primary school age *(HR_primary)*, basic–secondary school age *(HR_basic_sec)* and upper-secondary age *(HR_upper_sec)*. Column 2 displays the estimates including the male dummy. Column 3 and 4 show the fully interacted model with and without individual controls, respectively. All estimations include municipality-of-birth and year-of-birth fixed effects.

Estimates suggest that women completed more years of school if they were exposed to higher levels of violence between the ages of 15 and 17 *(HR_upper_sec)*. In general, men completed almost 0.5 fewer years of school. The estimates for the interactions between different levels of violence exposure and the dummy for men indicate that the exposure to violence is particularly important during the last three years of high school age and that the effect of violence is different for men and women *(HR_upper_sec*Male)*. These results do no change with the inclusion of individual controls (Column 4). In regions with higher levels of male-biased violence, males completed even fewer years of schooling than females. This estimate indicates that an increase of one standard deviation in violence when men were in upper-secondary school (0.092) implies that men accumulate 0.081 fewer years of school relative to women (column 4). To put these results in context, León (2012) looks at the effect of the civil war in Perú and finds that an additional year of exposure to violence before birth entails that a person will accumulate 0.07 fewer years of education. If the exposure to violence happens at the preschool age, the effect is 0.05.

Note that living in a municipality affected by violence during primary school age or the first part of secondary school age does not have a significant impact on years of education completed. This finding is in line with previous results and suggests that male-biased violence has a long-term gender differential effect on the schooling attained by children. The sample used in this analysis does not include individuals who died because of violence or that are members of guerrilla and paramilitary forces. As a result, if these individuals, mostly men, were also likely to complete fewer years of schooling because they did not go to school in order to participate in violence-related activities or join military forces, the estimates would be a lower bound of the actual effects of male-biased violence on education.

One limitation of this approach is that I cannot fully track individuals' migration histories due to data availability, which could affect violence exposure measures. As a result a restrict the sample to people that have not migrated. Furthermore, the census data include current information of the individuals, so I cannot consider the conditions experienced by adults in their childhood. Omitted characteristics such as education of the parents may be underestimating/overestimating the effects of homicide rate, depending on the correlation between the homicide rate and the omitted childhood characteristics.

1.8 Conclusions

This paper empirically assesses the effect that male-biased violence has on the gender gap in education in Colombia. To identify this effect, I exploit the variation in the male versus female homicide rate across time within municipalities in Colombia. In this study, I cover the two time periods with the highest levels of violence between the 1980s and 2000s. The increase in violence during these two periods is linked to the surge in narcotraffic, the increase of coca production, and the strengthening of guerrilla and paramilitary forces. To deal with the potential endogeneity of violence, I include year-municipality fixed effects, family fixed effects and I control for several variables that may also affect enrolment outcomes. Results are very robust to different sets of control variables and suggest that when male-biased violence increases, the gender gap in education increases in disfavor of men. This finding is consistent with a simple theoretical model in which families include violence as an additional factor in investment decisions in education. Using a gender-specific measure of violence allows me to pin down the potential mechanism through which violence is affecting educational decisions differentially between boys and girls. Male-biased violence affects school enrolment decisions through two channels. First, it increases the opportunity cost of attending school as it increases the value of getting involved in violence-related activities. Second, it may reduce the benefit of attending school as life expectancy and perceived safety falls.

The results in this paper suggest that the deterioration of human capital investments adds to the economic costs of violence. Families may reduce investments in education when confronted with violence if income falls and the returns to education decline. Furthermore, violence seems to influence schooling choices within the household, causing enrolment decisions to be gender specific. In Colombia, changes in crime affect young men more than women due to a greater risk of homicide as well as labor opportunities related to violence. I find evidence that male-biased violence has a negative effect on the education of boys relative to girls. This effect is found at the secondary school age, particularly when children are older and when they come from families with low levels of education. At this school age, a one standard deviation increase in the male versus female homicide rate differential leads to a 1.1 percentage point increase in the enrolment gender gap. This accounts for 14 percent of the gender gap in secondary school. Results suggest that this estimate is not driven by selective migration or coca production. Furthermore, I find evidence that violence has an effect on education conditional on gender, both in the short and long term.

My study focuses in Colombia, but crime and violence is a major concern among developed and developing countries, especially in Latin America today (IADB, 2013). I find that the type of male-biased violence experienced in Latin America is an important source of gender inequality, suggesting that violence may be one of the driving forces behind the widening of the "reverse" gender gap in education. With an annual average of 25 murders per 100,000 inhabitants, Latin America is among the most violent regions in the world. Countries like Mexico, Brazil, and El Salvador among others face a high incidence of violence, drug trafficking, and the proliferation of violent youth gangs (Heinemann and Verner, 2006). As a consequence, understanding how violence and crime affect family educational investment decisions has compelling policy relevance. The results of this paper suggest that one path to promote gender equality in school is to reduce the opportunity cost of schooling of boys living in municipalities with high levels of violence. This could be achieved by increasing penalties for young people who engage in violence-related activities or by having subsidies for families that send their children to school in violent areas.

References

[1] Akbulut-Yuksel, M. (2009). Children of war: The long-run effects of large scale physical destruction and warfare on children. IZA Discussion Paper No.4407.

[2] Akresh, R., and De Walque, D. (2008). Armed conflict and schooling: Evidence from the 1994 Rwandan genocide. World Bank Policy Research Working Paper Series.

[3] Alderman, H., Hoddinott, J., and Kinsey, B. (2006). Long-term consequences of early childhood malnutrition. Oxford Economic Papers, 58(3), 450-474.

[4] Angrist, J.D., and Kugler, A. (2008). Rural windfall or a new resource curse? Coca, income, and civil conflict in colombia. Review of Economics and Statistics, 90(2): 191-215.

[5] Annan, J., Blattman, C., Mazurana, D., and Carlson, K. (2009). Women and girls at war: "Wives", mothers, and fighters in the Lord's Resistance Army. HICN Working Paper 63.

[6] Becker, G. S. (1962). Investment in human capital: A theoretical analysis. Journal of Political Economy, 70(5), 9-49.

[7] Becker, G. S. (1968). Crime and punishment: An economic approach. Journal of Political Economy, 76(2), 169-217.

[8] Behrman, J. R., Pollak, R. A., and Taubman, P. (1982). Parental preferences and provision for progeny. Journal of Political Economy, 90(1), 52-73.

[9] Behrman, J. R., Pollak, R. A., and Taubman, P. (1990). The wealth model: Efficiency in education and distribution in the family. Working Papers 90-16. Department of Economics at the University of Washington.

[10] Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. Journal of Political Economy, 75(4), 352-365.

[11] Blattman, C., and Annan, J. (2010). The consequences of child soldiering. Review of Economics and Statistics, 92(4), 882-898.

[12] Blattman, C., and Miguel, E. (2010). Civil war. Journal of Economic Literature, 48(1), 3-57.

[13] Burger, R. Booysen, F, Van der Berg, S., and Von Maltitz, M. (2006). Marketable wealth in a poor african country. Research paper No. 2006/136, United Nations University World Institute for Development Economics Research.

[14] Chamarbagwala, R., and Morán, H. E. (2011). The human capital consequences of civil war: Evidence from Guatemala. Journal of Development Economics, 94(1), 41-61.

[15] Cornwell, C., and Trumbull, W. N. (1994). Estimating the economic model of crime with panel data. Review of Economics and Statistics, 76(2), 360-66.

[16] Dahan, M., and Gaviria, A. (2003). Parental actions and sibling inequality. Journal of Development Economics, 72(1), 281-297.

[17] Duryea, S., Galiani, S., Ñopo, H., and Piras, C. (2007). The educational gender gap in Latin America and the Caribbean (Inter-American Development Bank WP 600). Washington, DC: Inter-American Development Bank.

[18] Evans, W. N., Garthwaite, C., and Moore, T. J. (2012). The white/black educational gap, stalled progress, and the long term consequences of the emergence of crack cocaine markets (No. w18437). National Bureau of Economic Research.

[19] Formisano, M., Sanchez, F., and Solimano, A. (2005). Conflict, violence and crime in Colombia. Economics and Politics of Civil Wars, Nicholas Sambanis (Ed), Yale University/World Bank.

[20] Galor, O., and Moav, O. (2004). From physical to human capital accumulation: Inequality and the process of development. Review of Economic Studies, 71(4), 1001-1026.

[21] Gaviria, A. (2000). Increasing returns and the evolution of violent crime: The case of Colombia. Journal of Development Economics, 61(1), 1-25.

[22] Hanratty, D., and Meditz, S. (1988). Colombia: A country study. Washington, DC: GPO for the Library of Congress.

[23] Hausmann, R., Tyson, L., and Zahidi, S. (2012). The global gender gap report. World Economic Forum, Geneva, 2012.

[24] Heinemann, A., and Verner, D. (2006). Crime and violence in development: A literature review of Latin America and the Caribbean. World Bank Policy Research—Working Paper No. 4041.

[25] Hjalmarsson, R., and Lochner, L. (2012). The impact of education on crime: International evidence. CESifo DICE Report, 10(2), 49-55.

[26] IADB. (2013). The costs of crime and violence in Latin America and the Caribbean: Methodological innovations and new dimensions. Washington, DC: IADB.

[27] Ibáñez, A. M., and Moya, A. (2010). Do conflicts create poverty traps? Asset losses and recovery for displaced households in Colombia. In the economics of crime: Lessons for and from Latin America (pp. 137-172). University of Chicago Press.

[28] Jayachandran, S., and Lleras-Muney, A. (2009). Life expectancy and human capital investments: Evidence from maternal mortality declines. Quarterly Journal of Economics, 124(1), 349-397.

[29] Krug, E., Dahlberg, L., Zwi, A., Mercy, J., and Lozano, R. (2002). World report on violence and health. Geneva: World Health Organization.

[30] León, G. (2012). Civil conflict and human capital accumulation: The long-term effects of political violence in Perú. Journal of Human Resources, 47(4), 991-1022.

[31] Levitt, S., and Rubio, M. (2000). Understanding crime in Colombia and what can be done about it. Working Papers Series. Documentos de trabajo 003735, Fedesarrollo.

[32] Lochner, L. (2004). Education, work, and crime: A human capital approach. International Economic Review, 45(3), 811-43.

[33] Lochner, L., and Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. American Economic Review, 94(1), 155-189.

[34] Machel, G. (1996). Impact of Armed Conflict on Children. New York: UNICEF.

[35] Moser, C.O, and McIlwaine, C. (2000). Urban poor perceptions of violence and exclusion in Colombia. Washington, DC.: World Bank..

[36] Oster, E., Shoulson, I., and Dorsey, E. (2013). Limited life expectancy, human capital and health investments. American Economic Review, 103(5), 1977-2002.

[37] Pecaut, D., and González, L. (1997). Presente, pasado y futuro de la violencia en Colombia. Desarrollo Económico, 36(144), 891-930.

[38] Rabasa, A., and Chalk, P. (2001). Colombian labyrinth: The synergy of drugs and insurgency and its implications for regional stability. Los Angeles: RAND.

[39] Rangel, A. (1998). Colombia: Guerra en el fin de siglo. Bogotá: Tercer Mundo Editores, p. 12.

[40] Rodriguez, C., and Sánchez, F. (2012). Armed conflict exposure, human capital investments, and child labor: Evidence from Colombia. Defence and Peace Economics, 23(2), 161-184.

[41] Serafino, N. (2002). Colombia: Summary and tables on U.S. assistance (FY 1989–FY 2003). CRS Report for Congress.

[42] Shemyakina, O. (2011). The effect of armed conflict on accumulation of schooling: Results from Tajikistan. Journal of Development Economics, 95(2), 186-200.

[43] Swee, E. L. (2009). On war and schooling attainment: The case of Bosnia and Herzegovina. Households in Conflict Network Working Paper, 57.

[44] Sohnesen, T., and Blom, A. (2006). Is formal lifelong learning a profitable investment for all of life? How age, education level, and flexibility of provision affect rates of return to adult education in Colombia. World Bank Policy Research Working Paper (3800).

[45] United Nations Refugee Unit. (2002). Aprenderas a no llorar: Niños combatientes en Colombia.

[46] United Nations. (2001). Global illicit drug trends. New York: United Nations Office for Drug Control and Crime Prevention.

[47] United Nations. (2006). Violencia, crimen y trafico illegal de armas. Oficina de crimen. Oficina contra la droga y el delito. Bogota-Colombia

[48] United Nations. (2009). World drug report 2009. United Nations Office on Drugs and Crime.

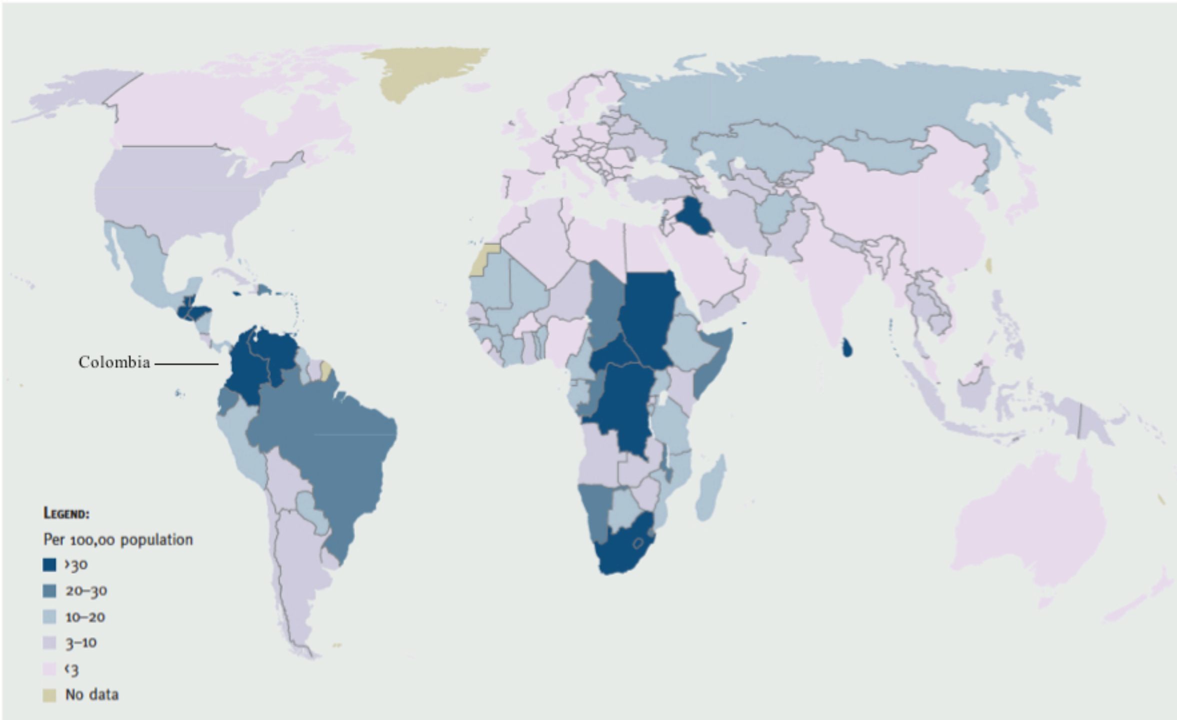
[49] United Nations. (2011). Global study on homicide. United Nations Office on Drugs and Crime.

[50] Valente, C. (2013). Education and civil conflict in Nepal. World Bank Economic Review.

[51] Vargas, J.F. (2009). Military empowerment and civilian targeting in civil war. (HiCN Working Papers 56). Households in Conflict Network. Available: http://www.hicn.org/papers/wp56.pdf. Accessed 4 July 2011.

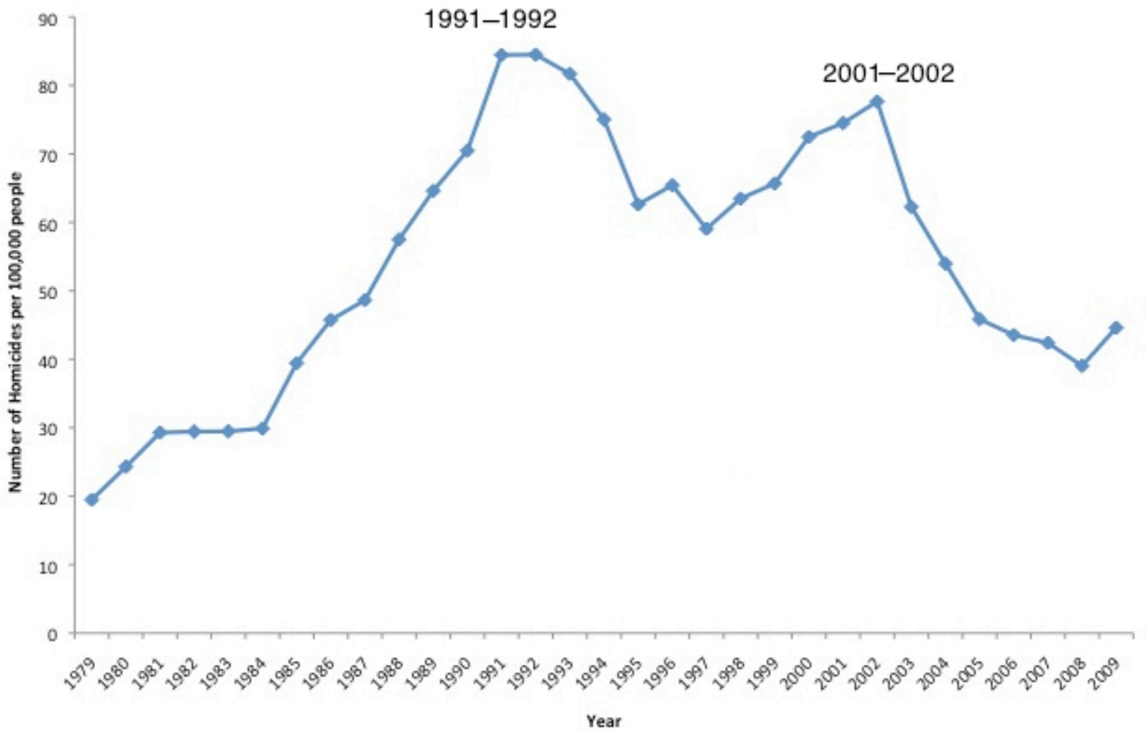
Figures and Tables

Figure 1.1: Average Annual Violent Death Rates per 100,000 (2004–2009)



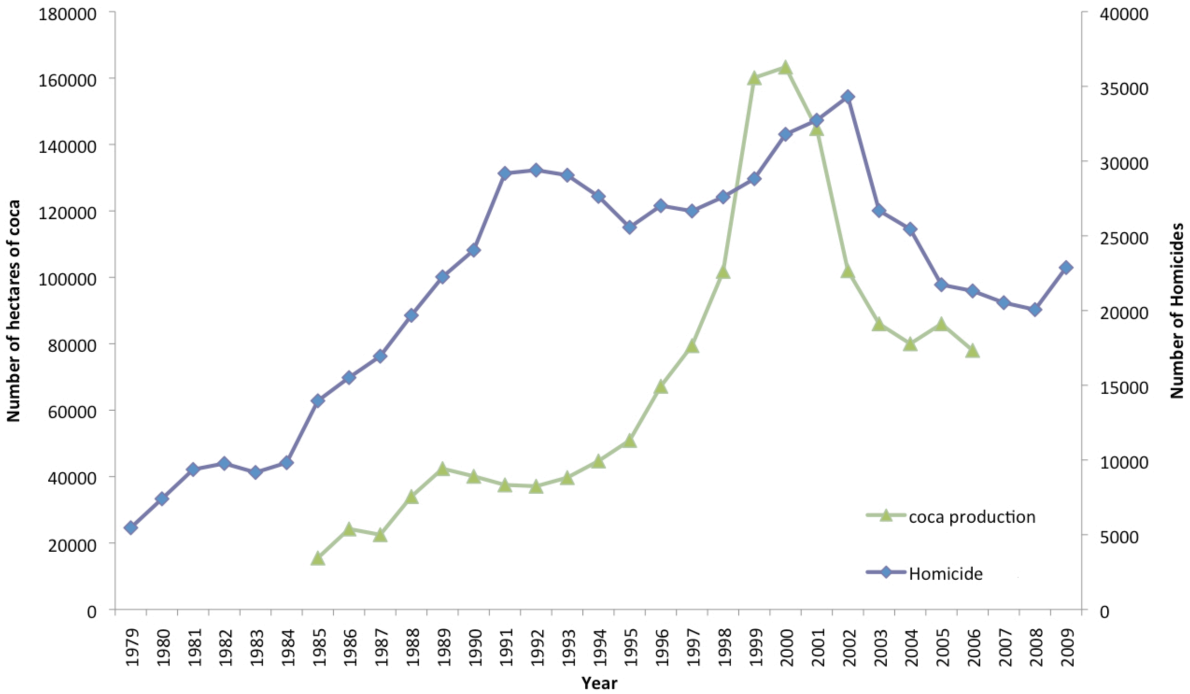
Notes: This graph is taken from "The Global Burden of Armed Violence, 2011" and describes the distribution of the homicide rate around the world.

Figure 1.2: Homicide Rate per 100,000 Inhabitants in Colombia



Notes: Own calculations using data from Departamento Administrativo Nacional de Estadística (DANE).

Figure 1.3: Coca Leaf Production in Colombia



Notes: Own calculations using data on homicides from Departamento Administrativo Nacional de Estadística (DANE) and on coca production from Sistema Integrado de Monitoreo de Cultivos Ilícitos (SIMCI).

Figure 1.4: Variation in the Male versus Female Homicide Rate Differential Across Colombia

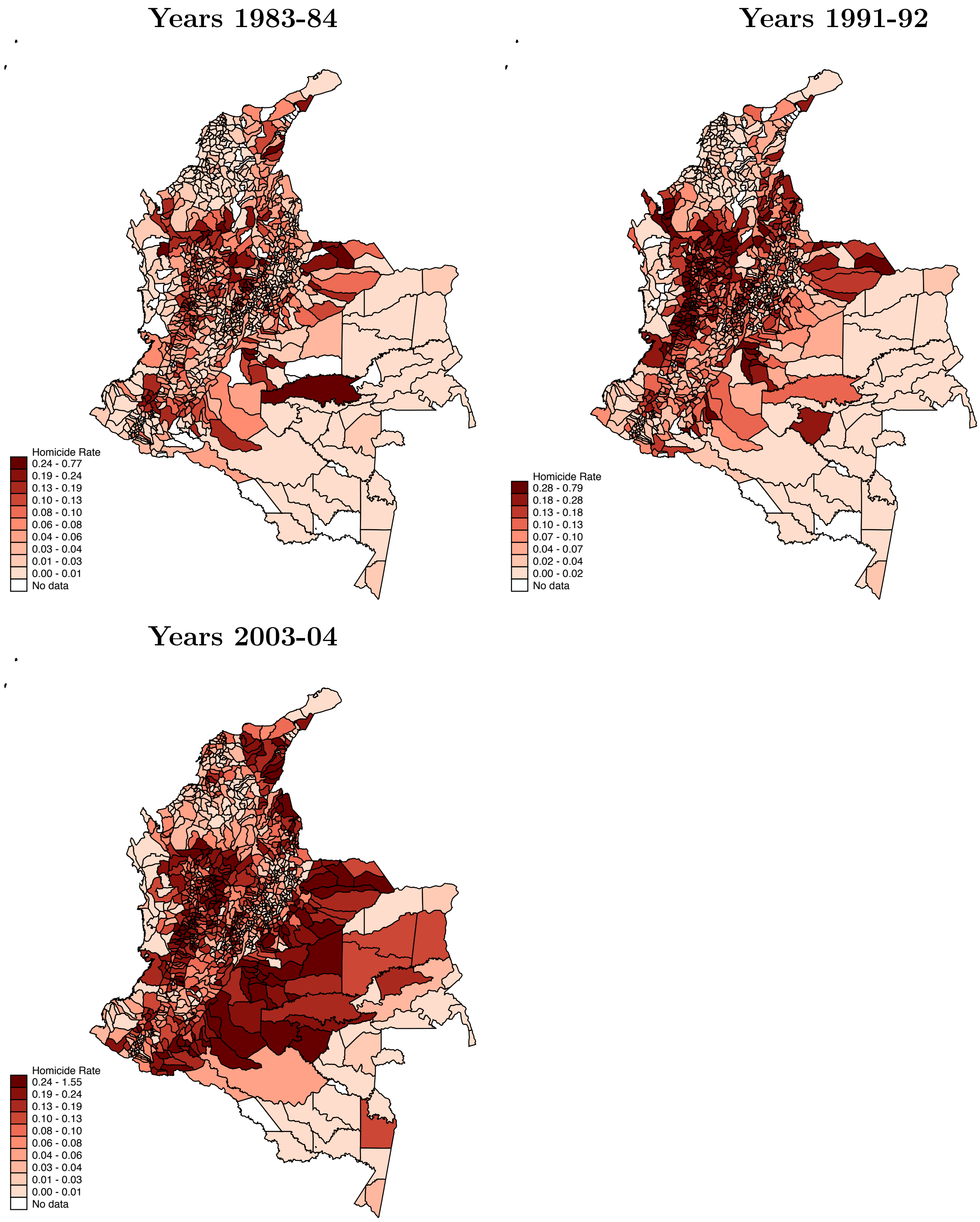


Table 1.1: Education Levels in Colombia

Level	Grades	Age
Preschool	Prekinder, Jardín, Transición	3 to 5 years old
Primary School	First to Fifth Grade	6 to 10 years old
Basic Secondary	Sixth to Eighth Grade	11 to 14 years old
Upper-Secondary	Ninth to Eleventh Grade	15 to 17 years old

Notes: Ministerio de Educación Nacional de Colombia.

Table 1.2: Descriptive Statistics

Variable	Mean	Std. Dev.
Diff. Male HR - Female HR (HR)	0.105	0.112
Male	0.514	0.5
Urban	0.636	0.481
Migrate in the previous 5 years	0.101	0.302
Head of household male	0.774	0.418
Head of household employed	0.726	0.446
Head of household: Less than primary	0.504	0.5
Head of household: Primary completed	0.339	0.473
Head of household: Secondary completed	0.11	0.313
Head of household: University completed	0.028	0.165
Mother present in the household	0.879	0.327
Mother: Less than primary	0.457	0.498
Mother: Primary completed	0.377	0.485
Mother: Secondary completed	0.127	0.333
Mother: University completed	0.023	0.15
Mother: Education Missing	0.016	0.125
Relation to head of household: Child	0.829	0.377
Relation to head of household: Grandchild	0.111	0.314
Relation to head of household: Sibling	0.006	0.055
Relation to head of household: Other relative	0.054	0.226
Father present in the household	0.724	0.447
Ownership of dwelling: Owned	0.676	0.468
Ownership of dwelling: Rented	0.211	0.408
Ownership of dwelling: Free	0.113	0.316
Electricity in the dwelling	0.841	0.365
Flush toilet in the dwelling	0.711	0.453
Non-flush toilet in the dwelling	0.087	0.281
No toilet in the dwelling	0.202	0.402
Number of Observations	2,428,282	

Notes: The table shows summary statistics for children aged 6–17 (10% sample of the Colombian Census 1985,1993 and 2005. Data source: IPUMS-International).

Table 1.3: Descriptive Statistics (Enrolment by school age and gender)

Variables	Census 1985		
	Girls	Boys	Diff. in enrolment
	Mean	Mean	
Primary school age enrolment	0.768	0.747	**0.022**
Secondary school age enrolment	0.725	0.685	**0.040**
Upper-secondary school age enrolment	0.615	0.570	**0.045**
	Census 1993		
	Girls	Boys	Diff. in enrolment
	Mean	Mean	
Primary school age enrolment	0.838	0.817	**0.021**
Secondary school age enrolment	0.784	0.732	**0.052**
Upper-secondary school age enrolment	0.675	0.591	**0.084**
	Census 2005		
	Girls	Boys	Diff. in enrolment
	Mean	Mean	
Primary school age enrolment	0.933	0.921	**0.012**
Secondary school age enrolment	0.856	0.813	**0.043**
Upper-secondary school age enrolment	0.760	0.698	**0.062**

Notes: The table shows descriptive statistics for enrolment of different school age groups: Primary school age (6–10), Secondary school age (11–17) and Upper-secondary school age (15–17). Difference in enrolment is equal to the mean enrolment of girls minus the mean enrolment of boys in the specified school age group (10% sample of the Colombian Census 1985,1993, and 2005. Data source: IPUMS-International). Boldface indicates statistically significant differences at the 1% level.

Table 1.4: Homicide Victim Characteristics

Gender of the Victim	Number of homicides 1984	%	Number of homicides 1992	%	Number of homicides 2004	%
Male	9,427	91.92	28,192	92.21	22,788	90.72
Female	829	8.08	2,381	7.79	2,318	9.23
No Information	0	0.00	0	0.00	14	0.06
Total	10,256		30,573		25,120	
Age of the Victim						
0-5	61	0.59	98	0.32	113	0.45
6-10	33	0.33	98	0.32	48	0.19
11-17	396	3.86	1,975	6.46	1,339	5.33
18-25	3,000	29.25	9,560	31.27	7,566	30.12
26-35	3,356	32.72	9,692	31.7	7,230	28.78
36-45	1,648	16.07	4,913	16.07	4,552	18.12
46-55	878	8.56	1,969	6.44	2,233	8.89
56-65	412	4.02	890	2.91	814	3.24
65+	190	1.85	452	1.48	500	1.99
No Information	283	2.76	926	3.03	726	2.89
Total	10,256	100	30,573	100	25,120	100
Marital Status						
Single	6,071	59.19	17,826	58.31	8,051	32.05
Married	3,075	29.98	6,968	22.79	3,308	13.17
Widow	173	1.69	453	1.48	309	1.23
Cohabitating	189	1.84	2,272	7.43	6,157	24.51
Divorced	748	7.29	3,054	9.99	417	1.66
No Information	0	0.00	0	0.00	6,878	27.38
Total	10,256	100.00	30,573	100.00	25,120	100.00
Education Level*						%*
No Education					48	0.37
Preschool					3,287	25.10
Primary Complete					3,501	26.73
Primary Incomplete					2,137	16.32
Secondary Complete					2,612	19.94
Secondary Incomplete					413	3.15
University Complete					249	1.90
University Incomplete					850	6.49
No Information					12,023	
Total					25,120	

Notes: This table describes the characteristics of the homicide victims in Colombia in the years 1984, 1992, and 2004. * Information available just for 2004 and *% without considering missing information. Data source: Estadisticas vitales - Departamento Administrativo Nacional de Estadistica (DANE).

Table 1.5: Enrolment by School Age (Total Sample 1985–2005)

	Primary School Age: 6–10				Secondary School Age: 11–17				Upper-Secondary School Age: 15–17			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
HR	0.0490***	0.0228	0.0237	0.0304	0.0747***	0.0409***	0.0461***	0.0220	0.0900***	0.0615***	0.0678***	0.0458**
	(0.0186)	(0.0202)	(0.0231)	(0.0203)	(0.0125)	(0.0130)	(0.0138)	(0.0142)	(0.0179)	(0.0176)	(0.0183)	(0.0191)
Male	-0.0202***	-0.0186***	-0.0189***	-0.0189***	-0.0387***	-0.0317***	-0.0359***	-0.0358***	-0.0544***	-0.0429***	-0.0521***	-0.0520***
	(0.00290)	(0.00270)	(0.00261)	(0.00261)	(0.00300)	(0.00208)	(0.00214)	(0.00214)	(0.00384)	(0.00330)	(0.00322)	(0.00322)
HR*Male	0.0154*	0.0136	0.0138	0.0136	-0.0389*	-0.0517***	-0.0495***	-0.0500***	-0.0832***	-0.103***	-0.0971***	-0.0978***
	(0.00910)	(0.00873)	(0.00849)	(0.00852)	(0.0204)	(0.0178)	(0.0179)	(0.0180)	(0.0312)	(0.0285)	(0.0279)	(0.0280)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Individual Controls B	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Municipality Controls	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Observations	1,045,820	1,045,820	1,045,820	1,045,820	1,382,462	1,382,462	1,382,462	1,382,462	549,893	549,893	549,893	549,893
R-Squared	0.106	0.193	0.204	0.205	0.084	0.209	0.230	0.230	0.106	0.200	0.231	0.231

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in *t*-1 and *t*-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain year fixed effects, municipality fixed effects, and a linear regional trend. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, missing information on income tax, and unemployment rate. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table 1.6: Municipality Year Fixed-Effects (Sample 1985–2005)

	Primary School Age: 6–10			Secondary School Age: 11–17			Upper-Secondary School Age: 15–17		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Male	-0.0201***	-0.0186***	-0.0189***	-0.0386***	-0.0318***	-0.0359***	-0.0542***	-0.0430***	-0.0522***
	(0.00289)	(0.00270)	(0.00261)	(0.00299)	(0.00209)	(0.00215)	(0.00381)	(0.00329)	(0.00321)
HR*Male	0.0145	0.0130	0.0132	-0.0393*	-0.0517***	-0.0496***	-0.0828***	-0.101***	-0.0959***
	(0.00912)	(0.00876)	(0.00852)	(0.0205)	(0.0179)	(0.0181)	(0.0311)	(0.0282)	(0.0277)
Municipality-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes	No	No	Yes
Observations	1,045,820	1,045,820	1,045,820	1,382,462	1,382,462	1,382,462	549,893	549,893	549,893
R-Squared	0.118	0.201	0.212	0.089	0.212	0.233	0.112	0.205	0.236

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain municipality-year fixed effects. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table 1.7: Family Fixed-Effects by Education Level

	Primary School Age	Secondary School Age	Upper-Secondary School Age
	(1)	(2)	(3)
Male	-0.00940***	-0.0300***	-0.0555***
	(0.00213)	(0.00233)	(0.0153)
MH*Male	-0.00681	-0.0555***	-0.109***
	(0.0115)	(0.0153)	(0.0407)
Family Fixed Effects	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes
Observations	76,806	150,882	19,884
R-Squared	0.050	0.101	0.066
Number of Families	38,403	75,441	9,942

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). Primary school age (6–10), Secondary school age (11–17) and Upper-Secondary school age (15–17). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in *t*-1 and *t*-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain family fixed effects and age as an individual control. Data source: IPUMS 1985, 1993 and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table 1.8: Municipalities with Coca Production

	Secondary School Age: 11-17			Upper-Secondary School Age: 15-17		
	(1)	(2)	(3)	(4)	(5)	(6)
HR	0.0770***	0.0420***	0.0214	0.100***	0.0705***	0.0526***
	(0.0109)	(0.0119)	(0.0143)	(0.0167)	(0.0158)	(0.0174)
Male	-0.0386***	-0.0317***	-0.0357***	-0.0544***	-0.0430***	-0.0521***
	(0.00278)	(0.00192)	(0.00193)	(0.00374)	(0.00334)	(0.00320)
HR*Male	-0.0278	-0.0452***	-0.0410**	-0.0797**	-0.106***	-0.0980***
	(0.0173)	(0.0169)	(0.0161)	(0.0323)	(0.0332)	(0.0313)
Coca*HR	-0.0206	-0.0108	-0.00254	-0.0624	-0.0492	-0.0379
	(0.0461)	(0.0366)	(0.0299)	(0.0534)	(0.0426)	(0.0360)
Coca*HR*Male	-0.0423	-0.0249	-0.0344	-0.0117	0.0154	0.00161
	(0.0431)	(0.0348)	(0.0327)	(0.0625)	(0.0543)	(0.0519)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1,382,462	1,382,462	1,382,462	549,893	549,893	549,893
R-Squared	0.084	0.209	0.230	0.106	0.200	0.231

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in *t*-1 and t-2. Male is a dummy equal to 1 if the child is a boy. Coca is a dummy variable that indicates if the municipality has ever cultivated coca. All regressions contain year fixed effects, municipality fixed effects, and a linear regional trend. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, missing information on income tax, and unemployment rate. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table 1.9: Survival Ratio

	Secondary School Age: 11-17			Upper-Secondary School Age: 15-17		
	(1)	(2)	(3)	(4)	(5)	(6)
MFSR	-0.0717***	-0.0399***	-0.0221*	-0.0862***	-0.0594***	-0.0447**
	(0.0116)	(0.0118)	(0.0130)	(0.0165)	(0.0160)	(0.0174)
Male	-0.0757***	-0.0810***	-0.0834***	-0.133***	-0.140***	-0.145***
	(0.0180)	(0.0160)	(0.0160)	(0.0278)	(0.0255)	(0.0250)
MFSR*Male	0.0369*	0.0491***	0.0474***	0.0788***	0.0970***	0.0926***
	(0.0196)	(0.0172)	(0.0174)	(0.0300)	(0.0275)	(0.0271)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1,382,462	1,382,462	1,382,462	549,893	549,893	549,893
R-Squared	0.084	0.209	0.230	0.106	0.200	0.231

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). Columns represent OLS coefficients. MFSR is defined as the average of (1–Male homicide rate) /(1–Female homicide rate) in *t*-1 and *t*-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain year fixed effects, municipality fixed effects, and a linear regional trend. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, missing information on income tax, and unemployment rate. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.10: Probability of Migration

Dependent Variable: Migration	Secondary School Age: 11–17			Upper-Secondary School Age: 15–17		
	(1)	(2)	(3)	(4)	(5)	(6)
MH	-0.00118	-0.00565	0.00214	0.00892	0.00198	0.00987
	(0.0148)	(0.0155)	(0.0120)	(0.0138)	(0.0141)	(0.0119)
Male	-0.00677***	-0.00605***	-0.00360***	-0.00885***	-0.00808***	-0.00369***
	(0.000882)	(0.000809)	(0.000825)	(0.00125)	(0.00118)	(0.00113)
MH*Male	-0.00509	-0.00458	-0.00802	-0.0111	-0.00962	-0.0145
	(0.00624)	(0.00606)	(0.00572)	(0.0102)	(0.0104)	(0.00950)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1,381,267	1,381,267	1,381,267	549,461	549,461	549,461
R-Squared	0.031	0.055	0.072	0.030	0.056	0.084

Notes: The dependent variable is being a migrant when within the specified age ranges (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain year fixed effects, municipality fixed effects, and a linear regional trend. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, missing information on income tax, and unemployment rate. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table 1.11: Initial Sample versus Non-Migrant Sample

	All Sample		Non-Migrants	
	Secondary School Age	Upper-Secondary School Age	Secondary School Age	Upper-Secondary School Age
	(1)	(2)	(3)	(4)
MH	0.0220	0.0458**	0.0247*	0.0540***
	(0.0142)	(0.0191)	(0.0138)	(0.0192)
Male	-0.0358***	-0.0520***	-0.0353***	-0.0505***
	(0.00214)	(0.00322)	(0.00212)	(0.00315)
MH*Male	-0.0500***	-0.0978***	-0.0594***	-0.114***
	(0.0180)	(0.0280)	(0.0177)	(0.0276)
Year Fixed Effects	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes
Individual Controls A	Yes	Yes	Yes	Yes
Individual Controls B	Yes	Yes	Yes	Yes
Municipality Controls	Yes	Yes	Yes	Yes
Observations	1,382,462	549,893	1,246,791	496,859
R-Squared	0.230	0.231	0.232	0.234

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in *t*-1 and *t*-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain year fixed effects, municipality fixed effects, and a linear regional trend. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, missing information on income tax, and unemployment rate. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table 1.12: Heterogeneous Effects of Violence by Level of Education of the Head of the Household

	Secondary School Age: 11–17			Upper-Secondary School Age: 15–17		
	(1)	(2)	(3)	(4)	(5)	(6)
HR	0.0640***	0.0530***	0.0359**	0.0685**	0.0697***	0.0599**
	(0.0178)	(0.0161)	(0.0178)	(0.0292)	(0.0257)	(0.0274)
Male	-0.0224***	-0.0210***	-0.0243***	-0.0373***	-0.0335***	-0.0407***
	(0.00191)	(0.00167)	(0.00169)	(0.00336)	(0.00335)	(0.00324)
Less than Primary	-0.136***	-0.0994***	-0.0946***	-0.190***	-0.148***	-0.141***
	(0.00838)	(0.00491)	(0.00412)	(0.00940)	(0.00569)	(0.00478)
HR*Male	-0.0239**	-0.0298***	-0.0297***	-0.0574**	-0.0624***	-0.0632***
	(0.0107)	(0.0100)	(0.0107)	(0.0232)	(0.0228)	(0.0234)
HR*Less than Primary	0.0346	-0.0250	-0.0245	0.0601	-0.0205	-0.0239
	(0.0314)	(0.0245)	(0.0237)	(0.0471)	(0.0364)	(0.0364)
Male*Less than Primary	-0.0290***	-0.0214***	-0.0225***	-0.0273***	-0.0188***	-0.0225***
	(0.00309)	(0.00302)	(0.00298)	(0.00452)	(0.00473)	(0.00453)
HR*Male*Less than Primary	-0.0495**	-0.0554***	-0.0508***	-0.0764**	-0.0933***	-0.0808***
	(0.0232)	(0.0201)	(0.0187)	(0.0321)	(0.0305)	(0.0287)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1,355,699	1,355,699	1,355,699	539,232	539,232	539,232
R-Squared	0.110	0.206	0.223	0.142	0.193	0.218

Notes: The dependent variable is school enrolment of children of different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in *t*-1 and *t*-2. Male is a dummy equal to 1 if the child is a boy. Less than primary indicates if the head of the household has less than primary education. All regressions contain year fixed effects, municipality fixed effects, and a linear regional trend. Individual controls A include age, ownership of the home, type of toilet in the household, type of electricity, and metropolitan location. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, child's relationship with the head of the household, employment status of the household head, a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, a dummy variable for missing information on income tax, and unemployment rate. Data source: IPUMS 1985, 1993 and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Table 1.13: Probability of Employment

Dependent Variable: Employ	Age: 10–17			Age: 15–17		
	(1)	(2)	(3)	(4)	(5)	(6)
MH	-0.0296	-0.0115	0.00763	-0.0610	-0.0430	-0.0207
	(0.0279)	(0.0252)	(0.0230)	(0.0469)	(0.0444)	(0.0415)
Male	0.0952***	0.0919***	0.0917***	0.177***	0.170***	0.170***
	(0.0113)	(0.0107)	(0.0104)	(0.0183)	(0.0175)	(0.0170)
MH*Male	0.0830	0.0882	0.0892	0.139	0.151	0.154
	(0.0659)	(0.0645)	(0.0641)	(0.106)	(0.104)	(0.103)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1,544,356	1,544,356	1,544,356	535,975	535,975	535,975
R-Squared	0.090	0.161	0.172	0.141	0.188	0.207

Notes: The dependent variable is being in employed when within the specified age ranges (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain year fixed effects, municipality fixed effects, and a linear regional trend. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, missing information on income tax, and unemployment rate. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table 1.14: Mortality Unrelated to Homicides

	Secondary School Age: 11–17			Upper-Secondary School Age: 15–17		
	(1)	(2)	(3)	(4)	(5)	(6)
GM	-0.0224	0.0118	-0.0119	0.0142	0.0281	0.0200
	(0.0298)	(0.0247)	(0.0229)	(0.0419)	(0.0344)	(0.0340)
Male	-0.0416***	-0.0347***	-0.0389***	-0.0596***	-0.0578***	-0.0579***
	(0.00311)	(0.00209)	(0.00216)	(0.00395)	(0.00327)	(0.00326)
GM*Male	-0.0146	-0.0298	-0.0267	-0.0417	-0.0524	-0.0513
	(0.0250)	(0.0210)	(0.0207)	(0.0378)	(0.0335)	(0.0335)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1,382,462	1,382,462	1,382,462	549,893	549,893	549,893
R-Squared	0.084	0.209	0.230	0.106	0.200	0.231

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. GM is defined as the average between the male non-homicide mortality rate minus the non-homicide female mortality rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain municipality-gender fixed effects and linear a regional trend. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table 1.15: Years of School Completed

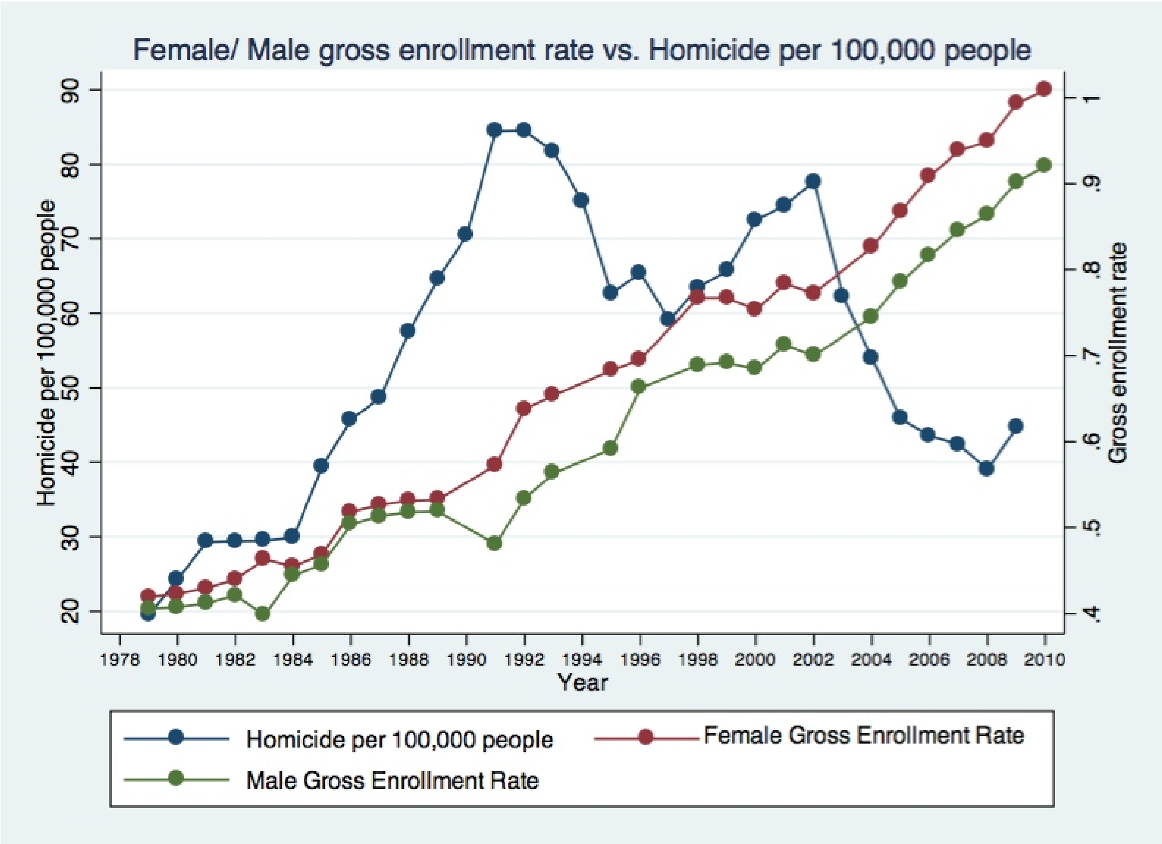
Dependent Variable: Years of school	(1)	(2)	(3)	(4)
HR Primary age	-0.215	-0.184	-0.135	-0.162
	(0.384)	(0.378)	(0.435)	(0.428)
HR Lower-Sec age	0.151	0.189	0.000949	0.0130
	(0.247)	(0.248)	(0.312)	(0.309)
HR Upper-Sec age	0.154	0.155	0.584**	0.587**
	(0.233)	(0.232)	(0.264)	(0.263)
Male		-0.443***	-0.408***	-0.403***
		(0.0380)	(0.0528)	(0.0524)
HR Primary age*Male			0.00640	-0.0128
			(0.252)	(0.254)
HR Lower-Sec age*Male			0.389	0.392
			(0.364)	(0.360)
HR Upper-Sec age*Male			-0.874*	-0.882*
			(0.482)	(0.477)
Birth Mun. Fixed Effects	Yes	Yes	Yes	Yes
Year of Birth Fixed Effect	Yes	Yes	Yes	Yes
Regional Cubic Trend	Yes	Yes	Yes	Yes
Individual Control	Yes	Yes	No	Yes
Mean Dependent Variable		7.52		
Observations	745,499	745,499	745,499	745,499
R-Squared	0.191	0.195	0.188	0.195

Notes: The dependent variable is Number of years of school completed (up to 11). HR Primary age is defined as the average homicide rate between the ages of 6 and 10. HR Lower \-Sec age the average homicide rate between the age of 11 and 14, and HR Upper-Sec age is the average homicide rate between ages of 15 and 17. Male is a dummy equal to 1 if the child is a boy. All regressions contain municipality-of-birth fixed effects, year-of-birth fixed effects, and a regional cubic trend. Race is included as a control variable. Data source: IPUMS 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Appendix A1

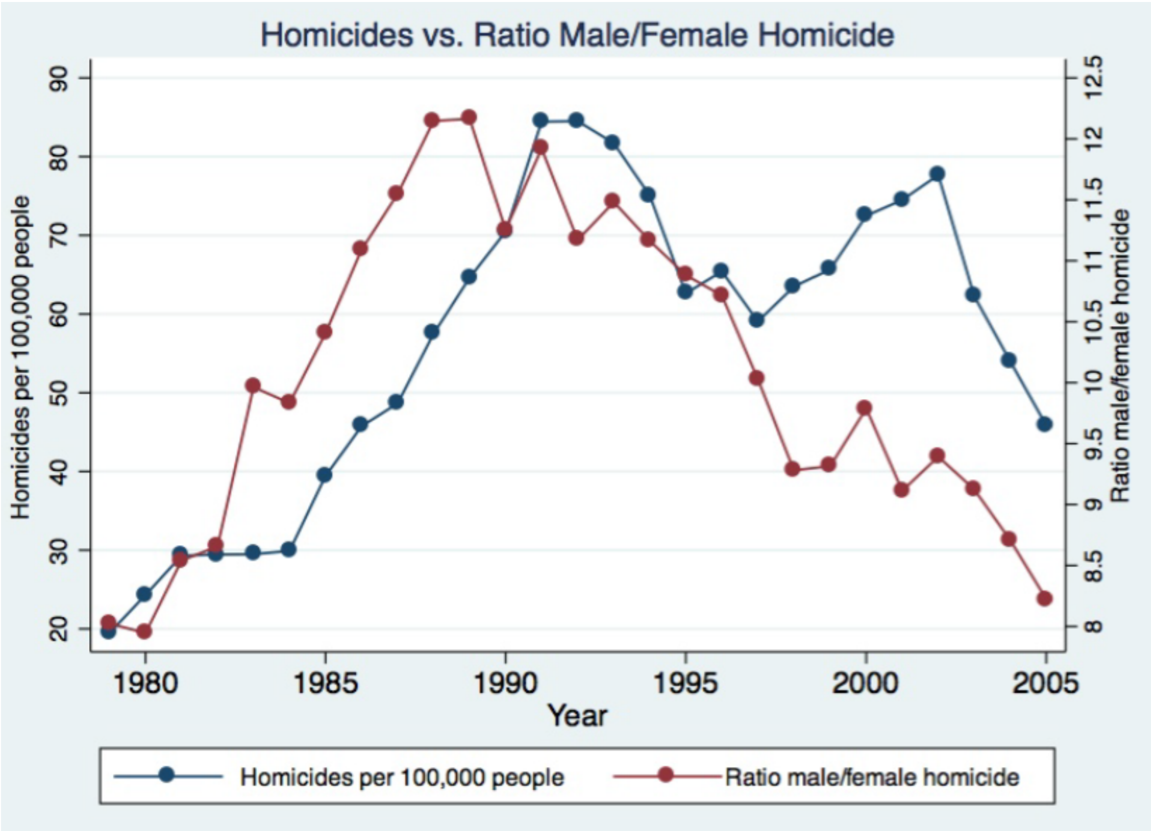
Figures and Tables

Figure A1.1: Enrolment Rates in Secondary School and Homicide Rates in Colombia



Notes: Own calculations using data on homicides from Departamento Administrativo Nacional de Estadística (DANE) for homicides and World Bank data for enrolment.

Figure A1.2: Ratio Male/Female Homicides



Notes: Own calculations using data on homicides from Departamento Administrativo Nacional de Estadística (DANE).

Table A1.1: School Enrolment of Children of Secondary School Age

Panel A : Year 1985-1993

	Secondary School Age: 11–17				Upper-Secondary School Age: 15–17			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HR	0.0733***	0.0599***	0.0648***	0.0721***	0.119***	0.121***	0.124***	0.125***
	(0.0175)	(0.0120)	(0.0117)	(0.0132)	(0.0270)	(0.0213)	(0.0192)	(0.0234)
Male	-0.0410***	-0.0326***	-0.0360***	-0.0360***	-0.0589***	-0.0445***	-0.0522***	-0.0522***
	(0.00265)	(0.00219)	(0.00214)	(0.00213)	(0.00397)	(0.00422)	(0.00409)	(0.00407)
HR*Male	-0.0258	-0.0413***	-0.0391***	-0.0391***	-0.0692**	-0.0923***	-0.0887***	-0.0887***
	(0.0163)	(0.0146)	(0.0144)	(0.0144)	(0.0295)	(0.0273)	(0.0259)	(0.0259)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Individual Controls B	No	No	Yes	Yes	No	No	Yes	Yes
Municipality Controls	No	No	No	Yes	No	No	No	Yes
Observations	850,023	850,023	850,023	850,023	338,202	338,202	338,202	338,202
R-Squared	0.083	0.218	0.239	0.239	0.109	0.215	0.248	0.248

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain year fixed effects, and municipality fixed effects. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, missing information on income tax, and unemployment rate. Data source: IPUMS 1985 and 1993. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Panel B : Year 1993-2005

	Secondary School Age: 11–17				Upper-Secondary School Age: 15–17			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
HR	0.0650***	0.0182	0.0243	-0.0171	0.0542*	0.00144	0.0105	-0.0329
	(0.0217)	(0.0196)	(0.0204)	(0.0206)	(0.0321)	(0.0291)	(0.0295)	(0.0313)
Male	-0.0397***	-0.0337***	-0.0388***	-0.0387***	-0.0558***	-0.0462***	-0.0572***	-0.0572***
	(0.00365)	(0.00266)	(0.00292)	(0.00292)	(0.00516)	(0.00419)	(0.00468)	(0.00468)
HR*Male	-0.0357*	-0.0464**	-0.0422**	-0.0426**	-0.0791**	-0.0927***	-0.0839***	-0.0843***
	(0.0213)	(0.0186)	(0.0188)	(0.0190)	(0.0338)	(0.0299)	(0.0297)	(0.0299)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Individual Controls B	No	No	Yes	Yes	No	No	Yes	Yes
Municipality Controls	No	No	No	Yes	No	No	No	Yes
Observations	984,996	984,996	984,996	984,996	384,650	384,650	384,650	384,650
R-Squared	0.075	0.195	0.219	0.220	0.095	0.186	0.220	0.221

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain year fixed effects, and municipality fixed effects. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, missing information on income tax, and unemployment rate. Data source: IPUMS 1993 and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table A1.2: Municipality-Year Fixed Effects

Panel A : Year 1985-1993

	Secondary School Age: 11-17			Upper-Secondary School Age: 15-17		
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.0409***	-0.0328***	-0.0360***	-0.0589***	-0.0455***	-0.0522***
	(0.00266)	(0.00220)	(0.00213)	(0.00393)	(0.00441)	(0.00406)
HR*Male	-0.0261	-0.0411***	-0.0390***	-0.0678**	-0.0908***	-0.0866***
	(0.0164)	(0.0143)	(0.0144)	(0.0286)	(0.0254)	(0.0250)
Municipality-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Observations	850,023	850,023	850,023	338,202	338,202	338,202
R-Squared	0.086	0.235	0.241	0.114	0.244	0.252

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain municipality-year fixed effects. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Data source: IPUMS 1985 and 1993. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Panel B : Year 1993-2005

	Secondary School Age: 11-17			Upper-Secondary School Age: 15-17		
	(1)	(2)	(3)	(4)	(5)	(6)
Male	-0.0396***	-0.0342***	-0.0388***	-0.0557***	-0.0478***	-0.0572***
	(0.00364)	(0.00262)	(0.00293)	(0.00517)	(0.00405)	(0.00469)
HR*Male	-0.0354*	-0.0457**	-0.0420**	-0.0784**	-0.0906***	-0.0831***
	(0.0214)	(0.0184)	(0.0190)	(0.0340)	(0.0290)	(0.0299)
Municipality-Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Observations	984,996	984,996	984,996	384,650	384,650	384,650
R-Squared	0.080	0.215	0.223	0.101	0.213	0.225

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). Columns represent OLS coefficients. HR is defined as the average between the male homicide rate minus female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain municipality-year fixed effects. Individual controls A include: age, ownership of the household, type of toilet in the household, type of electricity, metropolitan location and education of the head of the household. Individual controls B include: rural or urban location, migration status, missing information on migration, education of the mother, missing information on the mother, relation with household head, employment status of the household head, missing information on employment. Data source: IPUMS 1993 and 2005. Number of municipalities: 515. Standard errors clustered at municipality level reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table A1.3: Municipality-Gender Fixed Effects

	Secondary School Age: 11–17			Upper-Secondary School Age: 15–17		
	(1)	(2)	(3)	(4)	(5)	(6)
HR	0.0710***	0.0418***	0.0178	0.0783***	0.0577***	0.0360*
	(0.0135)	(0.0158)	(0.0157)	(0.0184)	(0.0195)	(0.0198)
HR*Male	-0.0312***	-0.0406***	-0.0413***	-0.0590***	-0.0764***	-0.0776***
	(0.0101)	(0.0118)	(0.0119)	(0.0216)	(0.0223)	(0.0225)
Municipality-Gender Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1,382,462	1,382,462	1,382,462	549,893	549,893	549,893
R-Squared	0.086	0.210	0.231	0.108	0.201	0.233

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain municipality-gender fixed effects and linear a regional trend. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table A1.4: Number of Children per Family

Number of children per family	Number of children	Number of families	Percent	Cum. Per.
1	518,107	518,107	21.34	21.34
2	733,708	366,854	30.22	51.55
3	566,748	188,916	23.34	74.89
4	334,564	83,641	13.78	88.67
5	166,885	33,377	6.87	95.54
6	70,764	11,794	2.91	98.46
7 and +	37,506	5,009	1.54	100
Total	2,428,282	1,207,698	100	100

Table A1.5: Families with Two Children

	Secondary School Age: 11-17			Upper-Secondary School Age: 15-17		
	Boy and Girl	Boys	Girls	Boy and Girl	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
HR	0.0312**	-0.00363	0.0179	0.0509**	-0.00226	0.00650
	(0.0122)	(0.0154)	(0.0133)	(0.0216)	(0.0321)	(0.0237)
Male	-0.0250***			-0.0391***		
	(0.00270)			(0.00617)		
HR*Male	-0.0501***			-0.109***		
	(0.0149)			(0.0291)		
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls B	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Controls	Yes	Yes	Yes	Yes	Yes	Yes
Observations	196,694	103,330	92,472	76,684	40,375	36,022
R-Squared	0.213	0.234	0.195	0.210	0.238	0.203

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain year fixed effects, municipality fixed effects, and a linear regional trend. Individual controls A include: age, ownership of the home, type of toilet in the household, type of electricity, metropolitan location, and education of the head of the household. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, education of the mother, a dummy variable for missing information on the mother, child's relationship with the head of the household, employment status of the household head, and a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, missing information on income tax, and unemployment rate. Data source: IPUMS 1985, 1993, and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table A1.6:Municipalities with No Coca Production

	Secondary School Age: 11-17			Upper Secondary School Age: 15-17		
	(1)	(2)	(3)	(4)	(5)	(6)
HR	0.0688***	0.0347**	0.0157	0.0896***	0.0588***	0.0413**
	(0.0110)	(0.0140)	(0.0168)	(0.0161)	(0.0173)	(0.0192)
Male	-0.0377***	-0.0319***	-0.0361***	-0.0533***	-0.0432***	-0.0526***
	(0.00314)	(0.00223)	(0.00227)	(0.00433)	(0.00383)	(0.00370)
HR*Male	-0.0320	-0.0446**	-0.0402**	-0.0851**	-0.105***	-0.0971***
	(0.0196)	(0.0178)	(0.0171)	(0.0360)	(0.0348)	(0.0330)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1,034,195	1,034,195	1,034,195	414,150	414,150	414,150
R-Squared	0.076	0.197	0.219	0.094	0.189	0.221

Notes: The dependent variable is enrolment in different school ages (dummy equals 1). Columns represent OLS coefficients. HR is defined as the average between the male homicide rate minus female homicide rate in t-1 and t-2. Male is a dummy equal to 1 if the child is a boy. All regressions contain year fixed effects, municipality fixed effects and a linear regional trend. Individual controls A include: age, ownership of the household, type of toilet in the household, type of electricity, metropolitan location and education of the head of the household. Individual controls B include: rural or urban location, migration status, missing information on migration, education of the mother, missing information on the mother, relation with household head, employment status of the household head, missing information on employment. Municipality controls include: population density, municipality income tax per capita, missing information on income tax and unemployment rate. Data source: IPUMS 1985, 1993 and 2003. This sample is restricted to municipalities with no coca production. Number of municipalities: 354. Standard errors clustered at municipality level reported in brackets. Significance levels *** p<0.01, ** p<0.05, * p<0.1

Table A1.7: Selective Migration of Boys

Dependent Variable: Boys/Girls Ratio	Age: 11–14			Age: 15–17		
	(1)	(2)	(3)	(4)	(5)	(6)
Diff. Homicide Rate	0.0111	0.0349	0.0430	0.143	0.0816	0.119
	(0.105)	(0.114)	(0.111)	(0.0873)	(0.0944)	(0.0966)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	No	Yes	Yes	No	Yes	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1545	1545	1545	1545	1545	1545
R-Squared	0.443	0.454	0.458	0.519	0.541	0.544

Cluster standard errors in parentheses (clustered at Municipality-level)
*** p<0.01, ** p<0.05, * p<0.1
No. of Municipalities: 515

Table A1.8: Heterogeneous Effect of Violence by Level of Education of the Mother

	Secondary School Age: 11–17			Upper-Secondary School Age: 15–17		
	(1)	(2)	(3)	(4)	(5)	(6)
HR	0.0805***	0.0671***	0.0470***	0.100***	0.0941***	0.0793***
	(0.0167)	(0.0150)	(0.0167)	(0.0300)	(0.0279)	(0.0299)
Male	-0.0268***	-0.0249***	-0.0248***	-0.0449***	-0.0425***	-0.0422***
	(0.00262)	(0.00224)	(0.00211)	(0.00461)	(0.00449)	(0.00427)
Less than Primary_Mother	-0.161***	-0.117***	-0.109***	-0.212***	-0.168***	-0.157***
	(0.00965)	(0.00601)	(0.00512)	(0.0113)	(0.00732)	(0.00622)
HR*Male	-0.0242**	-0.0294***	-0.0300***	-0.0579**	-0.0616**	-0.0635***
	(0.0120)	(0.0113)	(0.0111)	(0.0248)	(0.0250)	(0.0240)
HR*Less than Primary_Mother	0.0180	-0.0375	-0.0337	0.0238	-0.0485	-0.0443
	(0.0316)	(0.0262)	(0.0245)	(0.0519)	(0.0439)	(0.0439)
Male*Less than Primary_Mother	-0.0301***	-0.0231***	-0.0222***	-0.0342***	-0.0248***	-0.0231***
	(0.00400)	(0.00406)	(0.00394)	(0.00660)	(0.00688)	(0.00674)
HR*Male*Less than Primary_Mother	-0.0535**	-0.0583**	-0.0593***	-0.0762	-0.0881*	-0.0888*
	(0.0262)	(0.0234)	(0.0227)	(0.0475)	(0.0462)	(0.0455)
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Linear Regional Trend	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls A	No	Yes	Yes	No	Yes	Yes
Individual Controls B	No	No	Yes	No	No	Yes
Municipality Controls	No	No	Yes	No	No	Yes
Observations	1,172,883	1,172,883	1,172,883	454,193	454,193	454,193
R-Squared	0.129	0.216	0.223	0.167	0.214	0.224

Notes: The dependent variable is school enrolment of children of different school ages (dummy equals 1). The columns present OLS coefficients. HR is defined as the average between the male homicide rate minus the female homicide rate in *t*-1 and *t*-2. Male is a dummy equal to 1 if the child is a boy. Less than primary_mother indicates if the mother has less than primary education. All regressions contain year fixed effects, municipality fixed effects, and a linear regional trend. Individual controls A include age, ownership of the home, type of toilet in the household, type of electricity, and metropolitan location. Individual controls B include rural or urban location, migration status, a dummy variable for missing information on migration, child's relationship with the head of the household, employment status of the household head, a dummy variable for missing information on employment. Municipality controls include population density, municipality income tax per capita, a dummy variable for missing information on income tax, and unemployment rate. Data source: IPUMS 1985, 1993 and 2005. Number of municipalities: 515. Standard errors clustered at municipality level are reported in brackets. Significance levels: *** p<0.01, ** p<0.05, * p<0.1

Appendix B1

B1.1 A simple model : The Returns to Education versus the Returns to Violence

I use a simple model of a family's decision on whether to enrol children in school. The model has two periods. School payoffs (r_s) are obtained in the second period, whereas the payoffs of violence ($r_v(v)$) are obtained in both periods. Payoffs in the second period are discounted by a discount factor $\beta(v)$. This discount factor represents the life expectancy of the child as well as how impatient the family is. Both the payoffs and the discount factor depend on the level of violence v. The family makes the decision to enrol the child in secondary school by comparing its expected gains from school and from violence. I assume violence can be measured on a continuous scale with $v \in [0,1]$.

The family utility function (U) is then

$$U = \begin{cases} \beta(v)r_s & if \quad S = 1 \\ (1+\beta(v))r_v(v) & if \quad S = 0 \end{cases}$$

where S represents the family decision of whether to enrol the child in school. S is equal to 1 if he/she decides to go and 0 otherwise. In this model, the family chooses S to maximize its expected utility. These assessments of utility in each period lead to two possible cases:

A. The family will decide to send the child to school ($S = 1$) if,

$$\beta(v)r_s \geq (1+\beta(v))r_v(v)$$

B. The child will decide not to send the child to school ($S_c = 0$) if,

$$\beta(v)r_s < (1+\beta(v))r_v(v)$$

I make 4 parameter assumptions which allow me to show a first proposition:

A1. The discount factor $\beta(v)$ is a decreasing function of violence in the municipality. Therefore,

$$\beta(v) \in (0,1) : \frac{\partial\beta(v)}{\partial v} < 0.$$

This implies that when there is an increase in violence, there is a decrease in the life expectancy and perceived safety of the children.

A2. When violence is lowest, the expected returns to education are higher than the returns to violence ($\beta(0)r_s > (1+\beta(0))r_v(0)$). This captures the fact that there are little opportunities to get involved in violence-related activities when violence is low.

A3. When violence is highest, the expected returns to education are lower than the returns to violence ($\beta(1)r_s < (1+\beta(1))r_v(1)$). In places where violence is very high and the payoffs of violence are also high, children will decide to enter into violence-related activities since they can get returns immediately.

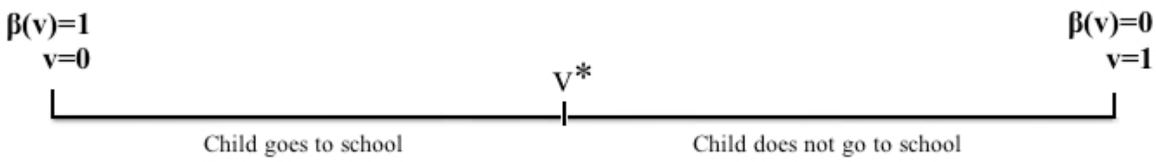
A4. $\beta(v)r_v(v)$ is continuous and strictly increasing in v. The payoffs for getting involved in violence-related activities increases faster with violence than the reduction in life expectancy decreases.

Proposition 1 There exists a unique $v^* \in (0,1)$ such that $\beta(v^*)r_s = (1 + \beta(v^*))r_v(v^*)$. If $v \leq v^*$, the child goes to school. On the contrary if $v^* > v$, the child will go to violence.

Proof: Follows directly from A1-A4.

This result is intuitive. When violence is very low ($v \to 0$), the returns to violence are lower than the returns to education, and the child goes to school. When violence increases, the returns to violence increase, until they equal the returns to education at $v = v^*$. Therefore, for every $v > v^*$, the child will not enrol in school and will get involved in violence-related activities instead. Diagram 1 summarizes graphically the family decision as a function of violence.

Diagram 1: Schooling Decision Based on Violence



As violence increases, the payoffs of violence also increase but the life expectancy of the child decreases. As a consequence, the child will decide to go to school up to the point where the returns to school are higher or equal to the returns to violence.

B1.2 Differences across Genders

Because the type of violence considered in this analysis affects mainly men, there will be differences in the returns to violence, and life expectancy and perceived safety between genders. In particular, I assume:

A5. $\frac{r^v_{boy}(v)}{r^s_{boy}} > \frac{r^v_{girl}(v)}{r^s_{girl}} \,\forall\, v.$

A6. $\beta_{boy}(v) < \beta_{girl}(v) \,\forall\, v.$

Assumption A5 refers to the relative return to violence for boys being larger than the relative return to violence for girls. Violence is a more male activity, and the likelihood of women getting involved in these activities is low. The net returns to violence for women are low relative to their returns to school. Nevertheless, I assume as well that the absolute returns to education are lower for girls than for boys.[37]

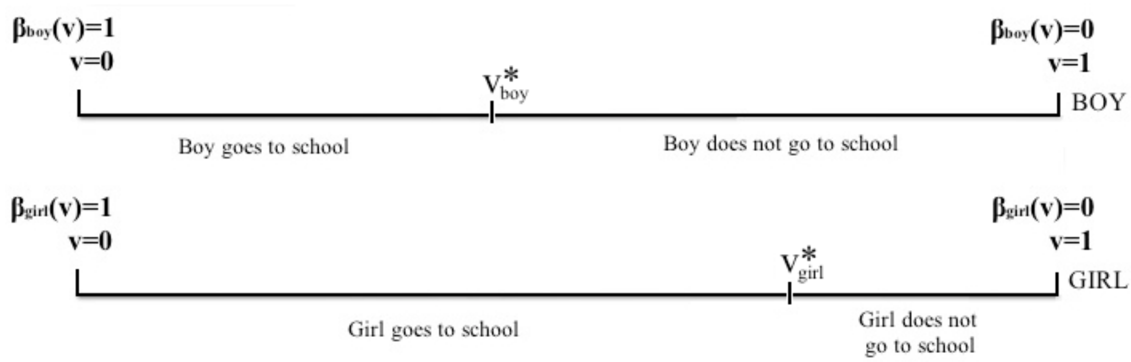
Assumption A6 deals with the life expectancy (the perception of danger) being lower for boys relative to girls. Life expectancy and perceived safety for men is lower than for women since most homicide victims are young men aged 18 to 35 (around 80 percent).

Proposition 2

Boys will always get first involved in violence-related activities in comparison to girls. Under assumptions A5 and A6, I get that $v^*_{boy} < v^*_{girl}$. Diagram 2 illustrates the violence thresholds where boys and girls, respectively, will abandon school. Boys will go to school up to point v^*_{boy}, and girls will enrol in school up to point v^*_{girl}. Since point v^*_{girl} is higher than point v^*_{boy} (with respect to violence), girls are more inclined to go to school relative to boys.

[37] In a study about education in Colombia, Sohnesen and Blom (2006) find that the returns to education for men are higher than for women. The gender difference affects returns to secondary education the least (less than 1 percentage point) and primary education the most (4 percentage points), and the gender difference for tertiary education is around 3 percentage points. The gender difference is established by using different measures: higher hourly wage, more working hours, and a lower unemployment ratio for men.

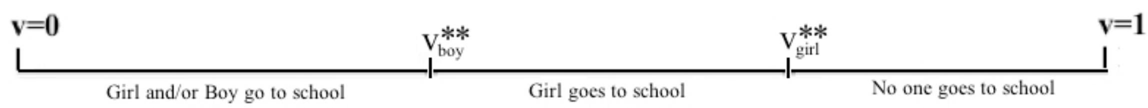
Diagram 2: Differences between Boys and Girls



B1.2.1 Families with Two Children

Now, I look at families with two children: a boy and a girl. In the case that resources are limited and families have to decide whom to educate, they will compare their children's future returns when deciding who will be enrolled in school. Diagram 3 illustrates these possible choices as a function of the level of violence. When violence is very low, the returns to violence are low as well, and the life expectancy (or the perception of safety) is high. In this case, a family will be neutral (with respect to violence) on whether to send the girl versus the boy to school. In principal since the returns to education are higher for the boy than for the girl, without violence one will expect the boy to enrol in school. When violence increases, the opportunities in violence-related activities for boys increase and the life expectancy/perceived safety decreases, there will be a threshold level (v^{**}_{boy}) at which the returns to violence are so high the family prefers to enrol the girl in school and not the boy. Finally, when violence passes an even higher threshold (v^{**}_{girl}), families do not send any of their children to school.

Diagram 3: Families with Two Children



Chapter 2

The Effect of Violence on Fertility Decisions and the Sex Composition Preference of the Offspring

Joint with Adriana Camacho, Universidad de los Andes

2.1 Introduction

The demographic transition theory suggest that as a country develops, its population pattern adjust (e.g., Becker et al., 1990; Galor and Weil, 2000). As a result of improvements in health coverage and technology, the child mortality rate declines, which in turn leads to a decrease in fertility as a consequence of longer life expectancies (Soares, 2005). The demographic changes in Latin America present an interesting case. On the one hand, this region has faced one of the most rapid demographic transitions in the world in the recent years. The fertility rate decreased from almost 5.98 in 1960 to 2.37 children per woman six decades later (CEPAL, 2014). This same reduction took twice as long in the United States and Europe. On the other hand, the region has the highest youth homicide rate in the world, with around 90 percent of these victims are men (United Nations, 2011). This reduces the life expectancy of males and may induce additional demographic changes.

Our paper explores the role of male-biased violence as a factor that affects women's fertility choices. Although violence has been identified as a deterrent for economic growth, more research is needed before conclusions can be drawn about the specific effects that violence has on individual and household welfare (Blattman and Miguel, 2010). We contribute to the literature on the costs of violence by analyzing the related demographic effects via changes in fertility decisions and gender preferences.

Specifically, our study uses the case of Colombia to investigate the effect of male-biased violence on the actual and desired number of children, and the ideal gender composition. This country has suffered from a very long internal conflict and high levels of violence, with variation in intensity and geographic location across time. Furthermore, the economic costs of conflict in Colombia have been enormous, not only in terms of economic growth but also in the death and victimization of civilians (Arias et al., 2014). According to the Centro de Memoria Histórica (2013), in the period between 1985 and 2013, 220,000 people were killed in the conflict, of which 81.5 percent were civilians and 18.5 percent combatants. This implies that one in three deaths in Colombia is caused by violence. Youth in Colombia are directly affected by crime with young males being the main victims and perpetrators of violent crimes according to the United Nations Refugee Unit (2002). Colombian men between the ages of 15 and 35 are 15 times more likely to be homicide victims than women in the same cohort (Velez, 2002). Figure 2.1 compares sex ratios calculated from census data for Colombia in 1993 and 2005 and for other countries in a similar stage of development (Brazil, Ecuador, Mexico, and Venezuela). Colombia shows an atypical sex ratios, particularly in 1993, which coincides with the main peak of violence over the last two decades. The marked dip in the cohorts of 15 to 24 year olds is evidence of males missing from the population. The disparity in sex ratios provides evidence of disproportionate deaths among males in cohorts that are more exposed to violence and war. We further find that the unbalanced sex ratios might are related to the level of violence (see Figure 2.2). Looking at 1993, as violence increases, there is an increase in the shortage of men relative to the women among the youth cohort.

Using high-quality data on violence by municipality and women's reproductive history, we compare the fertility choices of women who have been exposed to different levels of male-biased violence over different periods within their reproductive years. The identification strategy exploits the temporal and spatial variation of homicides rates between 1979 and 2009 in the municipalities of Colombia, and it relies on different geographic and time-fixed effects, along with department-specific trends. We argue that once we partial out municipality- and year-specific variation, violence is no longer correlated with other unobserved determinants of the fertility decisions of women.

Following Soares (2005), we consider that the utility women derive from having children depends on the number of children, the child mortality rate, and the life expectancy and quality of life that each child will enjoy as an adult. In turn, how mortality factors into the value that parents place on each child has important implications for fertility decisions. As a result, the net effect of violence on reproduction choices is ambiguous. On the one hand, the risk insurance approach suggests that children may be seen as a means of insuring income (Pörtner, 2001; Schultz, 1997). Furthermore, there may be an insurance or hoarding effect in response to perceptions of high mortality risk of children, leading to a preventive substitution behavior (Montgomery and Cohen, 1998; Atella and Rosati, 2000). As a result, one would expect an increase in fertility if women demand a higher number of children due to a higher probability of losing them. On the other hand, a decrease in fertility could also be expected. First, women may perceive a disutility from losing their children. Second, women may delay births in response to uncertainty about the future. Prior research shows that violence not only inflicts a direct cost in society through mortality but also brings indirect costs through negative effects on economic activity and, consequently, in individual welfare (e.g., Hoeffler and Reynal-Querol, 2003; Collier and Hoeffler, 2004; Camacho and Rodriguez, 2013). Finally, fertility may also be negatively affected by the shortage of men in places where violence is high (Greene and Rao, 1995; La Mattina, 2014). Jones and Ferguson (2006) show the effect of the shortage of men on the marriage market in Colombia, which makes finding a stable partner and having children more difficult for women.

In addition to the changes in the demand for children, male-biased violence can also affect the gender preferences for children. Such a preference can lead to discriminatory practices that can affect the health and well-being of children. For instance, pronounced gender preferences can lead to sex-selective abortions (e.g., Das Gupta et al., 2003) and, hence, to skewed sex ratios at birth (Goodkind, 1996). Parental gender bias can also negatively affect the provision of basic care, such as breastfeeding, childcare, nutrition, immunizations, medical treatment, and schooling (e.g., Gupta, 1987; Jensen, 2005; Jayachandran and Kuziemko; 2009; Wang, 2005, and Barcellos et al., 2010). Extreme cases of differential investments in children based on gender have also led to excess gender-specific infant/child mortality (Pande 2003, Oster 2009).

To investigate the impact that violence has on gender preferences for children, we study its effect on the desired gender composition of children. The assumption is that women have an ideal gender composition, that is likely to translate into prenatal gender discrimination, differential fertility behavior based on the gender composition of surviving children, and/or postnatal discriminatory practices.

From a theoretical perspective, it is unclear how male-biased violence will affect child gender preferences among women. On one side, women who want to reach a certain sex composition might demand more boys due to a higher probability that they would die in violent environments. Furthermore, women that live in patriarchal environment, sons may be desired for the continuity of the family line and for future economic returns. In Colombia, men have higher returns in the labor market than women, and due to violence, they have new job opportunities in the criminal market.[38]

On the other side, families might prefer girls if they perceive a disutility from losing their children. Since most of the victims in Colombia are young men, girls might be seen as a less risky long-term investment. Women might regard violence as a proxy changes in life expectancy and changes in perceived safety, which in turn would affect the expected income returns. Since males' probability of dying is higher, especially when they are young, there would be a reduction in the future income; hence, sons may be seen as a riskier asset than daughters. As result of these competing possibilities, the effect of violence on gender preferences is an empirical question.

When comparing women of the same age exposed to different levels of violence, we find that those who experienced higher levels of violence, particularly during their most fertile years, have and want fewer children. Furthermore, we show that women exposed to higher levels of violence have a lower preference for boys. These results are in line with those of Gerardino (2013), who finds that households in Colombia that experience higher levels of violence over time decide to invest more in the education of girls relative to boys. Gerardino attributes this result to two main factors. First, in areas with high male-biased violence, investing in female education brings less risk than investing in male education. Second, criminal activities increase the opportunity cost of going to school for males relative to females.

Other papers in the literature have investigated the effects of wars on fertility decisions. Vandenbroucke (2012) finds that birth rates in European countries such as France, Germany, the U.K., Belgium and Italy fell by almost 50 percent during the First World War. Our paper complements Vandernbrouchke´s channel of fertility reduction, as both papers find that the change in demographics comes from deaths during war as well as a decrease in births. Papers by Woldemicael (2008) and Schindler and Brück (2011) look at the decline in fertility for the postconflict period

[38]In a study about education in Colombia, Sohnesen and Blom (2006) find that the returns to education for men are higher than for women. The gender difference affects returns to secondary education the least (less than 1 percentage point) and primary education the most (4 percentage points), and the gender difference for tertiary education is around 3 percentage points. The gender difference is established by using different measures: higher hourly wage, more working hours, and a lower unemployment rate for men.

at the end of the 1990s in Eritrea and the genocide in Rwanda, respectively. Similar to our paper, both of these studies use the DHS survey for their respective countries to create a pooled cross-section to examine changes in fertility. Woldemicael finds a remarkable drop in fertility, suggesting that marriage delay partly explains the decline of births. Schindler and Bruck find that the genocide affected fertility in the short term, with a strong replacement effect for lost children. In the long-term, fertility declined due to the large group of widows in Rwanda, demonstrating the importance of the institution of marriage in determining fertility. The main contribution of our paper to the existing literature is that, to the best of our knowledge, this is the is the first study to look at the relationship between reproductive choices and the level of exposure to violence over time during an ongoing conflict in which the main victims are young men. Additionally, we are the first to study the impact of violence on desired fertility and the desired gender composition of children.

This paper is divided in seven sections, including this introduction. In Sections 2.2 and 2.3, we describe the historical context and the data, respectively. Section 2.4 explains the empirical methodology, and Section 2.5 presents the baseline results. Section 2.6 describes the results of additional estimations. Finally, section 2.7 concludes.

2.2 Historical Context

2.2.1 Violence in Colombia

High levels of violence have become an important social concern, not only because they undermine the social and economic climate of an area, but also because they weaken economic incentives and conditions that foster development (Collier et al., 2003). This dramatic effect becomes more evident in a country like Colombia, which has suffered one of the longest ongoing domestic conflict in the world.[39] Over the last two decades, the conflict has grown both in intensity and size, becoming a national issue.

Violence in Colombia changed from remote guerilla activity, originated in the early 1960s, to a country in war involving multiple paramilitary groups, drug cartels, and two guerrilla groups. In the early, Colombia's 1970s national homicide rates were similar to those of its neighbors. Thereafter, homicides increased dramatically, and by 1991 the rate had grown more than threefold. During the past four decades, the

[39]This section is mainly drawn from Echandia (2006).

scale and intensity of violence changed from a marginal conflict to a national issue.

The rise of different illegal armed groups spurred this intensified conflict and led to increased levels of crime. The causes of Colombia's crime problem are undoubtedly complex. All these violent actors, in addition to engaging in direct fighting with the army, also perpetrate crimes and terrorist attacks against the civil population and public infrastructure. In addition they have used crime such as kidnapping and extortion, drug production and trafficking as sources of financing.

Colombia has been the largest cocaine exporter in the world since the 1980s, fueling the growth and strengthening of the Medellín and Cali cartels. Before 1993, Peru and Bolivia were the main producers of coca leaf, and Colombia was involved in the production and trac of cocaine. After 1993, the drug industry began to change in response to new antinarcotics policies in the coca leaf producer countries that made transportation between countries very costly. Thereafter, coca leaf cultivation and drug production, in addition to tracking, became one of the main drivers of the intensification and expansion of violent conflict in Colombia. Furthermore, aid for fighting cartels increased, causing the Medellín and Cali cartels to be dismantled, accompanied by high levels of violence. At the peak in 1991, nearly one in every 100,000 Colombians was murdered each year (see Figure 2.3). Once the two main cartels disappeared, there were several violent clashes within different cartels and against the guerrillas and paramilitaries that wanted to maintain control over this profitable business. Thus, guerrilla groups and the paramilitaries increased their participation in the illegal drug trade. These additional resources allowed them to expand their military capacity, intensifying the Colombian conflict from the mid-1990s onward.

The history of violence in Colombia over the last sixty years can be divided into four main blocks. A first period of low and stable violence occurred between 1965 and 1981. Then from 1982 to 1995, the country experienced a growing trend of violence due to the expansion of the guerrillas, the emergence of paramilitary groups, and the strengthening of drug trafficking. In the following years (1996-2002), violence remained high due to the military strengthening of the guerrillas, the geographical expansion of paramilitary groups, and the link between drug trafficking and armed groups. From 2003 until today, violence has decreased with the consolidation of the military presence in the country, the retreat of the guerrillas, and the demobilization of paramilitary groups (Arias et al., 2014).

2.3 Data

This study uses two sources of information. The first is the individual and household data retrieved from the Demographic and Health Survey (DHS), which contains information on fertility decisions and sociodemographic characteristics. The second source corresponds to homicides reports in the vital statistics records from Departamento Administrativo Nacional de Estadística (DANE). We merge this data with the DHS data by municipality and year of occurrence.

2.3.1 Individual and Household Data

We use the Demographic Health Survey (DHS) for Colombia for the years 1995, 2000, 2005, and 2010. The survey includes detailed information on female reproductive health and the history of births for women between ages 13 and 49 for the years 1995 and 2000, and women between 15 and 49 years old for the years 2005 and 2010. Because we want to track the violence a woman experiences during her fertile years and the available data spans from 1979 to 2010, we restrict the analysis to women born later than 1964 and that were at least 15 years old at the time of the survey. Once we pool the information for the available years, the data set consists of approximately 82,012 observations. These women live in 355 municipalities located in 28 departments in Colombia, which account for 32 percent of the total number of municipalities (1,123) and 87 percent of the total number of departments (32).

We present descriptive statistics in Table 2.1, including information on the sample of women and their partners' sociodemographic characteristics, respectively. Women in our sample are on average 26.1 years old; 13.4 and 12.6 percent of them have incomplete and complete primary education, respectively. The greatest proportion of women (30.7 percent) have incomplete secondary education, followed by women with complete secondary (23.9 percent) and higher education (19.2 percent). In our sample, 37.5 percent of the women have never been married, and 50 percent are either married or cohabiting. An additional 12 percent of females are widows or divorced. For the partners education, 20 percent have no education or less than primary, 18 percent have completed primary education, around 47 have completed some or all secondary education, and 14 percent have received some tertiary education. On average, the partners are 34 years old. Of the females in the sample, 11.4 percent are the head of the household, 37.5 are spouses of the head, 39.2 are daughters or daughters-in-law. Approximately three-fourths of our sample live in a urban areas.

Table 2.2 displays the summary statistics for the variables that describe the fertility choices of women. The number of children born in the sample is 1.43, and the number of children alive is 1.39. On average, these women have 0.70 girls and 0.73 boys. The women in sample desire, on average, 2.16 children, measured by the ideal number of children reported in the survey. Women were asked "If you could go back to the time [when] you did not have any children and could choose exactly the number of children to have in your whole life, how children would you have?". This number is higher than the number of children born perhaps because the women in our sample had not yet completed their years of fertility. In terms of gender preferences, women prefer to have more girls than boys: 1.12 versus 1.03, respectively. On average, around 18.6 percent of the women desire to have more girls than boys, 12.29 wish to have more boys than girls, and 69.10 are indifferent to desired child gender or want the same number of boys and girls.

2.3.2 Violence Data

We measure violence via municipal homicide rates per 100.000 inhabitants. The number of homicides by municipality is taken from vital statistics records from DANE for the years 1979 to 2009. This information on violent deaths is retrieved from the forensics office in each municipality and includes all deaths caused by a third party (homicide and guerrilla or governmental actions). Furthermore, it includes characteristics of the victims such as gender, educational level, and marital status. Table 2.3 summarizes the descriptive statistics of homicides during this period. Between the years 1979 and 2009, 628,103 homicides were committed. Most of the victims were male (around 92 percent) between the ages of 18 and 35 (61.28 percent). Hence, the violence in Colombia during this period, using homicides, was male-biased. Not every death record is complete, but among the subsample that has full information, 52 percent of victims were single, 36 percent were married or cohabitating. Data on the education level of the victims is only available for 2000 onwards. Around 5 percent of the victims had at most a preschool education; 57 percent had some primary school education (either incomplete or complete, 29 and 28 percent, respectively), and 34 percent had attended secondary school. Around 4 percent of the victims went to university.

To generate the main variable of interest—the homicide rate by municipality—we match the number of homicides with the population projections for each municipality each year.[40] We then combine the homicide data with the DHS data for a sample

[40]Censuses in Colombia were conducted in 1973, 1985, 1993, and 2005. Also this information

of years and Colombian municipalities based on where the survey was carried out.

The way we define periods of violence is very important for this analysis. First, we turn to the literature that examines the way individuals make decisions based on their memories, because we are not only going to use the number of children, but also the desired number of children. We use different definitions to identify the effect of violence on fertility decisions. First, to correctly establish the period of violence, we refer to the economics and psychology literature. The former heavily relies on the premise that individuals discount the present time much more than the distant future, a process known as hyperbolic discounting. Rational individuals will calculate the present discounted value of their investments, taking into account certain risk (expected utility). The prospect theory, developed by Daniel Kahneman and Amos Tversky (1979), incorporates loss aversion into the expected utility framework, which implies that people might react strongly to losses. The psychology literature examines how individuals make decisions based on their memories; evidence indicates that individuals are able to suppress bad memories more easily than good memories, which is known as the fading effect bias. In addition, psychologists have identified other types of bias, such as "duration neglect bias", which that shows that humans value painful situations based on the intensity and duration of their suffering.

We defined three different periods of violence considering important moments in the fertile cycle of a women. The first period, exposure to violence from age 16 to 24, corresponds to the interval from minus one standard deviation to plus one standard deviation around the mean sample age of women at the birth of their first child. The second period we look at is exposure to violence between the ages of 20 to 30: the former is the mean age when women in the sample had their first child, and the latter is the mean age when women in the sample had their oldest child. In this case, we restrict the sample to women older than 20 years of age. Exposure to violence between ages 15 and 36 corresponds to the period from the minimum age (15) to the mean age of women at the birth of their last child plus one standard deviation. We calculate the homicide rate in every municipality for each year between 1979 and 2009 $((\frac{Number\,of\,Homicides_{m,t}}{Population_{m,t}}) * 100)$ and then estimate the average for the different periods of exposure to violence defined above. On average, women were exposed during the three periods to an average rate of 66, 68, and 66 homicides per 100.000 inhabitants, respectively. We see a large variation in the exposure to violence of these women, with standard deviations of the average rate, respectively, being 53,

comes from DANE.

56, and 57 homicides per 100.000 inhabitants.

2.4 Empirical Strategy

2.4.1 Baseline Specification: Number of children and desired number of children

Our empirical strategy exploits the variation in homicide rates by municipality and over time between 1979 and 2009 in Colombia. All women in the sample are in their fertile years, and they have been exposed to different levels of violence. This allows us to capture whether being exposed to different levels of violence has differential effects on fertility decisions. We estimate all regressions using ordinary least squares (OLS). We first empirically test the relationship between violence exposure and the number of children born and desired. We estimate the following reduced form model:

$$Fertility_{i,m,t} = \beta_0 + \gamma_\tau HomicideRate_{m,\tau} + \chi_i + \mu_m + \lambda_y + \varphi_d(t) + \upsilon_t + \varepsilon_{i,m,t}\ , \quad (2.1)$$

where $Fertility_{i,m,t}$ corresponds to the number of children (real or desired) that a woman i in municipality m and in year t reports. The number of children refers to the number of births a woman had up to the time of the survey. The number of desired children is the number of children a woman would like to have if she could start her fertile period over. Our main variable of interest, $HR_{m,\tau}$, is the average homicide rate that each woman experienced during three different stages of her reproductive life (early stage, middle stage, complete fertile stage).[41] $HomicideRate_{m,\tau}$ includes three measures of exposure to violence. We aggregate the estimation of γ_τ into three different spans of ages during which women are exposed to violence $\tau\epsilon\{15–36, 20–30, 16–24\}$ corresponding to when women were ages 15-36, 20–30, and 16–24, respectively. To measure of exposure, we use the average of each municipality's homicide rate during the aforementioned age ranges. This methodology allows us to examine the age ranges during which changes in violence have a larger effect on fertility decisions. χ_i includes age-group fixed effects. We divide all women in 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.); the last group includes ages 42 and 43 only. This control variable is very important because it allows us to control for the effect of age in fertility behavior. μ_m are municipality fixed effects, which control for any specific characteristic of the municipality that might

[41] The homicide rate in each period is equal to $((\frac{Number\ of\ Homicides_{m,t}}{Population_{m,t}}) * 100)$

affect the number of children born or the preferences for a particular family size. Year-age-15 fixed effects (λ_y) incorporate time shocks and control for any cohort-specific characteristic that may have an effect on the fertility decisions, such as the introduction of birth control methods. To account for potential long-run differences in regional development, we include a cubic departmental–specific time trend $\varphi_d(t)$. We also include the year-of-the-survey fixed effects υ_t to account for any potential difference in the surveys across years. Finally, to consider any potential intraclass correlation across individuals living in the same municipality, standard errors $\varepsilon_{i,m,t}$ are clustered at the municipality of residence level.

To identify women's exposure to violence, we consider women of different ages, and we match their municipality of residence with the time series data of homicides by municipality. We compare the number of children born and the desired number of children within women residing in municipalities with varying levels of violence. Because we control for age-group fixed effect, year-age-15 fixed effects, municipality fixed effects, year of the interview fixed effects, and regional trends, we use changes in the number of children and the desired number of children over time within municipalities (beyond those predicted by national trends) to identify the effect of violence. As a result, estimates will measure the average effect of violence on fertility, conditional on age, in relation to municipality averages and year averages.

The main objective of this paper is to determine the impact of changes in violence on fertility decisions. According to equation (2.1), the main focus is the estimates of γ_τ. In accordance with the nature of violence in Colombia, women may regard violence as a proxy for factors such as changes in the life expectancy of children and partners, and changes in perceived safety, which in turn will affect families' economic well-being. The net effect of violence on total fertility is ambiguous. On the one hand, a decrease in fertility may be expected if i) women perceive a disutility from losing their children, ii) women delay births in response to uncertainty about the future, and iii) there is a shortage of men in places where violence is high (Green and Rao, 1995). On the other hand, one would expect an increase in fertility if i) children are seen as means of ensuring income, and ii) there are insurance or hoarding effects stemming from the perceptions of high mortality risk that may manifest as preventive substitution behavior (Montgomery and Cohen, 1998 ; Atella and Rosati, 2000). Parents would demand more children because they perceive a higher probability of losing them.

To consistently estimate the causal effect of violence on fertility decisions, the main identifying assumption is that within municipalities, changes in violence are not correlated with unobserved changes in other determinants of fertility. There is one main threat to this identification strategy. There could be other unobserved characteristics that affect fertility in places where violence is high. If this is the case, there may be time-varying omitted variables that will bias the estimates. For example, if municipalities with higher levels of violence also experience changes in job opportunities for women, we may be wrongly attributing the negative effect of violence on fertility.

Migration is an important concern when estimating the impact of violence on fertility. Since violence may induce migration, positive or negative selection into migration could bias the results. Estimates may be driven by selective migration of women affected by homicide. If women who migrate due to violence frequently have a lower preference for real and desired children, the effect of violence on fertility choices will be underestimated. On the other hand, if the women who migrate are among the ones with higher preferences for children and those most affected by violence, the effect of violence on fertility choices would be overestimated. In order to determine correctly the exposure to violence that a woman has experienced, it is important to consider the time a woman has been living in the current location of residence. The DHS data do not include individuals' entire migration histories. We only have information on how many years the women have been living in the current municipality at the time of the survey. As a result, we cannot precisely assign the level of violence a woman that migrated during his fertile year has experienced. Consequently, we restrict the sample to women who have not migrated or those who have been living in the municipality since they were 15 years old. We compare the sample of migrant women with the original sample of women.

An additional concern is the effect that violence may have on women conditional on age. Women might decide to postpone or anticipate having children, depending on their age when they are exposed to violence. On the one hand, women may want to wait until it is safer to have children. Biologically, women's fertility begins on average at age 12 and peaks in their early 20s. Fertility begins to decline in their late 20s and drops considerably after age 35 (Dunson et al, 2002). Thus, younger women can postpose having children longer than women close to the end of their reproductive cycle. On the other hand, women may want to have children sooner when fewer men are available. To deal with this issue, we interact the age group dummies with the different measures of violence and estimate equation (2.2):

$$Fertility_{i,m,t} = \beta_0 + \gamma_\tau HomicideRate_{m,\tau} + \varrho_i AgeGroup_i + \delta_\tau (HomicideRate_{m,\tau} * AgeGroup_i) + \mu_m + \lambda_y + \varphi_d(t) + \upsilon_t + \varepsilon_{i,m,t}\,, \quad (2.2)$$

where $HomicideRate_{m,\tau} * AgeGroup_{i,m}$ is the interaction between the average homicide exposure and a dummy variable for each age group.[42]The other control variables are the same as in equation (2.1). Using this approach, we can compare women in the same age group that were exposed to different levels of violence. Our focus is on the impact of high levels of violence and age, which is captured by δ_τ.

2.5 Baseline Results

Table 2.4 displays the results for equation (2.1), which measures the effect of violence on the number of children born. The top panel shows the estimates for the entire sample, and the second panel includes the non-migrant sample. We use the non-migrant sample in all specifications because we want to rule out the potential effect of selective migration and measurement error on the assignment of the violence in the time periods examined. We use the three different measures of life exposure to violence and present three different sets of regression for each of these measures. Controls are included sequentially on each regression. Columns 1, 4, and 7 include year, municipality, and age-group fixed effects. Columns 2, 5, and 8 add cohort fixed effects and sociodemographic control, and Columns 3, 5, and 9 further include a departmental cubic time trend. Overall, for all measures of exposure, the estimates of the main variable of interest (γ_τ)—the average homicide rate for different age groups during the fertile period of a woman's life—are very stable to the inclusion of all the controls variables. Results strongly suggest that being exposed to violence, particularly when women are young, influences reproduction decisions.

We find that exposure to violence between the ages of 16 and 24, 20 and 30, and 15 and 36 has a negative and significant effect of 1.091, 1.190 and 0.917 over the number of children born, respectively (Columns 3, 6, and 9). These findings imply that an increase of one standard deviation in the average homicide rate reduces the number of children born by 0.06, 0.07 and 0.052, respectively. .[43] This effect

[42]We divide all women in 10 brackets, each including three ages (e.g. 15-17, 18-20, 21-23, etc); the last group includes ages 42 and 43 only.

[43]The standard deviations of the violence exposure to violence depending on age are 0.0534,

represents a 5 percent reduction in the number of children born, from the average number of children born (1.4356) to the women in the sample. In a more extreme case, women that live in a municipality in the 90th percentile of the distribution of violence (0.12) will face a reduction of around 8 percent in the number of children relative to women who live in a municipality in the 10th percentile (0.018) of violence. When considering the sample of non-migrants, the effects are larger, which is consistent with our expectation that there might be some mis assignment in the violence perceived by migrant women. For these non-migrant women, an increase of one standard deviation in the measures of violence leads to a reduction in the number of children of 0.065, 0.098 and 0.046, respectively (Panel II, Columns 3, 6, and 9). This represents a reduction of around 4 to 7 percent in the average number of children born. Thus, women living in areas that fall within the 90th percentile of the distribution of violence exposure between the ages of 20 to 30 (Column 6) will, on average, have 0.17 fewer children than women living in areas in the 10th percentile of the distribution. These numbers equate to a 12 percent reduction in the average number of children born.

Results of equation (2.2), which addresses potential heterogenous effects of age on reproduction choices, are displayed in Table 2.5 for the full sample and Table 2.6 for the non-migrants. Although the effect of violence is negative for all age groups, it is not always statistically significant. For instance, there is not a differential effect of violence conditional on age when considering exposure between ages 20 to 30 (Columns 4 to 6). For this level of exposure, the younger girls in the sample are not included since they had not reached the age of exposure. In this case the omitted group is the group of women ages 42 to 43. When looking at the other periods of exposure (ages 15 to 24 and 15 to 36), we are able to include young girls. In this case, we use girls ages 15 to 17 as the omitted group. Results suggest that being exposed to higher rates of violence affect older women the most, implying that the effect increases with age. This evidence supports the idea that when violence is higher teenage pregnancy may increase (Millán, 2013), and young women may want to have children younger due to a shortage of men.

In addition to the number of children that the women have, the DHS survey asks about the desired number of children. Women are asked what would be the ideal number of children they would have today if they could start they fertile life over. We estimate the effect of violence on the ideal (desired) number of children using equation (2.1). Table 2.7 presents the results, employing the same control

0.0564 and 0.0569, respectively.

variables as in Table 2.4. Although violence exposure negatively affects the desired number of children, the effect is smaller than for the number of children born, and it is not always statistically significant. Panel II in Table 2.7 displays the results for the non-migrants sample. Estimates are smaller than for the full sample.

2.6 Gender Preferences for Children

The gender structure of violence is very unbalanced in Latin America. Males account for 90 percent of all homicide victims, compared to 81 percent in Africa and 73 percent in Europe (United Nations, 2011). In Colombia, over our study period, most of the victims were male (around 92 percent) between the ages of 18 to 35 (more than 60 percent). We want to further explore the changes in fertility decisions and how violence affects women's preferences for the gender of their children. The DHS survey asks women about their ideal (desired) number of children and also about the gender composition. Women are asked, "If you could go back to the time [when] you did not have any children and could choose exactly the number of children to have in your whole life, how many girls and boys would that be?"

Table 2.8 presents the results of the effect of violence on the difference between the ideal/desired number of boys minus the ideal/desired number of girls. We estimate equation (2.1) using as a dependent variable the difference between the ideal number of boys and girls per woman. In the case that women are indifferent, we assigned the total number of desired number of children divided by two to each gender. Control variables are the same as in equation (2.1). We find that violence has a negative effect on the difference between genders. The results for the sample of non-migrants is similar although the coefficients are larger (Panel II - Table 2.8).[44] These results may come from either an increase in the preference for girls or a decrease in the preference for boys.

An alternative strategy to study the effect that exposure to violence has on preferences for children is to explore the categorical nature of the answers given by women and estimate a multinomial logit model. For this approach, the dependent variable takes one of three values: (0) if the woman desires more girls; (1) if she does not have a preference for a specific gender; and (2) if the women prefers more boys. To estimate this model, we normalize on one category, which would be the reference group. In this analysis, the first category (preference for boys) is the

[44] Results of estimation 2.2 are shown in the Appendix A2.

reference category. Control variables are the same as in equation (2.1). By using this approach, we are assigning a level of indirect utility to each alternative and assuming that women choose the alternative that yields the greatest utility. The indirect utility is assumed to be a function of women's characteristics as well as other unobserved characteristics. The coefficient estimates of the various levels of exposure to violence represent the differential effects of this variable on the utility of the mother.

Results of the multinomial choice model are displayed in Table 2.9, and the marginal effects are presented in Table 2.10. Women's exposure to higher levels of violence increases the probability of choosing the alternative in which the mother does not has a gender preference (Panel III, Table 2.10). Furthermore, high levels of violence exposure also decreases the probability of having a preference for boys (Panel II, Table 2.10) and has no impact on the preferences for girls (Panel I, Table 2.10). The violence that has the largest impact on the probability of having a gender preference is when women are exposed to violence between the ages of 20 and 30. The inclusion of the cohort-fixed effect does not change the estimates (Columns 2, 5, and 7) whereas the inclusion of the cubic departmental time-trend makes the estimates smaller and less statistically significant (Columns 3, 6, and 9). This may be due to the fact that by including this variable, we might be taking out a lot of the variation of the homicides rates (Columns 3, 6, and 9). The marginal effects are very similar when looking at the sample of non-migrants, although the estimates are larger (Table 2.11).

To check for the robustness of our results to alternative estimation methods, we estimate equation (2.1) using as a dependent variable a dummy variables equals 1 if the woman does not have a preference for a boy or a girl, and 0, otherwise. The multinomial logit model produces estimates of the marginal effect of violence on not having a gender preference that are qualitatively identical to the OLS estimates. Table 2.12 displays the results.

Finally, we look at the effect of violence on the gender composition of actual births. Table 2.13 displays the results of equation (2.1) using the difference between the number of boys and girls born. Similar to the effect found on the desired number of girls and boys, the effect of violence is negative on this difference but not always statistically significant. Estimates are larger than when looking at the ideal number of children (Table 2.8). In Colombia, there is no evidence of selective abortion of boys or girls: the ratio of boys/girls is close to one between ages 0 and 4 (see Figure 2.2), suggesting the potential presence of a gender stopping rule in favor of girls.

Further research is needed to test this hypothesis.

2.7 Conclusions and Future Work

This article adds to the growing literature in economics on the effects of violence on individual and household welfare. We focus on the effect that male-biased violence has on fertility decisions, such as the number of children and the preferences for one gender over the other. These two issues are very important since can help to understand some of the patterns of the current demographics changes. Furthermore, understanding what affects the formation of gender preferences is key since this type of preferences can cause discriminatory practices that may lead to unfavorable social and demographic consequences.

In theory, the impact of male-biased violence on women's fertility decisions and the gender preferences of children is unclear. On the one hand, women might increase the number of children, particularly boys, that they choose to have in response to high mortality rates among males in violent regions. On the other hand, women in such areas might have a preference for girls due to a perceived disutility from losing a child. Because most of the victims of homicide in Colombia are young men, girls might be seen as a less risky long-term investment. Due to these competing potential outcomes, how violence affects fertility decisions is an empirical question.

To answer this question, we turn to Colombia. Using high-quality data on violence by municipality and data on women's reproductive choices, we compare the fertility choices of women that have been exposed to different levels of male-biased violence over different periods in their fertile age. The identification strategy relies on the temporal and spatial variation of homicides rates between 1979 and 2009 in the municipalities of Colombia. When comparing women of the same age exposed to different levels of violence, we find that those who experienced higher levels of violence, particularly during their most fertile years, have and want fewer children. Furthermore, we show that women exposed to higher levels of violence have a lower preference for boys and tend to be more indifferent about the gender of their children. The results of this paper suggest that violence imposes a demographic burden other than the direct loss of life during war. The uncertainty about life expectancy, particularly for males living in a conflict region, changes demographic patterns such as fertility and gender preferences. Understanding how violence and crime affect individual decisions has compelling policy relevance. To address the indirect costs of violence, policymakers must understand the mechanisms through which it affects people's lives, particularly in the developing world.

Further analysis is needed to understand the mechanism through which violence affects fertility decisions. The next step in this project is to explore the effect that violence has on family formation behavior such age of marriage, age difference between spouses, and the number of years between births. Another issue to explore is whether the gender preference translate into differential investment behaviors in terms of health and nutrition for boys versus girls.

References

[1] Arias, M.A., Camacho, A., Ibáñez, A.M., Mejía, D., and Rodríguez, C. (2014). Costos económicos y sociales del conflicto en Colombia: ¿Cómo construir un posconflicto sostenible? Universidad de los Andes, Ediciones Uniandes, Facultad de Economía.

[2] Atella, V., and Rosati, F. (2000). Uncertainty about children's survival and fertility: A Test Using Indian Microdata. Journal of Population Economics, 13(2): 263-278.

[3] Barcellos, S. H., Carvalho, L. S., and Lleras-Muney, A. (2010). Child Gender and Parental Investments in India.

[4] Becker, G. S., Murphy, K. M., and Tamura, R. (1990). Human capital fertility and economic growth. Journal of Political Economy, 98(5): 12-37.

[5] Blattman, C., and Miguel, E. (2010). Civil war. Journal of Economic Literature, 3-57.

[6] Camacho, A. and Rodriguez, C. (2013). Firm exit and armed conflict in Colombia. Journal of Conflict Resolution, 57(1): 89-116.

[7] Centro de Memoria Histórica (2013). Informes 2013. Bogotá D.C., Colombia

[8] Comisión Económica para América Latina (2014). Observatorio demográfico 2013: Proyecciones de población. CELADE - División de Población de la CEPAL.

[9] Collier, P., Elliott, V. L., Hegre, H., Hoeffler, A., Reynal-Querol, M., and Sambanis, N. (2003). Breaking the conflict trap: Civil war and development policy.

[10] Collier, P. and Hoeffler, A. (2004). Greed and grievance in civil war. Oxford Economic Papers, 56(4): 563-595.

[11] Das Gupta, M., Zhenghua, J., Bohua, L., Zhenming, X., Chung, W., and Hwa-Ok, B. (2003). Why is son preference so persistent in East and South Asia? A cross-country study of China, India and the Republic of Korea. The Journal of Development Studies, 40(2): 153-187.

[12] Dunson, D. B., Colombo, B., and Baird, D. D. (2002). Changes with age in the level and duration of fertility in the menstrual cycle. Human Reproduction, 17(5): 1399-1403.

[13] Echandía, C. (2006). Dos décadas de escalamiento del conflict armado en Colombia, 1986–2006. Universidad Externado de Colombia, pp. 309.

[14] Galor, O. and Weil, D. N. (2000). Population, technology, and growth: From malthusian stagnation to the demographic transition and beyond. American Economic Review, 90(4): 806-28.

[15] Gerardino, M. P. (2013). The effect of violence on the educational gender gap. Working Paper, Universitat Pompeu Fabra

[16] Greene, M. E., and Rao, V. (1995). The marriage squeeze and the rise in informal marriage in Brazil. Biodemography and Social Biology, 42(1-2): 65-82.

[17] Goodkind, D. (1996). On substituting sex preference strategies in East Asia: Does prenatal sex selection reduce postnatal discrimination? Population and Development Review, 111-125.

[18] Gupta, M. D. (1987). Selective discrimination against female children in rural Punjab, India. Population and Development Review, 77-100.

[19] Hoeffler, A. and Reynal-Querol, M. (2003). Measuring the costs of conflict. Washington, DC: World Bank.

[20] Jayachandran, S. and Kuziemko, I. (2009). Why do mothers breastfeed girls less than boys: Evidence and implications for child health in India. NBER Working Paper 15041.

[21] Jensen, R. (2005). Equal treatment, unequal outcomes? Generating sex inequality through fertility behavior. Unpublished Manuscript, Harvard University.

[22] Jones, J. H., and Ferguson, B. (2006). The marriage squeeze in Colombia, 1973– 2005: The role of excess male death. Biodemography and Social Biology, 53(3-4): 140-151.

[23] Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica: Journal of the Econometric Society, 47(2): 263-291.

[24] La Mattina, G. (2014). Civil conflict, sex ratio and intimate partner violence in Rwanda. HiCN Working Paper 175.

[25] Millán, J. A. (2013). Drugs, Conflict and Early Motherhood in Colombia. Working paper.

[26] Montgomery, MR., and Cohen B. (1998). From Death to Birth: Mortality Decline and Reproductive Change. Washington DC: National Academy Press.

[27] Oster, E. (2009). Proximate Sources of Population Sex Imbalance in India. Demography, 46(2): 325–339.

[28] Pande, R. (2003). Selective Gender Differences in Childhood Nutrition and Immunization in Rural India: The Role of Siblings. Demography, 40 (3): 395–418.

[29] Pörtner, C. C. (2001). Children as insurance. Journal of Population Economics, 14(1): 119-136.

[30] Schindler, K. and Brück, T. (2011).The effects of conflict on fertility in Rwanda. Policy research working paper WPS5715. Washington, DC: World Bank.

[31] Schultz, T.P. (1997). Demand for children in low income countries. Handbook of Population and Family Economics. Edited by M.R. Rosenzweig and 0. Stark.

[32] Sohnesen, T., and Blom, A. (2006). Is formal lifelong learning a profitable investment for all of life? How age, education level, and flexibility of provision affect rates of return to adult education in Colombia. World Bank Policy Research Working Paper (3800).

[33] Soares, R. 2005. Mortality Reductions, Educational Attainment, and Fertility Choice. American Economic Review, 95(3): 580-601.

[34] United Nations. (2011). Global study on homicide. United Nations Office on Drugs and Crime.

[35] United Nations Refugee Unit (2002). Aprenderas a no llorar: Niños Combatientes en Colombia.

[36] UNDOC (2006). Violencia, crimen y trafico illegal de armas. Oficina de crimen. Oficina contra la droga y el delito. Bogota-Colombia

[37] Vandenbroucke, G. (2012). Optimal fertility during World War I. Working paper series: Munich Personal RePEc Archive.

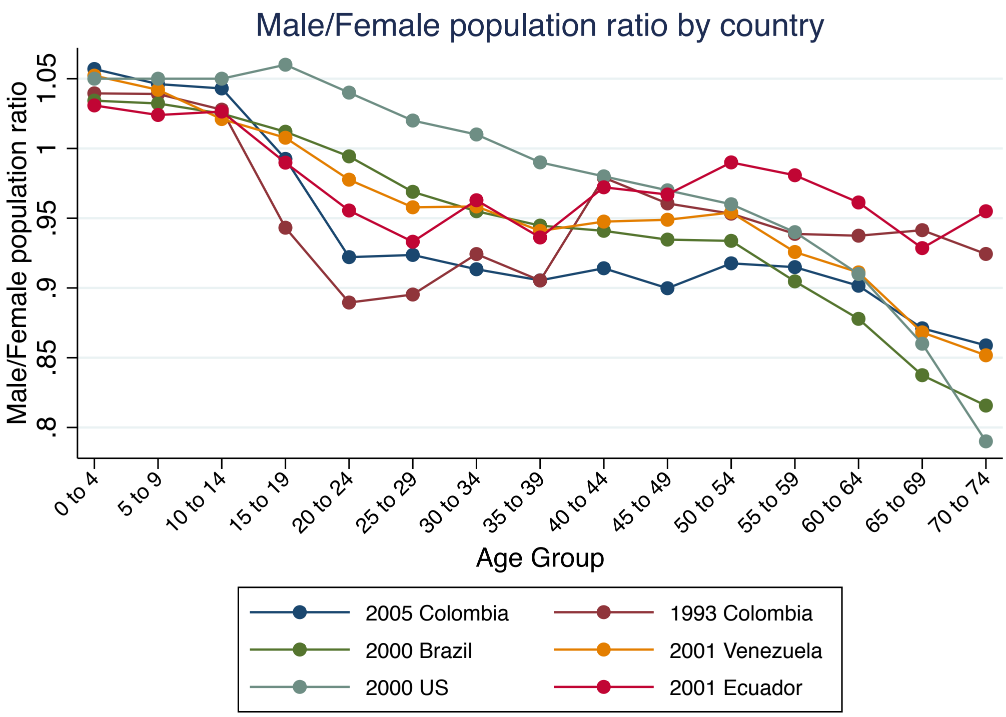
[38] Velez, C.E. (2002). Colombia Poverty Report. Washington, DC. World Bank.

[39] Wang, W. (2005). Son preference and educational opportunities of children in China— "I wish you were a boy!" Gender Issues, 22: 3-30.

[40] Woldemicael, G. (2008). Recent fertility decline in Eritrea: Is it a conflict led transition. Demographic Research, 18(2): 27-58.

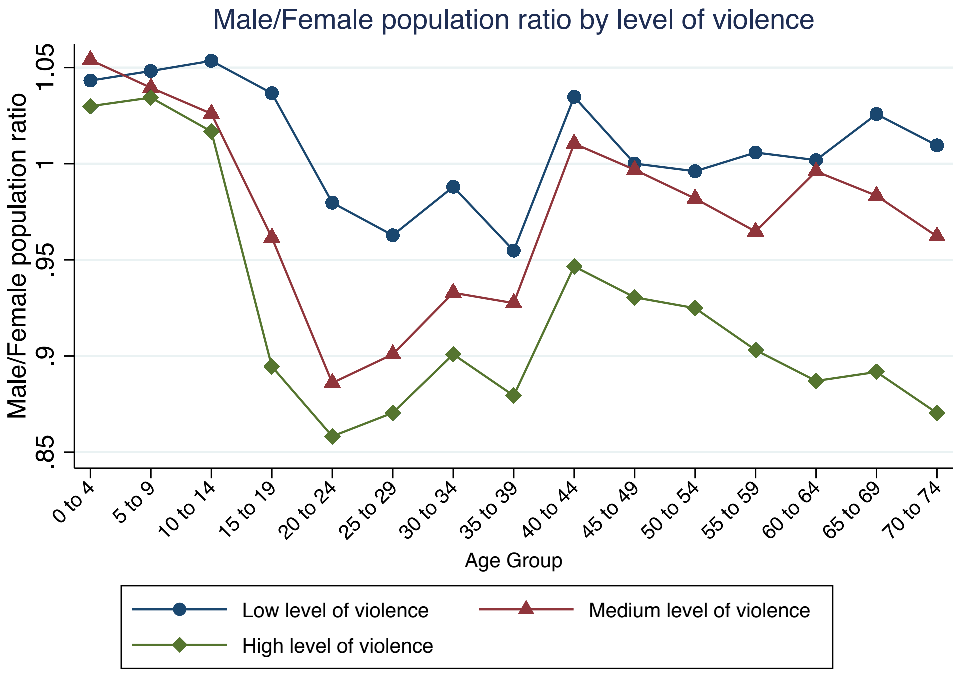
Figures and Tables

Figure 2.1: Sex Ratios Comparison by Country



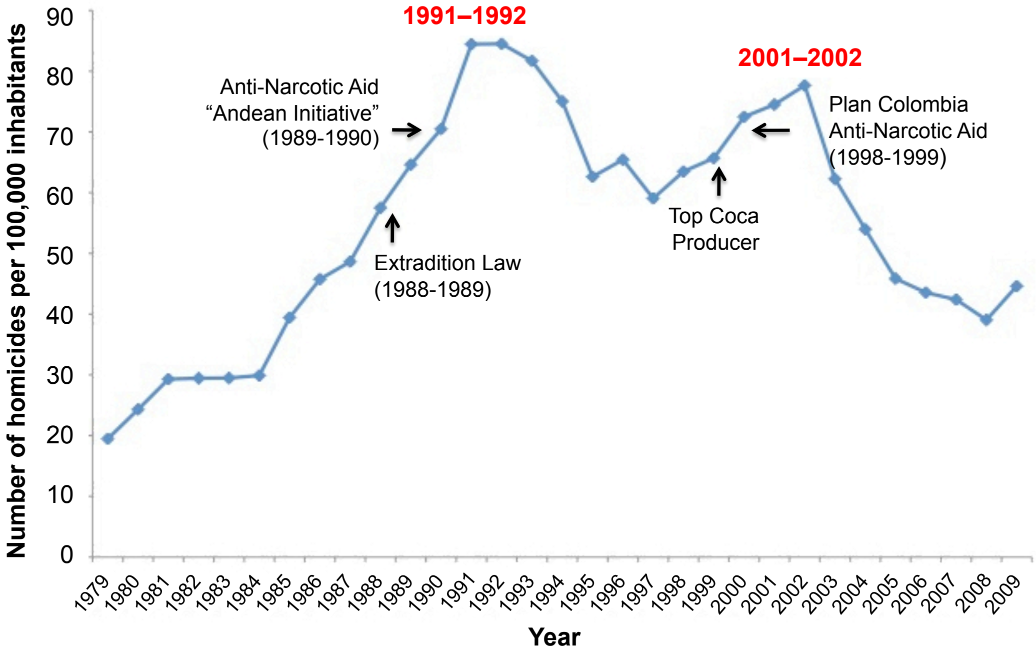
Notes: Own calculations using data from Departamento Administrativo Nacional de Estadística (DANE).

Figure 2.2: Sex Ratios Comparison in 1993 by Level of Violence



Notes: Own calculations using data from Departamento Administrativo Nacional de Estadística (DANE).

Figure 2.3: Homicide Rate per 100,000 Inhabitants in Colombia



Notes: Own calculations using data from Departamento Administrativo Nacional de Estadística (DANE).

Table 2.1: Summary Statistics

Variable	Obs.	Mean	St. Dev.	Min.	Max.
Violence exposure					
From 15 to 36 years old	82,012	0.0660	0.0534	0	0.4990
From 20 to 30 years old	59,783	0.0678	0.0564	0	0.4990
From 16 to 24 years old	75,632	0.0660	0.0569	0	0.5118
Women characteristics					
Woman Age	82,012	26.1881	7.7550	15	43
No education/Less than primary	82,012	0.1348	0.3415	0	1
Complete Primary	82,012	0.1261	0.3319	0	1
Incomplete Secondary	82,012	0.3071	0.4613	0	1
Complete Secondary	82,012	0.2394	0.4267	0	1
Higher Education	82,012	0.1926	0.3943	0	1
Never Married	82,012	0.3756	0.4843	0	1
Married	82,012	0.1556	0.3625	0	1
Living together	82,012	0.3427	0.4746	0	1
Widowed	82,012	0.0096	0.0977	0	1
Divorced/Living not together	82,012	0.1164	0.3207	0	1
Currently Working	82,012	0.4722	0.4992	0	1
Non Migrants[a]	82,012	0.6584	0.4743	0	1
Partner Education Level					
Partner's age	40,852	34.2901	8.9009	15	83
No education/Less than primary	51,184	0.2006	0.4004	0	1
Complete Primary	51,184	0.1764	0.3812	0	1
Incomplete Secondary	51,184	0.3402	0.4738	0	1
Complete Secondary	51,184	0.1288	0.3350	0	1
Higher Education	51,184	0.1376	0.3444	0	1
No Information on education	51,184	0.0164	0.1271	0	1
Head of Household Characteristics					
HofH Gender: Male	82,012	0.6892	0.4628	0	1
HofH Age	82,012	43.3640	13.8553	15	98
Relationship with HofH:					
HofH	82,012	0.1147	0.3186	0	1
Wife	82,012	0.3755	0.4843	0	1
Daughter/Daughter-in-law	82,012	0.3920	0.4882	0	1
Other Relative	82,012	0.0898	0.2858	0	1
Not Related	82,012	0.0281	0.1651	0	1
Dwelling Characteristics					
Urban	82,012	0.7463	0.4351	0	1

[a]Non-migrants are individuals who have been living in the same municipality for all of their fertile years.

Table 2.2: Summary Statistics: Fertility Characteristics

Variable	Obs.	Mean	St. Dev.	Min.	Max.
Current Children					
Number of children	82,012	1.4356	1.6009	0	14
Number of girls	82,012	0.7003	0.9847	0	8
Number of boys	82,012	0.7353	1.0175	0	12
Number of children alive	82,012	1.3890	1.5344	0	14
Fertility Preferences					
Number of desired children	82,012	2.1641	1.0681	0	20
Number of desired girls	82,012	1.1269	0.6812	0	20
Number of desired boys	82,012	1.0371	0.6703	0	20
Diff. between desired boys-desired girls	82,012	-0.0897	0.8282	-20	20
Preference for girls	82,012	0.1860	0.3891	0	1
Preference for boys	82,012	0.1229	0.3283	0	1
Women indifferent between boys and girls	82,012	0.6910	0.4620	0	1

Table 2.3: Homicide Victim Characteristics

Total number of homicides 1979-2009	628,103

Gender of the victim	Number of Homicides	%
Male	580,217	92.38
Female	47,685	7.59
No information	201	0.03
Total	628,103	100.00

Age of the Victim	Number of Homicides	%
0-5	2,717	0.43
6-10	1,797	0.29
11-17	37,496	5.97
18-25	194,587	30.98
26-35	190,315	30.30
36-45	103,670	16.51
45-55	48,069	7.65
56-65	20,420	3.25
65 +	11,287	1.80
No Info	17,745	2.83
Total	628,103	100.00

Marital Status Victim	Number of Homicides	%	%[1]
Single	288,611	45.95	51.87
Married	119,464	19.02	21.47
Cohabitation	80,629	12.84	14.49
Widow	7,012	1.12	1.26
Divorced	56,853	9.05	10.22
Separated	3,892	0.62	0.70
No information	71,642	11.41	
Total	628,103	100.00	

Education Level [2]	Number of Homicides	%	%[3]
No Education	5,716	2.43	4.95
Pre-School	394	0.17	0.34
Primary Complete	33,504	14.25	29.02
Primary Incomplete	31,986	13.60	27.71
Secondary Complete	18,747	7.97	16.24
Secondary Incomplete	20,193	8.59	17.49
University Complete	2,968	1.26	2.57
University Incomplete	1,937	0.82	1.68
No Information	119,746	50.91	
Total	235,191	100.00	115,445

[1] % without considering missing information
[2] Information available from 2000-2009
[3] % without considering missing information

Table 2.4: Effect of Violence on the Number of Children Born

Dependent variable: Children born									
Full sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	-1.258***	-1.218***	-1.091***						
16 to 24 years	(0.309)	(0.354)	(0.395)						
Exposure to violence from				-1.012**	-1.070**	-1.190**			
20 to 30 years				(0.406)	(0.421)	(0.538)			
Exposure to violence from							-1.353***	-1.377***	-0.917**
15 to 36 years							(0.349)	(0.427)	(0.355)
Observations	75,632	75,632	75,632	59,783	59,783	59,783	82,012	82,012	82,012
R-squared	0.374	0.381	0.386	0.277	0.285	0.288	0.407	0.414	0.419
Non-migrant sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	-1.332***	-1.347***	-1.230***						
16 to 24 years	(0.281)	(0.331)	(0.435)						
Exposure to violence from				-1.317***	-1.397***	-1.745***			
20 to 30 years				(0.369)	(0.386)	(0.673)			
Exposure to violence from							-1.293***	-1.378***	-0.801**
15 to 36 years							(0.314)	(0.386)	(0.377)
Observations	47,764	47,764	47,764	34,779	34,779	34,779	53,994	53,994	53,994
R-squared	0.385	0.391	0.397	0.285	0.290	0.294	0.422	0.427	0.434
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Socio-demographic Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table 2.5: Effect of Violence on the Number of Children Born by Age. Full Sample

Dependent variable: Children born	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from 16 to 24 years	-0.323	-0.0738	-0.197						
	(0.684)	(0.636)	(0.391)						
Exposure to violence from 20 to 30 years				-1.598**	-1.582**	-2.071**			
				(0.619)	(0.655)	(0.931)			
Exposure to violence from 15 to 36 years							-0.726	-0.562	-0.503
							(0.609)	(0.571)	(0.337)
Age18-20*Exposure	-0.712***	-0.882***	-0.757***	0.731	0.856	1.237	-0.556*	-0.689**	-0.551**
	(0.268)	(0.281)	(0.271)	(0.968)	(0.944)	(1.036)	(0.284)	(0.277)	(0.267)
Age21-23*Exposure	-0.581	-0.872*	-0.733	0.998	0.973	1.350	-0.370	-0.647	-0.542
	(0.522)	(0.509)	(0.513)	(0.862)	(0.882)	(0.946)	(0.589)	(0.560)	(0.555)
Age24-26*Exposure	-0.647	-0.924	-0.816	0.913	0.882	1.196*	-0.471	-0.680	-0.621
	(0.664)	(0.649)	(0.598)	(0.630)	(0.617)	(0.687)	(0.775)	(0.753)	(0.716)
Age27-29*Exposure	-0.712	-1.012	-0.927	0.656	0.509	0.690	-0.733	-0.993	-0.980
	(0.785)	(0.767)	(0.699)	(0.636)	(0.625)	(0.660)	(0.874)	(0.856)	(0.858)
Age30-32*Exposure	-1.227	-1.459**	-1.396**	0.0587	-0.148	-0.0733	-1.394	-1.655*	-1.695*
	(0.794)	(0.732)	(0.634)	(0.645)	(0.626)	(0.651)	(0.974)	(0.908)	(0.894)
Age33-35*Exposure	-1.186	-1.434*	-1.347**	0.357	0.200	0.240	-1.289	-1.523	-1.523
	(0.825)	(0.772)	(0.674)	(0.619)	(0.623)	(0.614)	(1.123)	(1.059)	(1.076)
Age36-38*Exposure	-0.963	-1.109	-0.956	1.073	0.934	0.963	-0.615	-0.839	-0.749
	(0.901)	(0.890)	(0.824)	(0.704)	(0.706)	(0.748)	(1.394)	(1.346)	(1.454)
Age39-41*Exposure	-1.619*	-1.657**	-1.512**	0.0883	0.170	0.246	-1.666	-1.704	-1.570
	(0.829)	(0.781)	(0.664)	(0.733)	(0.763)	(0.709)	(1.291)	(1.209)	(1.264)
Age42-43*Exposure	-1.438	-1.502	-1.101				-1.033	-1.142	-0.710
	(1.459)	(1.481)	(1.271)				(1.778)	(1.764)	(1.816)
Observations	75,632	75,632	75,632	59,783	59,783	59,783	82,012	82,012	82,012
R-squared	0.374	0.382	0.386	0.277	0.285	0.288	0.407	0.414	0.419
Year, Municipality and Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Socio-demographic Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table 2.6: Effect of Violence on the Number of Children Born by Age. Non-Migrant Sample

Dependent variable: Children born									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	0.0327	0.193	0.0480						
16 to 24 years	(0.599)	(0.574)	(0.418)						
Exposure to violence from				-2.348***	-2.477***	-4.052***			
20 to 30 years				(0.676)	(0.680)	(1.101)			
Exposure to violence from							-0.471	-0.392	-0.382
15 to 36 years							(0.533)	(0.516)	(0.360)
Age18-20*Exposure	-0.824***	-1.051***	-0.964***	1.943**	2.205**	3.378***	-0.623**	-0.793***	-0.716***
	(0.248)	(0.282)	(0.275)	(0.963)	(0.930)	(1.184)	(0.246)	(0.260)	(0.269)
Age21-23*Exposure	-0.606	-0.847*	-0.818	1.888**	2.014**	3.221***	-0.367	-0.588	-0.654
	(0.489)	(0.510)	(0.527)	(0.916)	(0.916)	(1.172)	(0.555)	(0.564)	(0.576)
Age24-26*Exposure	-0.287	-0.570	-0.654	1.970**	2.032***	2.978***	-0.0418	-0.253	-0.472
	(0.730)	(0.733)	(0.659)	(0.766)	(0.755)	(1.026)	(0.847)	(0.843)	(0.783)
Age27-29*Exposure	-0.772	-1.044	-1.167**	1.405*	1.346*	1.982**	-0.774	-1.015	-1.345*
	(0.702)	(0.675)	(0.572)	(0.791)	(0.773)	(0.887)	(0.764)	(0.738)	(0.715)
Age30-32*Exposure	-1.955**	-2.140***	-2.293***	0.0708	0.0281	0.502	-2.194**	-2.396***	-2.829***
	(0.770)	(0.708)	(0.555)	(0.893)	(0.854)	(0.932)	(0.950)	(0.886)	(0.783)
Age33-35*Exposure	-1.722**	-1.884***	-2.015***	0.834	0.889	1.309	-1.817*	-1.949**	-2.407**
	(0.711)	(0.672)	(0.569)	(0.872)	(0.868)	(0.966)	(1.022)	(0.978)	(0.970)
Age36-38*Exposure	-2.314***	-2.319***	-2.373***	0.823	0.931	1.260	-2.226*	-2.288*	-2.643**
	(0.782)	(0.798)	(0.732)	(0.874)	(0.875)	(1.020)	(1.256)	(1.236)	(1.284)
Age39-41*Exposure	-2.511***	-2.558***	-2.646***	0.372	0.533	0.939	-2.774**	-2.841**	-3.192**
	(0.856)	(0.835)	(0.814)	(1.016)	(1.015)	(1.021)	(1.303)	(1.251)	(1.312)
Age42-43*Exposure	-2.079	-2.257	-2.190				-1.991	-2.228	-2.399
	(1.589)	(1.577)	(1.421)				(1.910)	(1.875)	(1.913)
Observations	47,764	47,764	47,764	34,779	34,779	34,779	53,994	53,994	53,994
R-squared	0.386	0.392	0.398	0.285	0.291	0.295	0.423	0.428	0.435
Year, Municipality and Age-group fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table 2.7: Effect of Violence on the Ideal Number of Children

Dependent variable: Ideal children									
All sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	-0.392***	-0.388***	-0.290						
16 to 24 years	(0.145)	(0.147)	(0.265)						
Exposure to violence from				-0.273	-0.301	-0.401			
20 to 30 years				(0.231)	(0.222)	(0.334)			
Exposure to violence from							-0.359**	-0.353**	-0.0005
15 to 36 years							(0.173)	(0.170)	(0.286)
Observations	75,632	75,632	75,632	59,783	59,783	59,783	82,012	82,012	82,012
R-squared	0.097	0.098	0.100	0.098	0.099	0.102	0.096	0.097	0.099
Non-Migrant sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	-0.373**	-0.364**	-0.105						
16 to 24 years	(0.148)	(0.152)	(0.250)						
Exposure to violence from				-0.240	-0.239	-0.264			
20 to 30 years				(0.300)	(0.285)	(0.422)			
Exposure to violence from							-0.403**	-0.406**	0.078
15 to 36 years							(0.196)	(0.200)	(0.321)
Observations	47,764	47,764	47,764	34,779	34,779	34,779	53,994	53,994	53,994
R-squared	0.102	0.103	0.106	0.108	0.109	0.112	0.099	0.100	0.103
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Socio-demographic Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table 2.8: Effect of Violence on the Ideal Difference Boys-Girls

Dependent variable: Ideal difference boys-girls									
All sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	-0.251**	-0.288**	-0.164						
16 to 24 years	(0.117)	(0.122)	(0.205)						
Exposure to violence from				-0.390***	-0.399***	-0.259			
20 to 30 years				(0.135)	(0.136)	(0.233)			
Exposure to violence from							-0.093	-0.118	-0.126
15 to 36 years							(0.144)	(0.146)	(0.193)
Observations	75,632	75,632	75,632	59,783	59,783	59,783	82,012	82,012	82,012
R-squared	0.019	0.020	0.021	0.022	0.022	0.024	0.018	0.018	0.020
Non-Migrant sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	-0.208	-0.267*	-0.125						
16 to 24 years	(0.145)	(0.143)	(0.260)						
Exposure to violence from				-0.571***	-0.601***	-0.543			
20 to 30 years				(0.177)	(0.203)	(0.341)			
Exposure to violence from							-0.034	-0.084	-0.186
15 to 36 years							(0.142)	(0.142)	(0.219)
Observations	47,764	47,764	47,764	34,779	34,779	34,779	53,994	53,994	53,994
R-squared	0.022	0.023	0.025	0.027	0.028	0.030	0.020	0.021	0.023
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Socio-demographic Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table 2.9: Multinomial Logit Estimates of the Effect of Violence on the Gender Preferences for Children

Dependent variable: Gender preferences for children									
All sample:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Preference for boys: Reference group									
Preference for girls									
Exposure to violence from	1.019*	1.275**	0.702						
16 to 24 years	(0.586)	(0.623)	(1.020)						
Exposure to violence from				2.059***	2.180***	1.218			
20 to 30 years				(0.559)	(0.564)	(0.938)			
Exposure to violence from							0.811	1.008	-0.107
15 to 36 years							(0.711)	(0.725)	(1.057)
Same number of girls and boys									
Exposure to violence from	1.447**	1.549**	1.266						
16 to 24 years	(0.625)	(0.670)	(0.838)						
Exposure to violence from				1.977***	2.092***	1.293**			
20 to 30 years				(0.434)	(0.436)	(0.643)			
Exposure to violence from							1.533**	1.606**	0.896
15 to 36 years							(0.657)	(0.708)	(0.835)
Observations	75,632	75,632	75,632	59,783	59,783	59,783	82,012	82,012	82,012
Year Fixed Effects, Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects, Socio-demographic Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table 2.10: Multinomial Logit Estimates of the Effect of Violence on the Gender Preferences for Children (Marginal Effects, $\frac{dP_i}{dx}$)

Dependent variable: Gender preferences for children									
All sample:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Preference for boys									
Exposure to violence from	-0.145**	-0.159**	-0.122						
16 to 24 years	(0.066)	(0.070)	(0.089)						
Exposure to violence from				-0.218***	-0.230***	-0.139**			
20 to 30 years				(0.0492)	(0.0489)	(0.0705)			
Exposure to violence from							-0.145**	-0.156**	-0.0727
15 to 36 years							(0.0691)	(0.0737)	(0.087)
Preference for girls									
Exposure to violence from	-0.0312	-0.00655	-0.0549						
16 to 24 years	(0.0408)	(0.0456)	(0.0955)						
Exposure to violence from				0.0580	0.0613	0.0185			
20 to 30 years				(0.0549)	(0.0568)	(0.112)			
Exposure to violence from							-0.0720	-0.0528	-0.126
15 to 36 years							(0.0524)	(0.0532)	(0.102)
Same number of girls and boys									
Exposure to violence from	0.177**	0.166**	0.177						
16 to 24 years	(0.076)	(0.084)	(0.113)						
Exposure to violence from				0.160***	0.169***	0.121			
20 to 30 years				(0.0601)	(0.0621)	(0.115)			
Exposure to violence from							0.217***	0.208**	0.198*
15 to 36 years							(0.0757)	(0.0840)	(0.116)
Observations	75,632	75,632	75,632	59,783	59,783	59,783	82,012	82,012	82,012
Year Fixed Effects, Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects, Socio-demographic Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table 2.11: Multinomial Logit Estimates of the Effect of Violence on the Gender Preferences for Children. Non-Migrant Sample (Marginal Effects, $\frac{dP_i}{dx}$)

Dependent variable: Gender preferences for children									
Non-Migrant Sample:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Preference for boys									
Exposure to violence from	-0.118*	-0.139**	-0.0624						
16 to 24 years	(0.0641)	(0.0655)	(0.0957)						
Exposure to violence from				-0.239***	-0.265***	-0.210**			
20 to 30 years				(0.0595)	(0.0682)	(0.0967)			
Exposure to violence from							-0.105*	-0.122*	0.000370
15 to 36 years							(0.0634)	(0.0656)	(0.0828)
Preference for girls									
Exposure to violence from	-0.0575	-0.0304	-0.0319						
16 to 24 years	(0.0546)	(0.0603)	(0.127)						
Exposure to violence from				0.0729	0.0793	0.0903			
20 to 30 years				(0.0749)	(0.0837)	(0.149)			
Exposure to violence from							-0.118*	-0.0995	-0.147
15 to 36 years							(0.0619)	(0.0641)	(0.124)
Same number of girls and boys									
Exposure to violence from	0.176**	0.170**	0.0943						
16 to 24 years	(0.0806)	(0.0864)	(0.131)						
Exposure to violence from				0.166**	0.186**	0.120			
20 to 30 years				(0.0785)	(0.0793)	(0.161)			
Exposure to violence from							0.224***	0.221**	0.147
15 to 36 years							(0.0842)	(0.0911)	(0.129)
Observations	47,764	47,764	47,764	34,779	34,779	34,779	53,994	53,994	53,994
Year Fixed Effects, Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects, Socio-demographic Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table 2.12: Effect of Violence on the Probability of Being Indifferent Between the Gender of the Children

Dependent variable: Dummy for being indifferent									
All sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	0.177**	0.168**	0.157						
16 to 24 years	(0.0736)	(0.0804)	(0.109)						
Exposure to violence from				0.163***	0.171***	0.0971			
20 to 30 years				(0.0584)	(0.0607)	(0.109)			
Exposure to violence from							0.209***	0.201**	0.158
15 to 36 years							(0.0730)	(0.0805)	(0.110)
Observations	75,632	75,632	75,632	59,783	59,783	59,783	82,012	82,012	82,012
R-squared	0.022	0.023	0.024	0.021	0.022	0.023	0.021	0.023	0.024
Non-Migrant sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	0.181**	0.176**	0.0758						
16 to 24 years	(0.0790)	(0.0841)	(0.127)						
Exposure to violence from				0.182**	0.199**	0.113			
20 to 30 years				(0.0796)	(0.0804)	(0.156)			
Exposure to violence from							0.218***	0.218**	0.101
15 to 36 years							(0.0821)	(0.0892)	(0.120)
Observations	47,764	47,764	47,764	34,779	34,779	34,779	53,994	53,994	53,994
R-squared	0.026	0.027	0.029	0.028	0.030	0.032	0.024	0.026	0.028
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Socio-demographic Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table: 2.13 Effect of Violence on the Difference Between the Number of Boys and Girls Born

Dependent variable: Difference between the number of boys and girls born									
All sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	-0.204	-0.221	-0.235						
16 to 24 years	(0.162)	(0.164)	(0.223)						
Exposure to violence from				-0.340*	-0.341*	-0.486			
20 to 30 years				(0.194)	(0.201)	(0.320)			
Exposure to violence from							-0.308*	-0.307*	-0.260
15 to 36 years							(0.168)	(0.173)	(0.222)
Observations	75,632	75,632	75,632	59,783	59,783	59,783	82,012	82,012	82,012
R-squared	0.008	0.008	0.009	0.009	0.010	0.011	0.007	0.008	0.009
Non-Migrant sample:									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	-0.370*	-0.340*	-0.597*						
16 to 24 years	(0.196)	(0.201)	(0.308)						
Exposure to violence from				-0.409	-0.367	-0.556			
20 to 30 years				(0.270)	(0.268)	(0.453)			
Exposure to violence from							-0.375**	-0.328*	-0.479*
15 to 36 years							(0.186)	(0.190)	(0.285)
Observations	47,764	47,764	47,764	34,779	34,779	34,779	53,994	53,994	53,994
R-squared	0.012	0.013	0.015	0.015	0.017	0.019	0.010	0.012	0.014
0.014	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Municipality Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Socio-demographic Controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table A2.1 Effect of Violence on the Ideal Number of Children. All Sample

Dependent variable: Ideal number of children									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	0.201	0.263	0.204						
16 to 24 years	(0.286)	(0.284)	(0.289)						
Exposure to violence from				-0.272	-0.270	-0.472			
20 to 30 years				(0.384)	(0.376)	(0.572)			
Exposure to violence from							-0.120	-0.0828	-0.0156
15 to 36 years							(0.277)	(0.269)	(0.284)
Age18-20*Exposure	-0.301	-0.302	-0.274	-0.245	-0.169	-0.0238	-0.0835	-0.0480	0.0308
	(0.286)	(0.294)	(0.300)	(0.456)	(0.455)	(0.558)	(0.234)	(0.237)	(0.239)
Age2123*Exposure	-0.290	-0.358	-0.270	0.125	0.0639	0.228	-0.0584	-0.115	0.0211
	(0.309)	(0.313)	(0.326)	(0.499)	(0.501)	(0.552)	(0.238)	(0.239)	(0.264)
Age24-26*Exposure	-0.202	-0.274	-0.179	0.316	0.312	0.441	-0.0236	-0.0617	0.102
	(0.325)	(0.321)	(0.357)	(0.506)	(0.485)	(0.503)	(0.278)	(0.275)	(0.337)
Age27-29*Exposure	-0.741**	-0.826**	-0.751*	-0.0382	-0.0879	0.000687	-0.589*	-0.648*	-0.455
	(0.318)	(0.332)	(0.410)	(0.398)	(0.393)	(0.407)	(0.343)	(0.355)	(0.458)
Age30-32*Exposure	-0.819**	-0.870**	-0.835**	-0.349	-0.418	-0.436	-0.815**	-0.853**	-0.668
	(0.344)	(0.344)	(0.374)	(0.487)	(0.483)	(0.522)	(0.355)	(0.351)	(0.407)
Age33-35*Exposure	-0.915***	-0.985***	-0.982**	-0.289	-0.339	-0.457	-0.813**	-0.864**	-0.674
	(0.335)	(0.334)	(0.384)	(0.465)	(0.461)	(0.477)	(0.384)	(0.377)	(0.476)
Age36-38*Exposure	-0.602	-0.639*	-0.623	0.394	0.346	0.230	-0.118	-0.155	0.0510
	(0.386)	(0.383)	(0.453)	(0.464)	(0.461)	(0.478)	(0.510)	(0.488)	(0.641)
Age39-41*Exposure	-1.009***	-1.012***	-1.010**	-0.402	-0.378	-0.544	-0.898*	-0.889*	-0.697
	(0.365)	(0.367)	(0.403)	(0.504)	(0.508)	(0.544)	(0.507)	(0.494)	(0.569)
Age42-43*Exposure	-0.503	-0.521	-0.473				0.0897	0.0863	0.366
	(0.667)	(0.664)	(0.812)				(0.857)	(0.857)	(1.012)
Observations	75,632	75,632	75,632	59,783	59,783	59,783	82,012	82,012	82,012
R-squared	0.098	0.099	0.099	0.099	0.100	0.100	0.097	0.098	0.098
Year, Municipality and Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Table A2.2 Effect of Violence on the Ideal Number of Children. Non-Migrant Sample

Dependent variable: Ideal number of children	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Exposure to violence from	0.281	0.327	0.365						
16 to 24 years	(0.297)	(0.306)	(0.292)						
Exposure to violence from				-0.915	-0.925	-1.078			
20 to 30 years				(0.823)	(0.836)	(1.141)			
Exposure to violence from							-0.132	-0.114	-0.0389
15 to 36 years							(0.303)	(0.313)	(0.320)
Age18-20*Exposure	-0.240	-0.250	-0.199	0.700	0.799	0.902	0.0130	0.0346	0.110
	(0.323)	(0.337)	(0.349)	(1.001)	(1.012)	(1.134)	(0.271)	(0.277)	(0.286)
Age21-23*Exposure	-0.365	-0.426	-0.294	0.768	0.761	0.828	-0.107	-0.156	-0.0393
	(0.332)	(0.334)	(0.337)	(1.135)	(1.151)	(1.267)	(0.228)	(0.232)	(0.256)
Age24-26*Exposure	0.00981	-0.0349	0.103	1.368	1.421	1.504	0.213	0.192	0.325
	(0.422)	(0.406)	(0.427)	(1.239)	(1.212)	(1.266)	(0.379)	(0.367)	(0.419)
Age27-29*Exposure	-0.674*	-0.728*	-0.590	0.805	0.816	0.881	-0.515	-0.548	-0.385
	(0.399)	(0.424)	(0.510)	(0.995)	(0.987)	(0.986)	(0.445)	(0.471)	(0.600)
Age30-32*Exposure	-1.000**	-1.034**	-0.954**	0.457	0.431	0.310	-0.932**	-0.954**	-0.800
	(0.417)	(0.414)	(0.441)	(1.075)	(1.066)	(1.131)	(0.443)	(0.441)	(0.506)
Age33-35*Exposure	-0.692*	-0.722*	-0.667	0.813	0.804	0.545	-0.621	-0.649	-0.519
	(0.400)	(0.405)	(0.460)	(1.094)	(1.089)	(1.186)	(0.452)	(0.452)	(0.577)
Age36-38*Exposure	-1.000***	-0.998***	-0.961**	0.828	0.851	0.587	-0.656	-0.651	-0.633
	(0.373)	(0.377)	(0.444)	(1.009)	(1.012)	(1.093)	(0.501)	(0.498)	(0.666)
Age39-41*Exposure	-1.387**	-1.405**	-1.379**	0.0730	0.106	-0.232	-1.434*	-1.433*	-1.411*
	(0.686)	(0.672)	(0.598)	(1.565)	(1.569)	(1.729)	(0.790)	(0.783)	(0.762)
Age42-43*Exposure	-0.933	-0.959	-0.887				-1.066	-1.098	-1.145
	(1.739)	(1.759)	(1.984)				(1.600)	(1.616)	(1.923)
Observations	47,764	47,764	47,764	34,779	34,779	34,779	53,994	53,994	53,994
R-squared	0.103	0.104	0.106	0.108	0.109	0.113	0.100	0.101	0.103
Year, Municipality and Age-group Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Cohort Fixed Effects	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Departmental Cubic Trend	No	No	Yes	No	No	Yes	No	No	Yes

Notes: Clustered standard errors by municipality are in parentheses (355 municipalities). Significance levels are at 90(*), 95(**), and 99(***). Exposure to violence from ages 16 to 24 corresponds to the period from -1 sd to +1 sd around the mean age of women at the birth of their first child. Exposure to violence from ages 20 to 30 corresponds to the period from mean age of women at the birth of the first child to the mean age at the birth of their last child. Exposure to violence from ages 15 to 36 corresponds to the period from the minimum age (15) to mean age of women at the birth of their last child +1 sd. Year fixed effects are dummy variables for the years when the DHS survey was conducted. Municipality fixed effects are dummy variables for each municipality. Age-group fixed effects are dummy variables for each age group. We divide all women into 10 brackets, each including three ages (e.g., 15-17, 18-20, 21-23, etc.), the last group includes ages 42 and 43 only. Cohort fixed effects are dummies for the year of birth of the women. Socio-demographic controls include women characteristics (dummy for non-migrant). Departmental cubic trend is a departmental–specific time trend.

Chapter 3

Monitoring Public Procurement

(Evidence from a Regression Discontinuity Design in Chile)

Joint with Stephan Litschig and Dina Pomeranz,
IAE-Barcelona GSE and Harvard University-NBER

3.1 Introduction

The government is the largest buyer in many countries, typically accounting for a very substantial share of the economy (about 17 percent of GDP in the OECD, (OECD, 2011)). At the same time, public procurement is often thought to be rife with waste and corruption, particularly in developing countries. Despite its economic importance, there is very limited empirical evidence of how efficiently procurement is done in practice and how efficiency could be improved.

One of the key tools to limit waste and corruption in public procurement is external auditing by central government agencies. This paper investigates the impact of being audited on a public entity's subsequent procurement practices. We use detailed administrative data on the universe of public purchases of goods and services in Chile to study how the audit impacts the type of procurement modality. There are four main modalities for public procurement in Chile, each with distinct implications for the extent of transparency and competitiveness. Buying from pre-established framework agreements presents little scope for corruption because suppliers, prices and products are predefined. Running an auction potentially allows the buyer to find a more suitable product or service but also opens the door for more malpractice. Direct negotiation gives the buyer even more discretion in choosing a supplier. Finally, purchases in small amounts can also be made outside of the electronic procurement system altogether.

The data used in this project comes from two sources: the public procurement agency ChileCompra, which manages the online procurement platform through which almost all procurement in Chile is conducted, and the national auditing agency Contraloría-General de Chile, which conducts audits of the procurement process. From ChileCompra, we have information on all public procurement processes for goods and services that occurred between 2007 and 2013, representing about 4 percent of GDP each year (ChileCompra, 2013), including the type of good purchased, its purchase modality, number of bidders and final price. Contraloría provides data on audits in 2011, 2012 and 2013. The identification strategy is a fuzzy regression discontinuity design based on a scoring rule that is used by Contraloría to select public entities for audits.

Our findings for the year 2012 suggest that being audited leads to a temporary shift of procurement behavior toward less transparent purchase modalities. The share of total purchases through direct negotiations increases by about 20 percentage points at the expense of purchases through public auctions. The effect is most pronounced during months when the audit is taking place and continues for several months thereafter, but there are no effects in the following year. There is no evidence that the total amount of procurement increases as a result of the audits. We find the same pattern of effects in both 2011 and 2012, which lends credence to the external validity of these results. The analysis of audits in 2013 is underway.

In ongoing work, we are investigating the underlying mechanisms that led to this temporary shift toward less transparency. One possible mechanism is that public agents expect that being audited currently for their past procurement behavior implies a lower audit probability for their current procurement behavior. This could be based on the notion that audits will rarely happen in consecutive years, and typically cover the preceding year. Being currently audited would therefore imply a temporary drop in audit risk during the year of the audit. An alternative mechanism is that during the audits, agents learn that even though auctions are more transparent and therefore recommended by the auditing agency, they involve many more steps than direct negotiations, and that auditors therefore tend to find more problems with auctions than with direct negotiation. This experience could lead procurement officials to use more direct negotiation, in order to avoid being found at fault by the auditors.

While both interpretations seem a priori plausible, there is both evidence and institutional background to suggest that the temporary drop in audit risk is the more likely explanation than the learning story. First, audit findings are released to local officials at the end of the audit, while our results suggest that the shift from auctions to direct negotiations already starts during the audit. Second, if agents learn about costs and benefits of alternative procurement modalities, why would the effect disappear by the start of the following calendar year? Turnover of procurement officials is unlikely to account for such a short-lasting effect.

One piece of evidence that points to the reduced audit risk interpretation is the justification given by procurement agents for using direct negotiation. To use direct negotiation rather than an auction, procurement officers need to provide one of a list of about 20 possible justifications. We find that the increase in direct negotiations as a result of being audited is strongest for the justification that the purchase needs to be made due to emergency reasons, with no time to prepare an auction. Abuse of this particular justification triggers the highest sanctions in the event of an audit. It is the only one for which the head of the public entity can be personally held accountable.

Another result suggesting that the underlying mechanism may be a lower expected audit probability is the type of product for which we observe the shift toward direct negotiations: Small-scale construction of houses, apartments, and bridges account for about one third of the aggregate shift from auctions toward direct negotiations. These are the types of projects and procurement modalities for which collusion between procurement officers and providers is thought to be particularly common. Finally, we intend to do a subgroup analysis, looking at how impacts vary by previous experience with audits. If the audit risk mechanism is the most important, we would expect no difference in response to being audited between entities that had prior experience with audits and those that did not.

Our findings contribute to the literature on public procurement in several ways. First, our paper presents to the best of our knowledge the first causally identified analysis of the impact of an audit on subsequent procurement behavior. It complements a number of papers that find that an increased audit risk reduces corruption (Olken, 2007; Litschig and Zamboni, 2011). While it is important to understand the impact of audit risk, this effect may be substantially different from the impact of an actual audit. The effect is a priori ambiguous and depends on various factors such as prior and posterior beliefs about how likely and how thorough a future audit will be, as well as information that procurement agents acquire during the audit process. Procurement agents may learn about relevant procurement regulations and as a result compliance and efficiency may improve. At the same time, agents might also conclude that the audit is less thorough than expected, or that a repeat audit will now be less likely and waste and corruption may subsequently increase. Our finding that the audits lead to more use of less transparent modalities suggests that there may have been a decrease in compliance. However, additional analysis will be required to distinguish whether on average the audits had a positive or negative effect on efficiency in public procurement in Chile.

For example, we have begun to analyze to what extent the increased amount of direct purchases is a result of increased quantity or unit price. Increased unit price under directly negotiated contracts would be evidence of corruption. Unfortunately, however, the analysis of unit prices is unlikely to be informative due to several conceptual and data limitations. The first is that we have to restrict attention to homogeneous products in order to compare unit prices across purchases, but shirking or corruption are more likely with differentiated products because the purchasing process is both more involved and more discretionary. Moreover, homogeneous products are precisely those that are typically bought through framework agreements - rather than through auctions or direct negotiations - and are thus intrinsically poor candidates to account for the observed shift from auctions to direct negotiations. Last but not least, the data on product quality and units of measurement is sometimes of poor quality and so we had to rely on necessarily arbitrary imputation procedures.

This paper also relates to recent empirical work that investigates how aspects of auction design affect efficiency in public procurement. The closest study by Tran (2009) provides evidence based on one firm's internal records that best-price auctions reduced that firm's bribe payments and profit margins relative to best-value auctions. Another study (Decarolis, 2014) uses a difference-in-differences strategy to show that the use of first-price auctions in Italy reduces costs of public works at the awarding stage but worsens ex post performance compared with an awarding mechanism that automatically eliminates bids that are "too" low to be credible. Coviello and Mariniello (2014) provide evidence that publicity requirements reduce the cost of public procurement in Italy.

The remainder of the paper proceeds as follows. Section 3.2 provides institutional background on public procurement of goods and services and audits of public entities in Chile. Section 3.3 describes the data. Section 3.4 discusses the empirical strategy we use to estimate the impact of audits on subsequent procurement practices. Section 3.5 presents the results and Section 3.6 concludes.

3.2 Institutional Background, Procurement Modalities, and Audit Selection Process

3.2.1 Public Procurement in Chile

The procurement agency ChileCompra manages the online platform, on which most public procurement in Chile takes place.[45] Since its inception in 2003, the platform has grown to serve more than 850 public entities (such as municipalities) and 5500 sub-units (such as schools), with more than 100,000 firms providing goods and services, 90 percent of which are small and medium size firms.

Contraloría General de la República (short Contraloría) is the central government audit agency in charge of monitoring all public entities, including ministries, municipalities, public services, and state-owned enterprises. Contraloría's primary monitoring activity consists of audits, which account for more than 80 percent of its annual budget. Contraloría considers a universe of about 1,500 public entities and conducts audits of different types in about 30 percent of them every year. Audit types include some that are directly investigating public outlays on goods and services or capital projects, as well as some that investigate related aspects, such as transfers to individuals and firms. Most entities that are selected for audit receive at least one type of audit that directly or indirectly involves public procurements.

3.2.2 Procurement Modalities

Framework agreements for standardized goods

A growing share of public procurement is conducted through an online "supermarket" for public entities that ChileCompra is maintaining for standardized goods, such as office supplies or medicines. For this purpose, ChileCompra organizes auctions to establish framework agreements, which have common conditions for any interested buyer. Once winners are selected, the product becomes listed in the supermarket and procurement regulation requires buyers to purchase the product through the supermarket unless a better option is available in their region (Decree 250, Art. 8).[46] This allows public entities to purchase certain products directly at

[45]There are a few exceptions for transacting outside the platform, for example for purchases by the armed forces or for very small transactions. Large public works are not included in the ChileCompra procurement system and are handled by a different agency.

[46]Municipalities are exempt from this requirement. The military and the police are exempt from buying form the supermarket for certain goods and services.

the common conditions, without incurring the cost of running an auction. Price and availability may vary by region due to transportation costs. Over the past several years, ChileCompra has made an effort to increase the number of products, as well as the regions in which they are available. In 2012, about 16 percent of the value of purchases were made through framework agreements, with 1,500 firms offering 95,000 different products.

Open auctions

In 2012, almost 67 percent of the dollar value of government purchases was made through open auctions (licitación pública). Once a firm has created an account with ChileCompra's online platform, it can participate in any open auction. Open auctions are divided into three categories by size: less than 7,200 USD; between 7,200 USD and 72,000 USD; and more than 72,000 USD.[47] This classification allows small firms to quickly find auctions that are appropriate for their size, using the platform's search feature.

Procurement regulation requires buyers to use open auctions if they cannot find what they need from a framework agreement (Decree 250, Art. 9). Typically, open auctions attract several bidders. However, there is no minimum number of bidders. If no seller makes a bid by the end of the auction's open period, the auction is declared abandoned and the public entity must find a different solution, either by changing the specifications and running another open auction, or by using other modalities such as direct negotiation or a restricted auction (Decree 250, Art. 10).

Restricted auctions and direct negotiation

In a restricted auction (licitación privada), the buyer determines which firms are invited to make an offer. In contrast with open auctions, however, a minimum of three bidders must enter a restricted auction before a winner can be selected. In a direct negotiation (trato directo), a buyer solicits a good or service from at least three different firms (there also exist specific cases, where this is not required, for example if there is only one seller in the market). Chilean procurement regulation allows for a restricted auction or direct negotiation in a number of specific circumstances, and procurement officers must justify their use in the accompanying documentation. In addition to cases of abandoned public auctions, these modalities may also be used

[47]These cutoffs are defined in the regulation in terms of UTM (Unidad Tributaria Mensual), an inflation-adjusted unit used in Chile for tax and other official purposes. Currently, one UTM is equivalent to about 72 USD.

for example when the purchase requires a particularly experienced seller or when the cost of running an auction is deemed disproportionate. Restricted auction and direct negotiation are also permitted for small purchases, up to 720 USD. In 2012, less than one-half percent of auctions were restricted. About 17 percent of the value of purchases was made through direct negotiation.

3.2.3 Audit Selection Process

Contraloría selects entities to be audited using a scoring system with two dimensions: the entity's relative importance in the public sector and the risk of malpractice it presents. The scores are constructed by Contraloría as follows: The score of relative importance is a weighted average of several financial and operational indicators (size of budget or balance sheet, size of social programs budget, size of transfers to private sector, etc.). The risk score is a weighted average of several dimensions, along which the entity might pose a risk, for example based on prior audit results. Based on the two scores, entities are then classified as low, medium, or high in each of the two dimensions, resulting in 9 cells.[48] This calculation is done separately for each region and type of entity (municipalities, hospitals, other).

The selection for an audit is determined by the position of a public entity within the 9 cells. All public entities in the high risk category receive high priority for audits, all entities in the low risk category are low priority. For entities categorized as medium risk, their relative importance determines how they are prioritized for audits. Those with medium risk and high relative importance are high priority, those with low relative importance are low priority, and for those with medium relative importance it depends on available resources. Within each priority group, entities are selected for audit based on other criteria as well, such as time since their most recent audit and involvement in growing industries. Crossing the cutoff between low and high priority therefore increases the probability of being audited, but by less than 100 percentage points. If a public entity has been selected for an audit, Contraloría focuses its investigation on activities of the public entity that they deem of high risk.

We have access to the specific score an entity has for its relative importance, as well as its level of risk, but not its risk score. In our regression discontinuity analysis as described in Section , we will therefore use the thresholds in the relative

[48]Specifically, the range of each score is divided into three equally-sized parts. For example, if the score ranges from 2 to 8, the cutoffs would be at 4 and 6.

importance score among those categorized as medium risk.

3.3 Data

We combine data from ChileCompra's online procurement platform on auctions and purchases with data from Contraloría's audits.

3.3.1 Auctions

We have data on all auctions conducted via ChileCompra's online platform between January 2007 and August 2013. This includes the specifics of the auction: the type and size of the auction; the start and end dates of the auction; product codes (1-4 digits); a detailed description of each item in the auction (including make and model for products); and the quantity of each item solicited. On the bidders' side, the data includes the number and identity of bidders; the content of all bids (quality description, quantity, price); and the winning bid.

Table 3.1 Panel A presents summary statistics for auctions conducted in 2012 in the sample of public entities used in our estimations. Over 78 percent of auctions are open auctions smaller than 7,200 USD, 16 percent are between 7,200 USD and 72,000 USD, and about 6 percent are larger than 72,000 USD. On average, open auctions smaller than 7,200 USD receive 6.21 bids from 6.06 bidders and last 42 days. The number of bids/bidders per auction decreases with the size, while the duration increases. Restricted auctions have both the lowest number of bids/bidders (1.49/1.44) and the shortest duration (16 days). Table A3.1 Panel A presents these statistics among the full sample of public entities. The results are generally similar to those of the estimation sample.

3.3.2 Purchase Orders

Once the seller for a given item has been selected (through auctions, direct negotiation, or the online supermarket), a purchase order is submitted, completing the procurement process. We have data on all purchase orders from January 2007 to August 2013. An auction can result in several purchase orders, or none if the auction does not attract any bidders. A purchase order can include multiple items. The characteristics of the purchase order include: the purchasing entity; purchase modality (open auction, restricted auction, direct negotiation, or online supermarket); date of purchase; product code of each item in the order (1-4 digits); detailed description of each item (including make and model for products); name of the seller; and price and quantity of each item provided.

Table 3.1 Panel B presents summary statistics for purchase orders issued by public entities in our estimation sample in 2012. Framework agreements represent almost 29 percent of orders issued and 33 percent of items purchased, but only 16 percent of total dollars spent. Auctions make up about 48 percent of orders, 49 percent of items, and almost 67 percent of dollars spent. This is because some of the costliest services, such as building renovations and small-scale construction, are often too non-standardized to be offered through framework agreements. Direct negotiations make up around 17 percent of orders, 13 percent of items, and 17 percent of dollars spent. Less than one-half percent of the value of purchases was made outside ChileCompra. Table A3.1 Panel B presents these numbers for the full sample, which are again quite similar to the estimation sample

3.3.3 Audits

Contraloría provides the following data on audits in 2011, 2012, and 2013: the name of the audited entity; the type of audit and whether it is related to public purchases of goods, services, capital projects, or transfers according to Contraloría; and the start and end date of the audit. Table 3.1 Panel C presents summary statistics for audits conducted on entities in our estimation sample in 2012. A total of 164 audits were conducted, 114 of which had a high focus on public procurement (139 had a medium or high focus). Our empirical analysis uses all audits, irrespective of focus on procurement. Audits lasted 65 days on average and 40 days at the median. Table A3.1 Panel C shows these statistics for all audits, the only difference being that audits last longer on average in the full sample.

3.3.4 Timing of Audits

We explore the impact of an audit on purchasing behavior at the aggregate yearly level and also separately during the pre- during- and post-audit period within a given year. The advantage of the disaggregated approach is that it yields an additional internal validity check since we would not expect an audit to have an effect before the public entity even knows about the upcoming audit. It also allows us to examine whether the effect persists once the auditors are not physically present anymore.

For the aggregate approach we simply compute the total amount of purchases during the year in a certain modality. For public entities that were audited, we also compute the amount spent during the pre- during- and post-audit periods within the year. Public entities are informed of the selection for audit about one month in advance and so we consider the pre-period to be all the months from January to one month before the formal start of the audit process. Some audited entities were audited in January or February and these are excluded from the pre-audit analysis. The during-audit period goes from one month before the formal audit process started to the month when it finished. The post-audit period includes the months after the audit process took place until the end of the year. For a few public entities we do not know when the audit finished and these are dropped from the during- and post-audit analysis.

One concern with the disaggregated analysis is seasonality in the outcome variables. We find that while the share of the amount spent by modality does not vary much across months, the amount of purchases and the number of transactions changes considerably across months. During the last months of the year (i.e. November and December) the amount of purchases and the number of transactions are generally larger, while in January the amount of purchases and number of transactions are smaller than during the rest of the year (i.e. February-October) in both 2011 and 2012 (see Appendix Figures A3.2 through A3.5).

This type of seasonality could affect results if we compare transactions by audited entities in the pre- during- or post-audit period to aggregated transactions of entities that were not audited. To address this issue, we construct outcomes for three corresponding periods for the non-audited entities as well, and run three separate regressions that effectively hold seasonal factors constant. Specifically, we use the mean month of formal start of the audit minus one month (to be consistent with the pre-period in the audited units) and the mean month when the audit ended. The resulting pre-audit period for non-audited entities includes the months of January and February, the during-audit period goes from March to August and the post-audit period from September to December.

3.4 Empirical Strategy

3.4.1 Identification

As discussed in Section 3.2.3, Contraloría classifies public entities into 9 groups, based on two continuous underlying scores indicating the entities' relative importance within the public sector, and the risk of malpractice they present. Entities are classified as high, medium, or low along each dimension according to thresholds that vary from year to year. Among the medium risk entities, the probability of being audited changes discontinuously at the threshold of low and medium or medium and high relative importance, respectively in 2011 and 2012. This allows us to use a regression discontinuity design (RDD), which compares entities directly above and directly below the thresholds separating high from medium or medium from low relative importance for a medium level of risk.

The intuition behind an RDD in this context is as follows: among the medium risk entities, those directly below and above the threshold between two categories of relative importance have essentially the same relative importance. At the same time, the probability of being audited should jump discontinuously from one category of relative importance to the next if the cutoffs are indeed used in the audit selection process. This allows us to evaluate the effect of audits by comparing outcomes of entities just below and just above the cutoff.

Formally, let Y_{ij} denote an outcome for public entity i in region j; τ the effect of having been audited; D_{ij} the indicator for having been audited; X_{ij} the relative importance index; c_j the numerical value of the threshold separating two levels of relative importance within region j; $\mathrm{I}[X_{ij} \geq c_j]$ an indicator for an importance score above cutoff j; $f(X_{ij})$ and $g(X_{ij})$ polynomials in the importance score, typically linear splines; π the effect of crossing the cutoff on the audit probability; and U_{ij} and V_{ij} the influence of unobserved factors on outcomes and on the probability of receiving an audit, respectively. The model is then as follows:

$$Y_{ij} = \tau D_{ij} + f(X_{ij}) + U_{ij} \; , \quad (3.1)$$

$$D_{ij} = \pi \mathrm{I}[X_{ij} \geq c_j] + g(X_{ij}) + V_{ij} \; , \quad (3.2)$$

The difference in mean outcomes above and below the cutoff – the RD-gap in Y – is given by:

$$\lim_{X_{ij} \downarrow c_j} \mathrm{E}[Y_{ij} | X_{ij}] - \lim_{X_{ij} \uparrow c_j} \mathrm{E}[Y_{ij} | X_{ij}] \; , \quad (3.3)$$

The key assumption for interpreting the RD-gap in Y as a causal reduced form or intent-to-treat effect of being audited is the continuity of $\mathrm{E}[U_{ij}|X_{ij}]$. Intuitively, this requires that on average, unobserved factors do not exhibit discontinuities at the relevant cutoffs. As shown in Lee and Lemieux (2010), a sufficient condition for continuity of unobservables is that the density of the variable determining treatment assignment is continuous.

In our case, this means public entities can have at most imprecise control over their value of the importance index. This is plausible given that the details of how Contraloría calculates the index are unknown to the public entities. In addition, the cutoffs are determined after the indicators comprising the importance index have been calculated, so nobody knew where the cutoffs were going to be at the time the indicators were calculated. The smooth density assumption therefore seems plausible. While this assumption is not directly testable it has testable implications, which we examine in Section 3.5.2 below.

Identification of τ requires additionally that the exclusion restriction holds. This amounts to the assumption that crossing the cutoff affects outcomes only through the increased probability of receiving an audit, not through other channels. Because the cutoffs are specific to each region and type of entity (see Section 3.2.3), and calculated newly each year specifically for the internal purpose of assigning audits within Contraloría, it is unlikely that anything else also varies exactly at these cutoffs. A final requirement for our estimation strategy is that the RD-gap in D, or first stage, is non-zero and highly significant, in order to mitigate weak instrument bias and size distortion concerns.

If there is entity-specific heterogeneity in the effect of an audit, we additionally need to assume monotonicity, which says in our context that entities that were audited even when marginally below the cutoff would have also been audited if instead they had been marginally above the cutoff. This seems plausible. The ratio of RD-gaps would then identify a local average treatment effect for the subpopulation of entities that were audited because they were marginally above the cutoff. This excludes those that are always selected irrespective of their importance score (perhaps based on expert judgment), as well as those that are never selected, again irrespective of their score.

3.4.2 Estimation Approach

Following Hahn, Todd and Van der Klaauw (2001), Imbens and Lemieux (2008) and Lee and Lemieux (2010), we estimate local linear regressions in samples around the cutoffs between importance categories for entities with a medium level of risk. Following Lemieux (2010), we use OLS with a rectangular kernel, which in effect amounts to giving higher weight to observations closer to a given cutoff c_j. In particular, we estimate the following specifications for observations within a distance h of the cutoff between categories of relative importance:

$$D_{ij} = \pi \mathrm{I}[X_{ij} \geq c_j] + \alpha_0 + \alpha_1 (X_{ij} - c_j) + \alpha_2 (X_{ij} - c_j) \times \mathrm{I}[X_{ij} \geq c_j] + V_{ij}\ , \quad (3.4)$$

$$Y_{ij} = \tau\pi \mathrm{I}[X_{ij} \geq c_j] + \beta_0 + \beta_1 (X_{ij} - c_j) + \beta_2 (X_{ij} - c_j) \times \mathrm{I}[X_{ij} \geq c_j] + U_{ij}\ , \quad (3.5)$$

Where D_{ij}, Y_{ij}, X_{ij}, c_j, V_{ij}, and U_{ij} are as in Section 3.4.1. We show linear specifications for successively larger windows around the cutoff for robustness. We include quadratic specifications as a further robustness check when appropriate.

3.5 Estimation Results

3.5.1 First Stage Results

Figure 3.1 presents the first stage results for 2011 and 2012 in graphical form. For each threshold of relative importance among entities with a medium level of risk, the probability of being audited is plotted against the normalized distance from a unit's corresponding region-specific cutoff. Entities within a maximum distance of ± 10 are considered, and discontinuities are estimated using linear and quadratic splines. Each triangle represents the audit probability in a given bin. The bars indicate the 95 percent confidence intervals. Each graph shows the fitted lines for two polynomial and bandwidth specifications that seem appropriate given the graphical evidence.[49]

As expected for 2011, the discontinuity in the probability of audit occurs between low and medium levels of relative importance, while in 2012, it occurs between medium and high importance. Table 3.2 presents these results in regression form. The discontinuity in 2011 is statistically significant at the 5 percent level in all specifications. Crossing the threshold between low and medium importance for medium-risk entities is associated with a roughly 20 percentage point increase in the probability of being audited. In 2012, the discontinuity is significant at the 1 percent level and stronger in magnitude. Going from medium to high importance at medium risk, the probability of being audited increases by about 45 percentage

[49] In future analysis, we will include the Imbens and Kalyanaraman (2011) bandwidth in the analysis.

points. This indicates that we have a large and highly significant first stage for 2012, which will allow us to evaluate the impact of being audited on public entities' subsequent procurement behavior. Results for 2012 are reported below. The size and significance of the first stage for 2011 are not that strong. We present the results for 2011 in the appendix. The analysis for 2013 is underway.

3.5.2 Internal Validity Checks

Density test

Figure 3.2 shows the results of a McCrary density test (McCrary, 2008) on our estimation sample in 2012. The null hypothesis of the McCrary test is that the density of the treatment-determining variable - in our case the relative importance score - is smooth around the cutoff. The dashed line estimates the density on either side of the cutoff, while the solid lines provide a 95 percent confidence interval for the true density. There is no statistical evidence against the null hypothesis that the density is smooth around the cutoff. Figure A3.1 in the Appendix A3 shows the results for the full sample, with little difference.

Impacts on the share of the amount by modality of purchase in 2010-2011

Table 3.3 displays discontinuity estimates for the share of the amount purchased by modality during the two-year period 2010-2011 prior to the 2012 audits. For the shares of the amount purchased through framework agreements, outside Chile-Compra, and direct negotiation, most estimates are essentially zero and statistically insignificant in all but one case. For the share of the amount bought through auctions, estimates vary between -2 to -8 percentage points and are never significant statistically.

Figure 3.3 presents graphical evidence of these results. Each triangle represents the bin-average of the share of the amount purchased through a given modality in 2010-2011. The bars indicate the 95 percent confidence intervals. Each graph shows the fitted lines for two linear specifications. None of these graphs show evidence of discontinuities in the pre-audit shares of purchase amounts by modality.

3.5.3 Results on Impacts of the Audits

Share of the amount by purchase modality in 2012

Table 3.4 displays reduced form impact estimates on the distribution of the share of the amount purchased by modality during 2012. The estimates suggest that the share of the amount bought through auctions decreased by about 9 percentage points, whereas the share of the amount purchased through direct negotiation increased by about 10 percentage points – significant at 5 percent. Given a first stage estimate of about 45 percentage points, the implied impact estimate of an actual audit is therefore about 9 percent/45 percent equal to 20 percentage points. For the shares bought through framework agreements or outside ChileCompra, impact estimates are essentially zero and insignificant. Overall, there is no evidence that the total amount of procurement increased as a result of the audits (results not shown).

Figure 3.4 presents graphical evidence of these results. The two graphs on the right-hand side show the discontinuities in the shares of the amount bought through auctions and through direct negotiation, respectively. The graphs on the left-hand side show that there are no discontinuities for the shares of purchase amounts done through framework agreements or outside ChileCompra, respectively.

Share of the amount by purchase modality in 2012, Pre-audit period

Table 3.5 displays reduced form impact estimates on the distribution of the share of the amount purchased by modality in 2012 during the pre-audit period. Impact estimates are for the most part small and statistically insignificant, with the exception of outside-ChileCompra purchases that appear slightly higher already in the pre-audit period. Figure 3.5 presents graphical evidence of these results.

Share of the amount by purchase modality in 2012, During-audit period

Table 3.6 displays reduced form impact estimates on the share of the amount purchased by modality in 2012 for the during-audit period. The estimates suggest that the share of the amount bought through auctions decreased by about 14 percentage points, whereas the share of the amount purchased through direct negotiation increased by about 10 percentage points, significant at 5 percent. Outside-ChileCompra purchases also appear slightly higher for the during-audit period. For the shares bought through framework agreements impact estimates are essentially zero and insignificant.

Figure 3.6 presents graphical evidence of these results. The two graphs on the right-hand side show the discontinuities in the shares of the amount bought through auctions and through direct negotiation, respectively. The graphs on the left-hand side show that there are no discontinuities for the shares of purchase amounts done through framework agreements or outside ChileCompra, respectively.

Share of the amount by purchase modality in 2012, Post-audit period

Table 3.7 displays reduced form impact estimates on the share of the amount purchased by modality in 2012 for the post-audit period. None of the estimates are statistically significant. The point estimates suggest that the share of the amount bought through auctions decreased by about 5 percentage points, whereas the share of the amount purchased through direct negotiation increased by about 7 percentage points. For the shares bought through framework agreements or outside ChileCompra, impact estimates are essentially zero.

Figure 3.7 presents graphical evidence of these results. The two graphs on the right-hand side show some evidence of discontinuities in the shares of the amount bought through auctions and through direct negotiation, respectively. The graphs on the left-hand side show that there are no discontinuities for the shares of purchase amounts done through framework agreements or outside ChileCompra, respectively.

Share of the amount by modality of purchase in 2013

Table 3.8 displays reduced form impact estimates on the distribution of the share of the amount purchased by modality in 2013. Impact estimates are for the most part small and statistically insignificant, suggesting that the effect from 2012 has disappeared completely by the subsequent fiscal year. Figure 3.8 presents graphical evidence of these results.

Share of the amount spent on construction by modality of purchase in 2010 and 2011

Table 3.9 displays discontinuity estimates for the share of the amount spent on construction by modality during the two-year period 2010-2011. For the shares of the amount purchased through framework agreements and outside ChileCompra, most estimates are essentially zero and all are statistically insignificant. For the share of the amount bought through auctions and through direct negotiation, estimates vary somewhat more, depending on the specification, and are never significant statistically. Figure 3.9 presents graphical evidence of these results.

Share of the amount spent on construction by modality of purchase in 2012

Table 3.10 displays discontinuity estimates for the share of the amount spent on construction by modality during 2012. The estimates suggest that the share of the amount of construction spending done through auctions decreased by about 14-18 percentage points, whereas the share of construction spending through direct negotiation increased by about 14-18 percentage points, significant at 10 percent. Since construction accounts for about 20 percent of total spending, this shift corresponds to a 3 to 4 percentage point shift in the aggregate amount of purchases. There is also some evidence that construction spending was shifted away from framework agreements and done outside ChileCompra altogether, although impact estimates are smaller and often statistically insignificant. Figure 3.10 presents graphical evidence of these results.

Share of the amount by type of direct negotiation in 2010 and 2011

There are 20 distinct justifications for direct negotiation. The most common justifications are claiming that there is only a single supplier, that there is an emergency, or that a specific supplier is required for confidentiality reasons. Table 3.11 displays discontinuity estimates for the share of the amount spent through direct negotiation by type of justification during the two-year period 2010-2011. For most justifications, estimates are small and typically statistically insignificant. Figure 3.11 presents graphical evidence of these results.

Share of the amount by type of direct negotiation in 2012

Table 3.12 shows discontinuity estimates for the share of the amount spent through direct negotiation by type of justification during 2012. Impact estimates suggest that the share of direct negotiations justified on emergency grounds increased by about 4 percentage points, significant at 10 percent. Figure 3.12 presents graphical evidence of these results.

3.6 Conclusions

Despite the economic importance of public procurement for economies, there is limited evidence on how efficiently procurement is done and how efficiency could be improved. In this paper, we investigate the impact of audits by the Chilean auditing agency Contraloría on subsequent procurement practices by public entities. For identification, we exploit a scoring rule that Contraloría uses to allocate its audits, which allows for an RDD analysis. Our results suggest that the audits lead to a temporary shift toward less transparent modalities of procurement. The share of total purchases through direct negotiations increases by about 20 percentage points at the expense of purchases through public auctions. The effect is most pronounced during months when the audit is taking place and disappears by the subsequent fiscal year. Further analysis will aim to shed more light on mechanisms and investigate whether the same pattern of results holds for 2013.

References

[1] ChileCompra (2013). 6 de Octubre: Evento 10 años de ChileCompra, September 2013.

[2] Coviello, D., and Mariniello, M. (2014). Publicity requirements in public procurement: Evidence from a regression discontinuity design. Journal of Public Economics, 109, 76-100.

[3] Decarolis, F. (2014). Awarding price, contract Performance, and bids screening: Evidence from procurement auctions. American Economic Journal: Applied Economics, 6(1), 108-132.

[4] Hahn, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica, 69(1), 201-209.

[5] Imbens, G. W., and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. Journal of econometrics, 142(2), 615-635.

[6] Lee, D. S., and Lemieux, T. (2009). Regression discontinuity designs in economics. Journal of Economic Literature, 48 (2), 281-355.

[7] Zamboni, Y., and Litschig, S. (2011). Audit risk and rent extraction: Evidence from a randomized evaluation in Brazil. Economics working papers 1270, Department of Economics and Business, Universitat Pompeu Fabra.

[8] McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. Journal of Econometrics, 142(2), 698-714.

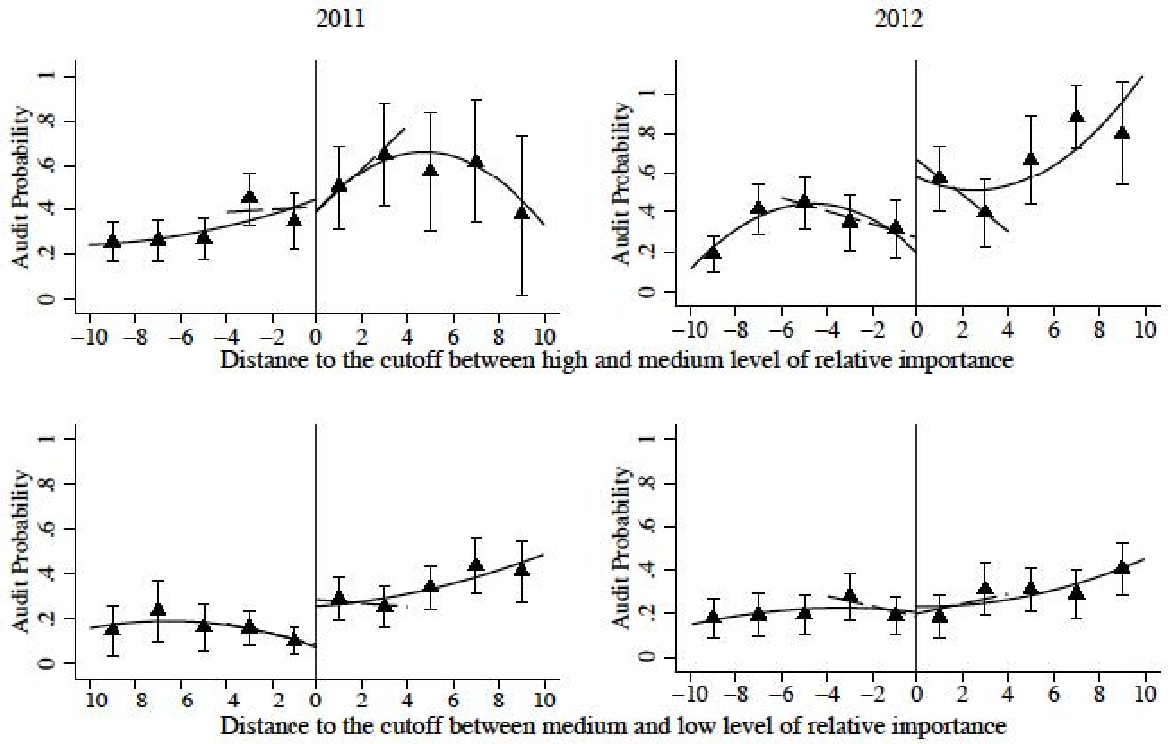
[9] OECD (2011). Government at a Glance 2011. Technical Report.

[10] Olken, B. A. (2007). Monitoring corruption: Evidence from a field experiment in Indonesia. Journal of Political Economy, 115, 200-249.

[11] Tran, A. (2009). Can procurement auctions reduce corruption? Evidence from the internal records of a bribe-paying firm. Working Paper, Harvard University.

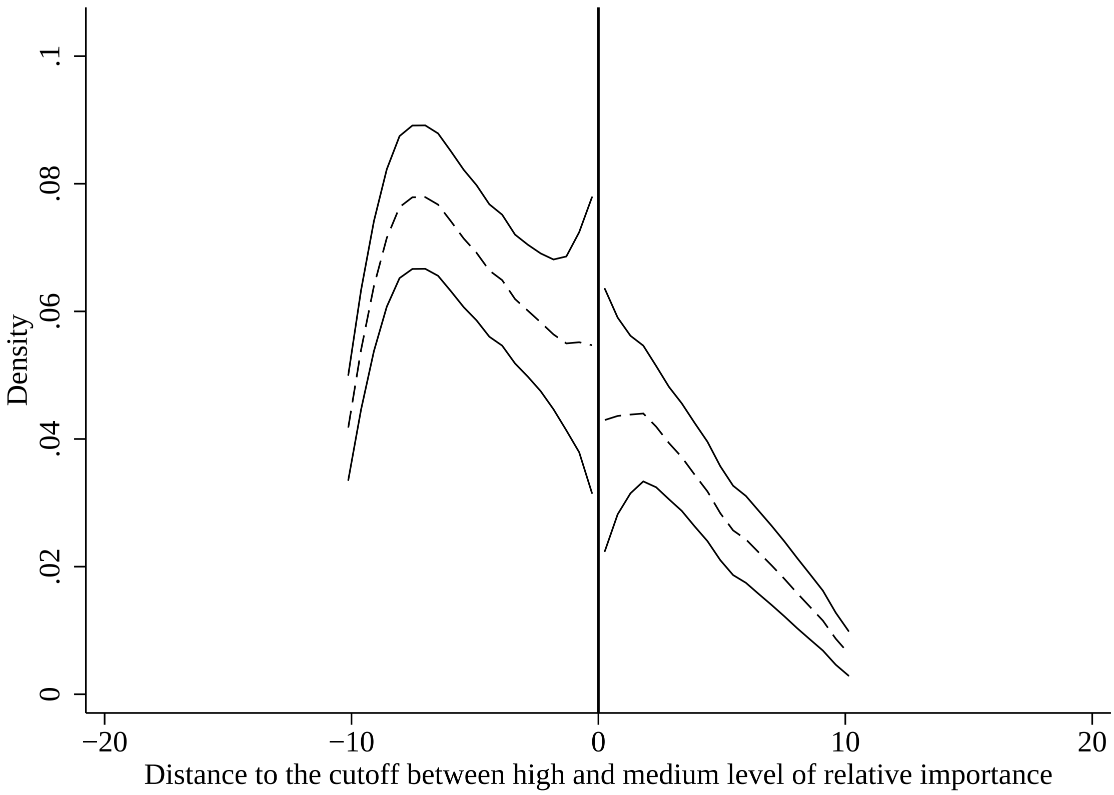
Figures and Tables

Figure 3.1: First Stage Estimates for 2011 and 2012



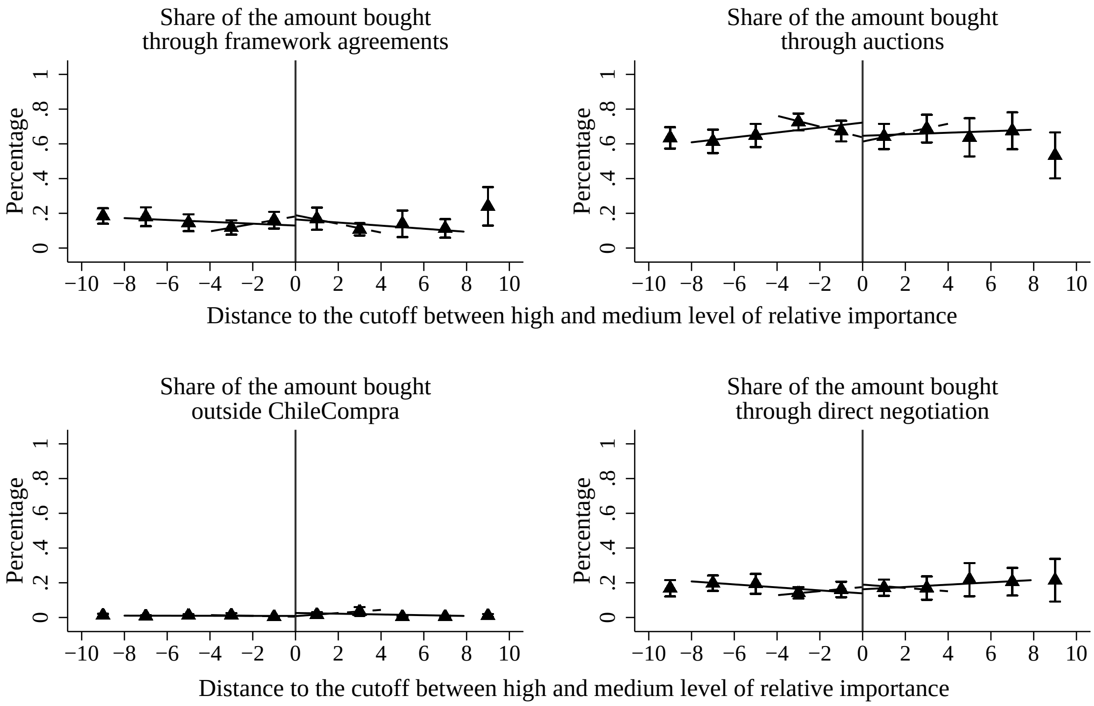
Notes: Dashed line shows linear fit and solid line shows quadratic fit. Triangles represent audit probability in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample in each column consists of entities with normalized importance scores within the ±10 range in 2011 and 2012, respectively.

Figure 3.2: Density test, Estimation Sample



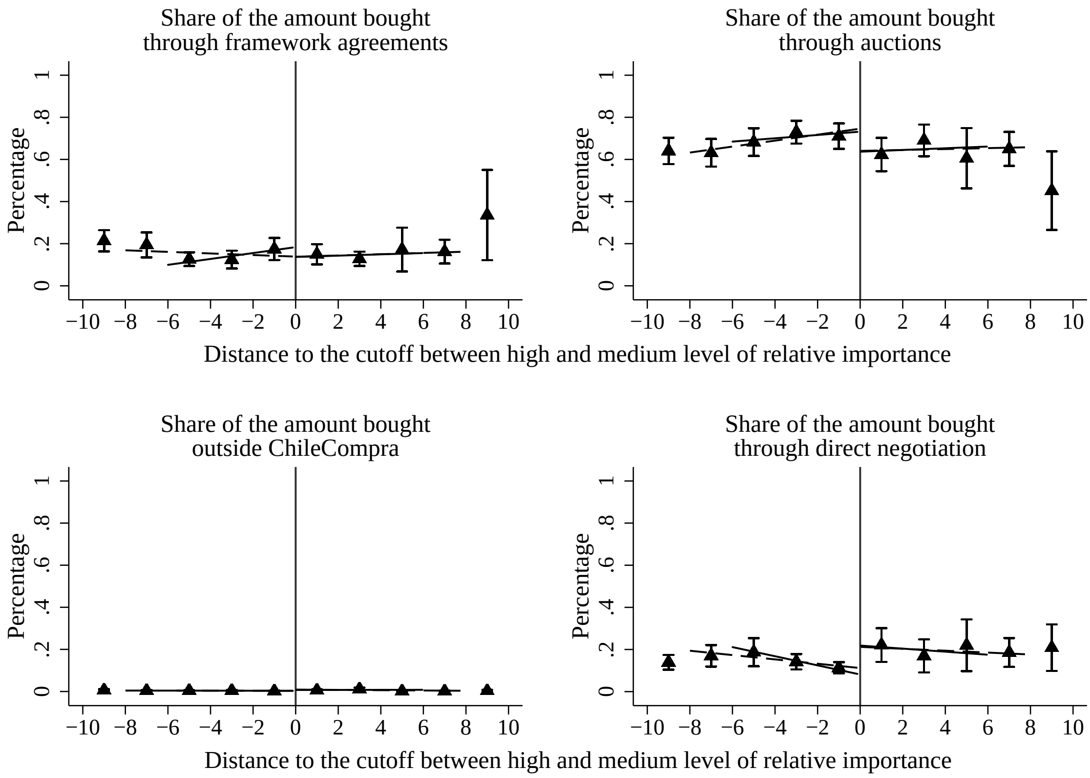
Notes: Dashed line indicates density estimate, and solid lines indicate 95% confidence interval. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012.

Figure 3.3: Impact on the Share of the Amount by Modality of Purchase in 2010 and 2011



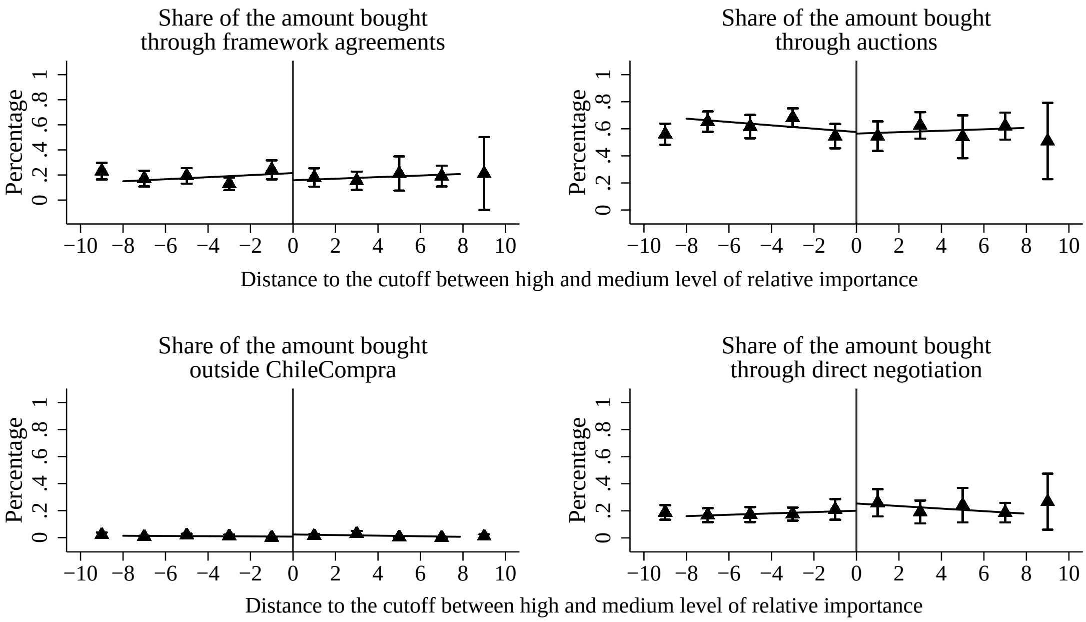
Notes: Dashed and solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012.

Figure 3.4: Impact on the Share of the Amount by Modality of Purchase in 2012



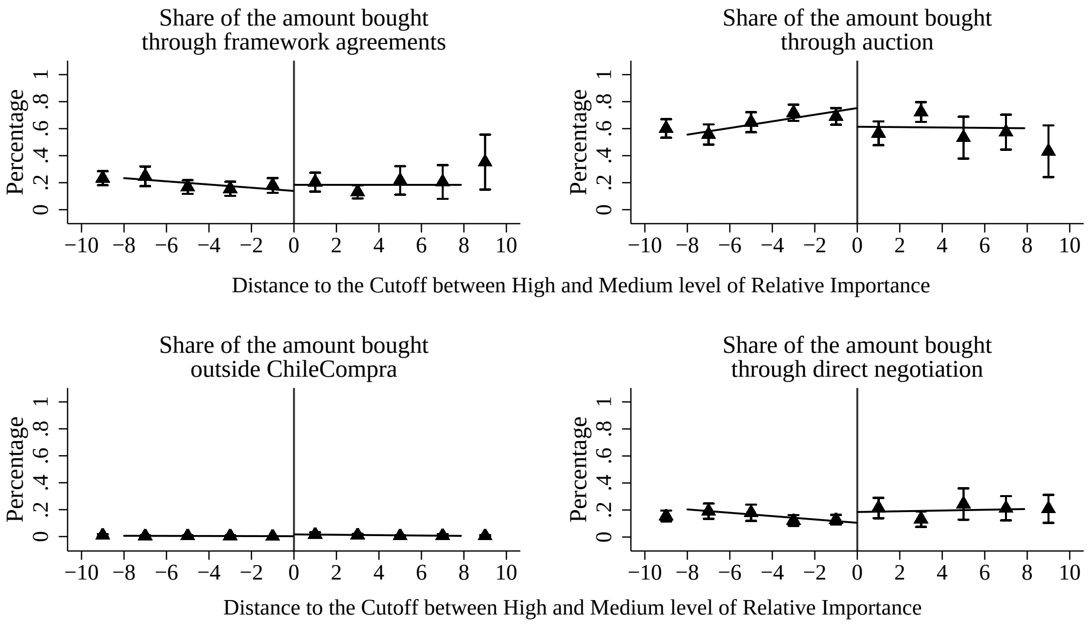
Notes: Dashed and solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012.

Figure 3.5: Impact on the Share of the Amount by Modality of Purchase in 2012, Pre-Audit Period



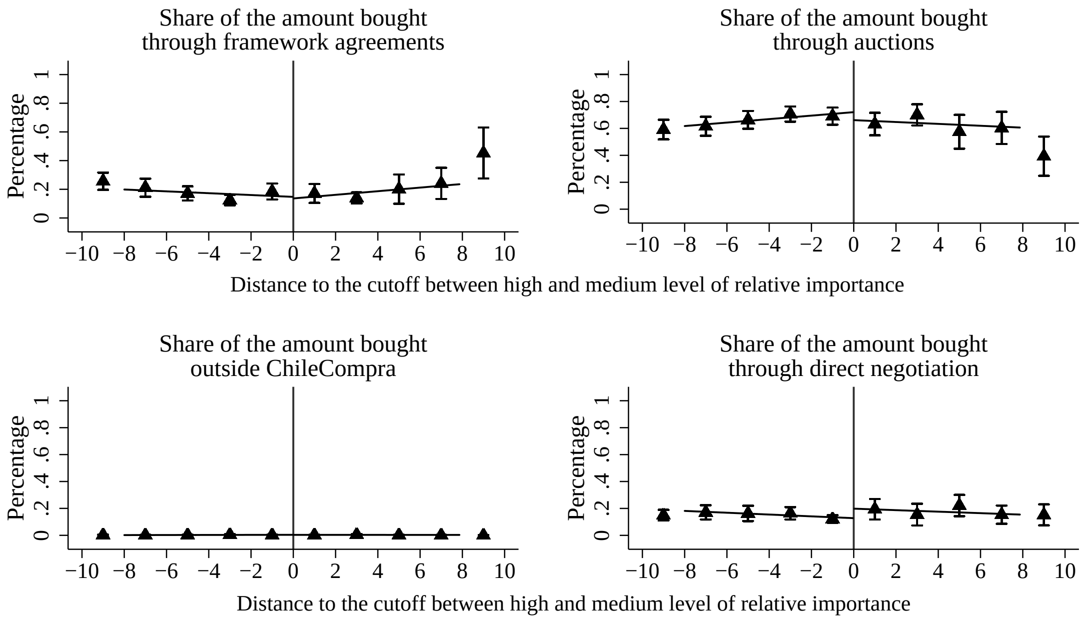
Notes: Dashed and solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012. For audited entities, the pre-audit period includes the months up to one month before the beginning of the audit. For not-audited entities, the pre-audit period corresponds to January and February.

Figure 3.6: Impact on the Share of the Amount by Modality of Purchase in 2012, During-Audit Period



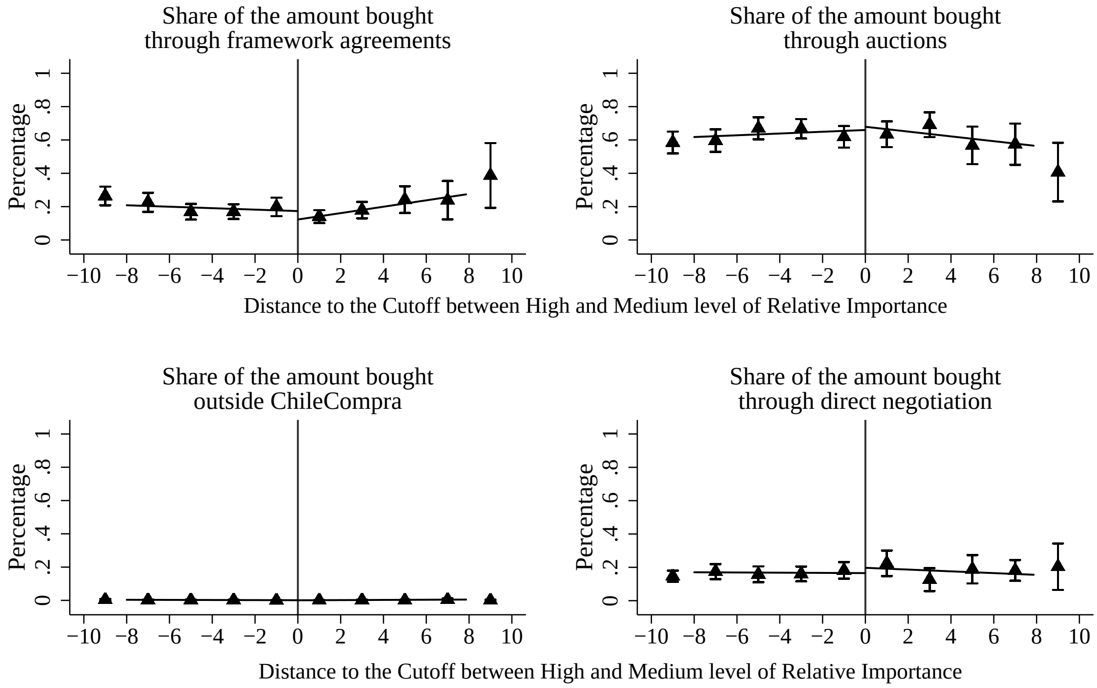
Notes: Dashed and solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012. For audited entities, the during-audit period includes one month before the beginning of the audit. For not-audited entities, the during-audit period corresponds to March through August.

Figure 3.7: Impact on the Share of the Amount by Modality of Purchase in 2012, Post-Audit Period



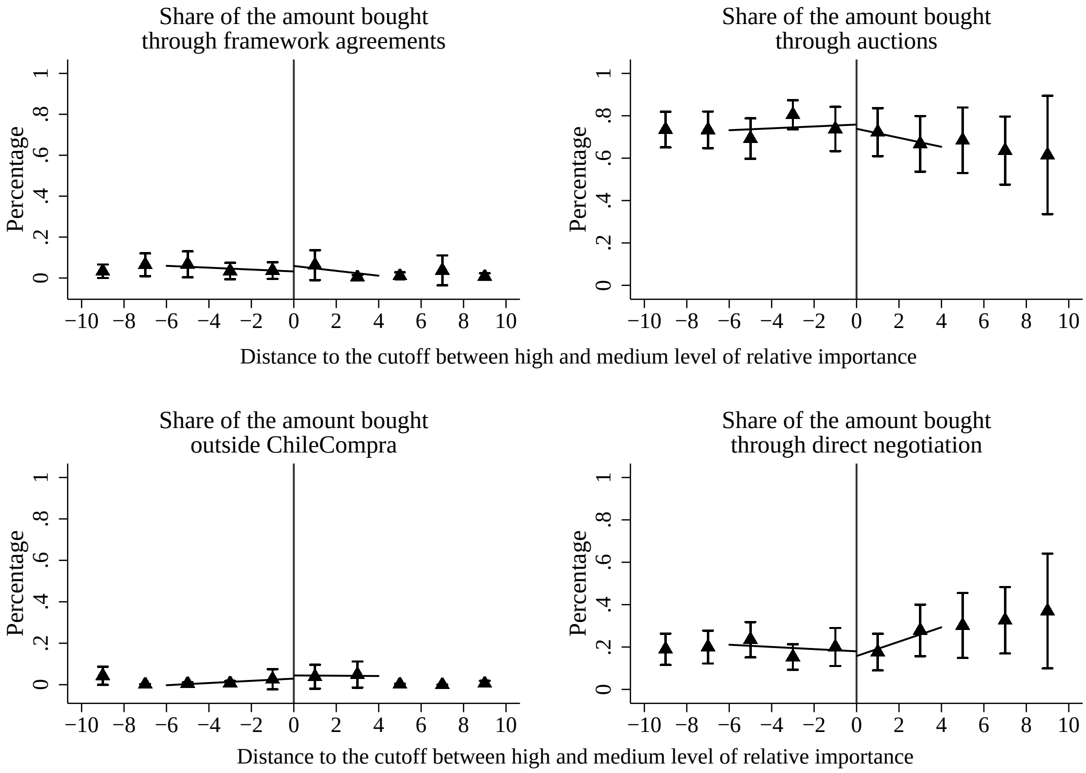
Notes: Dashed and solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012. For audited entities, the post-audit period includes the months after the audit. For not-audited entities, the post-audit period corresponds to September through December.

Figure 3.8: Impact on the Share of the Amount by Modality of Purchase in 2013



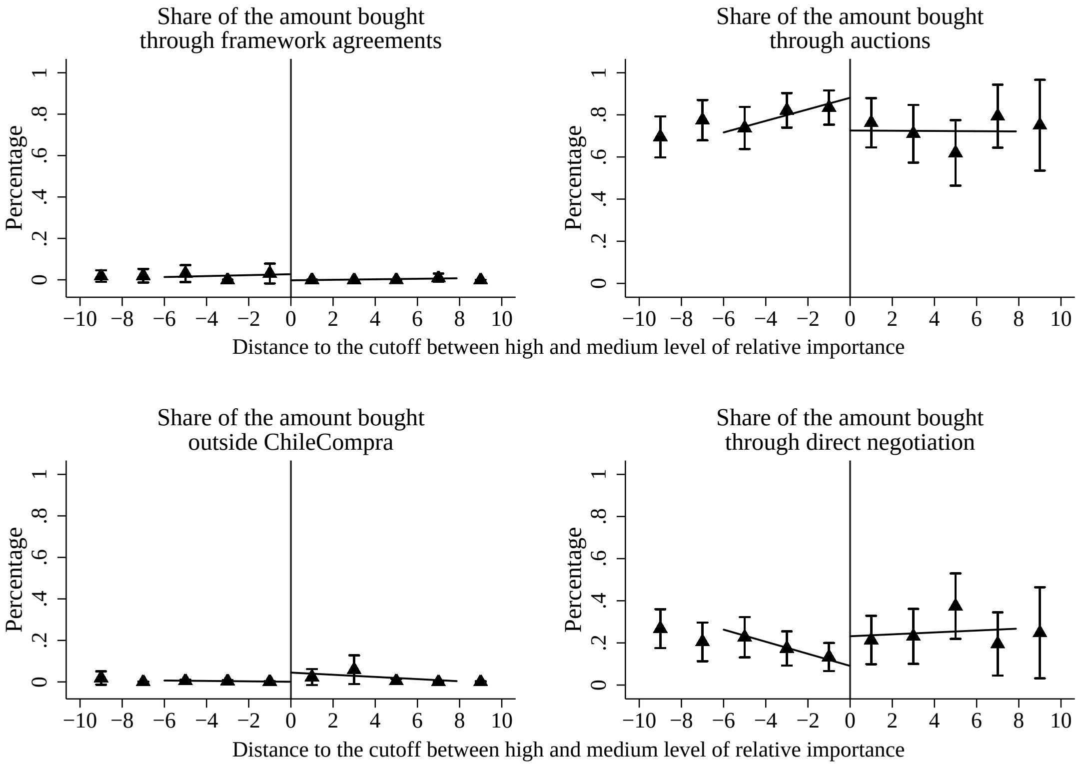
Notes: Dashed and solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012.

Figure 3.9: Impact on the Share of the Amount Spent in Construction by Modality of Purchase in 2010 and 2011



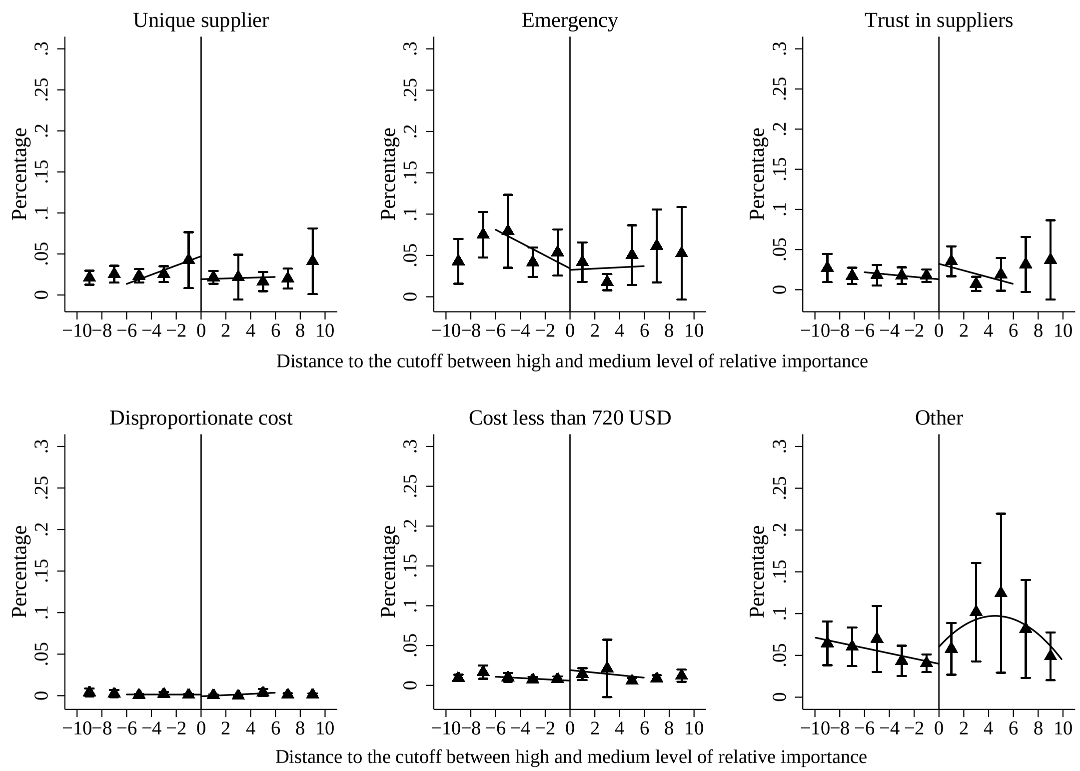
Notes: Construction includes improvements and repairs of houses, apartments, and bridges. Solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012.

Figure 3.10: Impact on the Share of the Amount Spent in Construction by Modality of Purchase in 2012



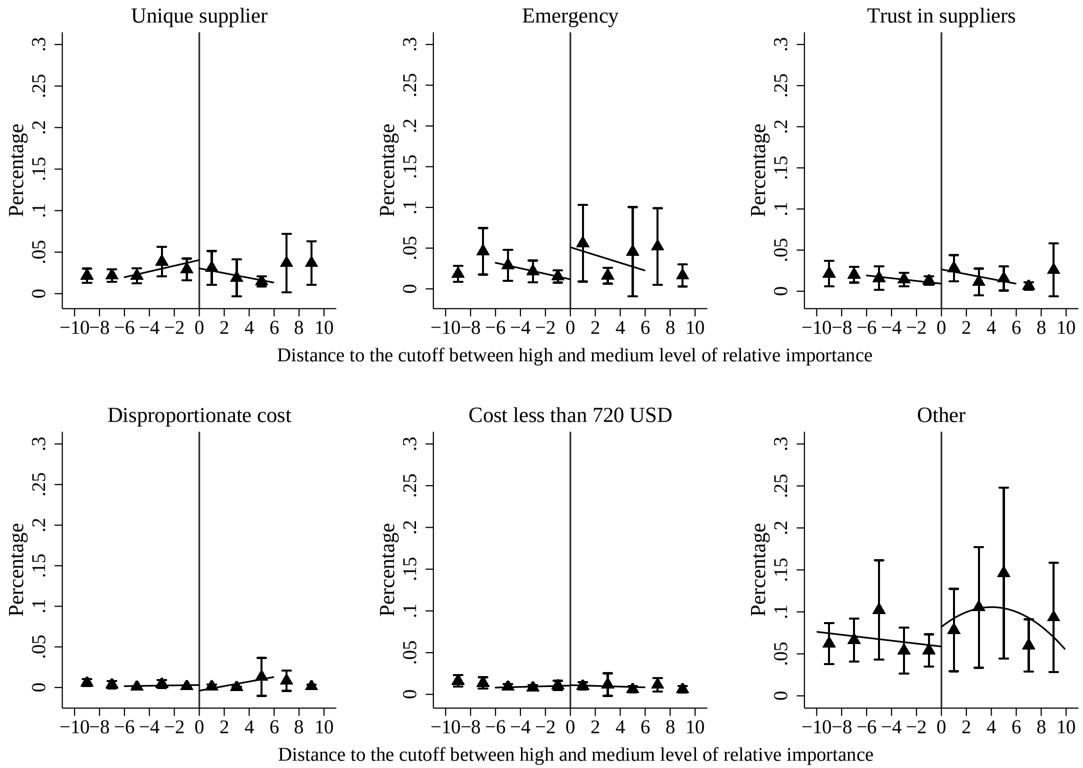
Notes: Construction includes improvements and repairs of houses, apartments, and bridges. Solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012.

Figure 3.11:
Impact on the share of the amount by type of direct negotiation in 2010 and 2011



Notes: Solid lines show linear or quadratic fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012.

Figure 3.12:
Impact on the share of the amount by type of direct negotiation in 2012



Notes: Solid lines show linear or quadratic fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2012.

Table 3.1: Summary Statistics, Estimation Sampae

Panel A: Auctions Type of Auction	Number of Auctions	Share of Auctions	Mean Number of Bids	Mean Number of Bidders	Mean Duration	
Open Auction, <7,200 USD	339,616	78.1%	6.21	6.06	42.16	
Open Auction, 7,200-72,000 USD	69,446	16.0%	5.14	4.98	56.98	
Open Auction, >72,000 USD	24,476	5.6%	4.28	4.01	91.54	
Restricted Auction	1,200	0.3%	1.49	1.44	15.95	
Other	9	0.0%	1.67	1.56	55.33	
Panel B: Purchase Orders Source of Purchase Order	Number of Orders	Share of Orders	Number of Items	Share of Items	Amount in Millions of USD	Share of Dollar Total
Framework Agreement	157,676	28.9%	503,909	32.8%	328	16.0%
Auction	258,850	47.5%	758,035	49.4%	1,360	66.6%
Direct Negotiation	94,273	17.3%	199,233	13.0%	350	17.1%
Outside ChileCompra	33,874	6.2%	72,917	4.8%	4.2	0.2%
Panel C: Audits Public Procurement Focus	Number of Audits	Mean Duration	Median Duration			
High	114	65	41			
High-Medium	139	67	42			
All	164	65	40			

Notes: Sample consists of public entities with normalized importance scores within the ±10 range in 2012. Appendix Table 3.A1 shows the same for the entire universe.

Table 3.2: First Stage Estimates for 2011 and 2012

	Cutoff between high and medium relative importance, medium risk					
	2011			2012		
Dependent variable	Audited (0/1)			Audited (0/1)		
$I[X \geq 0]$	-0.023	0.018	-0.054	0.449***	0.437***	0.424***
	(0.164)	(0.132)	(0.147)	(0.158)	(0.147)	(0.150)
Comparison mean	0.410	0.437	0.441	0.272***	0.285***	0.203**
Bandwidth	±4	±6	±10	±4	+4/-6	±10
Linear spline	✓	✓		✓	✓	
Quadratic spline			✓			✓
Observations	166	273	480	161	216	392
R-squared	0.034	0.046	0.056	0.056	0.040	0.106

	Cutoff between medium and low relative importance, medium risk					
	2011			2012		
Dependent variable	Audited (0/1)			Audited (0/1)		
$I[X \geq 0]$	0.199**	0.164**	0.182**	0.014	0.019	0.026
	(0.085)	(0.072)	(0.089)	(0.090)	(0.097)	(0.094)
Comparison mean	0.088	0.092	0.076	0.186	0.186	0.206
Bandwidth	±4	±6	±10	±4	±6	±10
Linear spline	✓	✓		✓	✓	
Quadratic spline			✓			✓
Observations	362	507	701	274	361	694
R-squared	0.033	0.039	0.058	0.009	0.004	0.023

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±10 range.

Table 3.3: Impact on the Share of the Amount by Modality of Purchase in 2010 and 2011

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
$I[X \geq 0]$	0.004	0.016	0.035	-0.021	-0.071	-0.078
	(0.049)	(0.043)	(0.040)	(0.064)	(0.059)	(0.053)
Comparison mean	0.185	0.15	0.129	0.635	0.715	0.725
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	160	233	313	160	233	313
R-squared	0.033	0.005	0.010	0.041	0.008	0.016
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
$I[X \geq 0]$	0.003	0.016	0.017**	0.014	0.040	0.026
	(0.014)	(0.010)	(0.008)	(0.047)	(0.046)	(0.040)
Comparison mean	0.005	0.007	0.009	0.176	0.127	0.137
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	160	233	313	160	233	313
R-squared	0.062	0.024	0.024	0.014	0.013	0.013

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±8 range in 2012.

Table 3.4: Impact on the Share of the Amount by Modality of Purchase in 2012

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
$\text{I}[X \geq 0]$	-0.019	-0.020	0.017	-0.088	-0.085	-0.108**
	(0.050)	(0.042)	(0.040)	(0.066)	(0.060)	(0.052)
Comparison mean	0.208	0.182	0.140	0.658	0.719	0.739
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	161	234	314	161	234	314
R-squared	0.033	0.018	0.004	0.061	0.026	0.024
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
$\text{I}[X \geq 0]$	-0.000	0.004	0.006**	0.107**	0.102**	0.085*
	(0.004)	(0.003)	(0.002)	(0.053)	(0.051)	(0.043)
Comparison mean	0.003	0.003	0.003	0.132	0.095	0.118
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	161	234	314	161	234	314
R-squared	0.090	0.040	0.028	0.040	0.030	0.017

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±8 range in 2012.

Table 3.5: Impact on the Share of the Amount by Modality of Purchase in 2012, Pre-audit Period

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
I[$X \geq 0$]	-0.082	-0.057	-0.027	0.013	-0.012	-0.042
	(0.063)	(0.054)	(0.053)	(0.083)	(0.070)	(0.067)
Comparison mean	0.236	0.215	0.187	0.559	0.576	0.615
Bandwidth	±6	±8	±10	±6	±8	±10
Linear spline	✓	✓	✓	✓	✓	✓
Observations	219	287	358	219	287	358
R-squared	0.012	0.006	0.002	0.013	0.013	0.003
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
I[$X \geq 0$]	0.018*	0.016**	0.017*	0.051	0.053	0.052
	(0.010)	(0.008)	(0.009)	(0.064)	(0.053)	(0.049)
Comparison mean	0.002	0.008	0.006	0.204	0.201	0.193
Bandwidth	±6	±8	±10	±6	±8	±10
Linear spline	✓	✓	✓	✓	✓	✓
Observations	219	287	358	219	287	358
R-squared	0.024	0.015	0.012	0.012	0.013	0.009

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±10 range in 2012. For audited entities, the pre-audit period includes the months up to one month before the beginning of the audit. For not-audited entities, the pre-audit period corresponds to January and February.

Table 3.6: Impact on the Share of the Amount by Modality of Purchase in 2012, During-audit Period

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
I[$X \geq 0$]	0.048	0.011	0.045	-0.170**	-0.129*	-0.138**
	(0.060)	(0.051)	(0.051)	(0.073)	(0.068)	(0.062)
Comparison mean	0.193	0.178	0.139	0.680	0.732	0.752
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	159	232	312	159	232	312
R-squared	0.025	0.002	0.011	0.066	0.023	0.034
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
I[$X \geq 0$]	0.013	0.014**	0.013***	0.109**	0.103**	0.080*
	(0.008)	(0.006)	(0.005)	(0.054)	(0.051)	(0.046)
Comparison mean	0.002	0.002	0.003	0.124	0.088	0.105
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	159	232	312	159	232	312
R-squared	0.041	0.038	0.034	0.039	0.029	0.021

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±8 range in 2012. For audited entities, the during-audit period includes one month before the beginning of the audit. For not-audited entities, the during-audit period corresponds to March through August.

Table 3.7: Impact on the Share of the Amount bl Modality of Purcease in 2012, Post-audit Period

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
$I[X \geq 0]$	-0.045	-0.025	-0.011	-0.045	-0.050	-0.059
	(0.052)	(0.047)	(0.046)	(0.071)	(0.063)	(0.057)
Comparison mean	0.221	0.178	0.147	0.637	0.704	0.721
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	156	227	306	156	227	306
R-squared	0.036	0.003	0.011	0.037	0.009	0.014
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
$I[X \geq 0]$	-0.003	-0.001	-0.000	0.093	0.077	0.071
	(0.003)	(0.002)	(0.002)	(0.059)	(0.051)	(0.045)
Comparison mean	0.003	0.004	0.004	0.139	0.114	0.128
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	156	227	306	156	227	306
R-squared	0.022	0.004	0.005	0.021	0.015	0.009

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±8 range in 2012. For audited entities, the post-audit period includes the months after the audit. For not-audited entities, the post-audit period corresponds to September through December

Table 3.8: Impact on the Share of the Amount by Modality of Purchase in 2013

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
$I[X \geq 0]$	-0.093**	-0.051	-0.052	0.051	0.019	0.038
	(0.044)	(0.042)	(0.042)	(0.062)	(0.056)	(0.052)
Comparison mean	0.209	0.173	0.159	0.626	0.660	0.666
Bandwidth	±6	±8	±10	±6	±8	±10
Linear spline	✓	✓	✓	✓	✓	✓
Observations	233	313	390	233	313	390
R-squared	0.026	0.019	0.044	0.007	0.008	0.025
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
$I[X \geq 0]$	0.000	-0.000	0.000	0.042	0.032	0.014
	(0.001)	(0.002)	(0.002)	(0.051)	(0.043)	(0.037)
Comparison mean	0.001	0.001	0.001	0.164	0.165	0.174
Bandwidth	±6	±8	±10	±6	±8	±10
Linear spline	✓	✓	✓	✓	✓	✓
Observations	233	313	390	233	313	390
R-squared	0.015	0.016	0.007	0.004	0.002	0.003

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±10 range in 2012.

Table 3.9: Impact on the Distribution of the Share of the Amount Spent on Construction by Modality of Purchase in 2010 and 2011

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
$I[X \geq 0]$	-0.004	0.027	0.010	0.093	-0.020	-0.035
	(0.050)	(0.054)	(0.043)	(0.106)	(0.102)	(0.084)
Comparison mean	0.063	0.032	0.032	0.646	0.759	0.759
Bandwidth	±4	-6/+4	-6/+8	±4	-6/+4	-6/+8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	155	208	243	155	208	243
R-squared	0.010	0.004	0.004	0.045	0.007	0.011
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
$I[X \geq 0]$	0.005	0.015	0.026	-0.094	-0.022	-0.001
	(0.047)	(0.037)	(0.029)	(0.092)	(0.088)	(0.074)
Comparison mean	0.040	0.029	0.029	0.251	0.18	0.18
Bandwidth	±4	-6/+4	-6/+8	±4	-6/+4	-6/+8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	155	208	243	155	208	243
R-squared	0.014	0.018	0.019	0.030	0.009	0.023

Notes: Construction includes improvements and repairs of houses, apartments, and bridges. OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±10 range in 2012.

Table 3.10: Impact on The Distribution of the Share of the Amount Spent on Construction by Modality of Purchase in 2012

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
I[$X \geq 0$]	-0.057**	-0.027	-0.030	-0.135	-0.177*	-0.155*
	(0.025)	(0.031)	(0.025)	(0.108)	(0.102)	(0.086)
Comparison mean	0.057	0.027	0.027	0.839	0.881	0.881
Bandwidth	±4	-6/+4	-6/+8	±4	-6/+4	-6/+8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	149	201	233	149	201	233
R-squared	0.061	0.009	0.010	0.022	0.021	0.023
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
I[$X \geq 0$]	0.020	0.020	0.044**	0.172*	0.184*	0.141*
	(0.035)	(0.028)	(0.022)	(0.102)	(0.097)	(0.082)
Comparison mean	0.001	0.001	0.001	0.103	0.091	0.091
Bandwidth	±4	-6/+4	-6/+8	±4	-6/+4	-6/+8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	149	201	233	149	201	233
R-squared	0.035	0.040	0.027	0.022	0.023	0.024

Notes: Construction includes improvements and repairs of houses, apartments, and bridges. OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±10 range in 2012.

Table 3.11: Impact on the Percentage of the Amount Spent by Type of Direct Negotiation in 2010-2011

	Cutoff between high and medium relative importance, medium risk								
Dependent variable	Unique supplier			Emergency			Trust in suppliers		
$I[X \geq 0]$	-0.028*	-0.021	-0.021	-0.002	-0.016	-0.008	0.019*	0.011	0.018*
	(0.017)	(0.013)	(0.015)	(0.028)	(0.024)	(0.026)	(0.011)	(0.010)	(0.010)
Comparison mean	0.047	0.04	0.04	0.034	0.041	0.041	0.013	0.014	0.014
Bandwidth	±6	±8	-8/+6	±6	±8	-8/+6	±6	±8	-8/+6
Linear spline	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	233	313	296	233	313	296	233	313	296
R-squared	0.020	0.013	0.013	0.024	0.024	0.026	0.014	0.005	0.011
Dependent variable	Disproportionate cost			Cost less than 720 USD			Other		
$I[X \geq 0]$	-0.002	-0.001	-0.002	0.013	0.015*	0.015	0.039	0.038	0.028
	(0.001)	(0.002)	(0.002)	(0.011)	(0.009)	(0.010)	(0.033)	(0.027)	(0.030)
Comparison mean	0.001	0.001	0.001	0.006	0.004	0.004	0.025	0.036	0.036
Bandwidth	±6	±8	-8/+6	±6	±8	-8/+6	±6	±8	-8/+6
Linear spline	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	233	313	296	233	313	296	233	313	296
R-squared	0.020	0.003	0.006	0.008	0.010	0.010	0.034	0.023	0.027

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±10 range in 2012.

Table 3.12: Impact on the Percentage of the Amount Spent by Type of Direct Negotiation in 2012

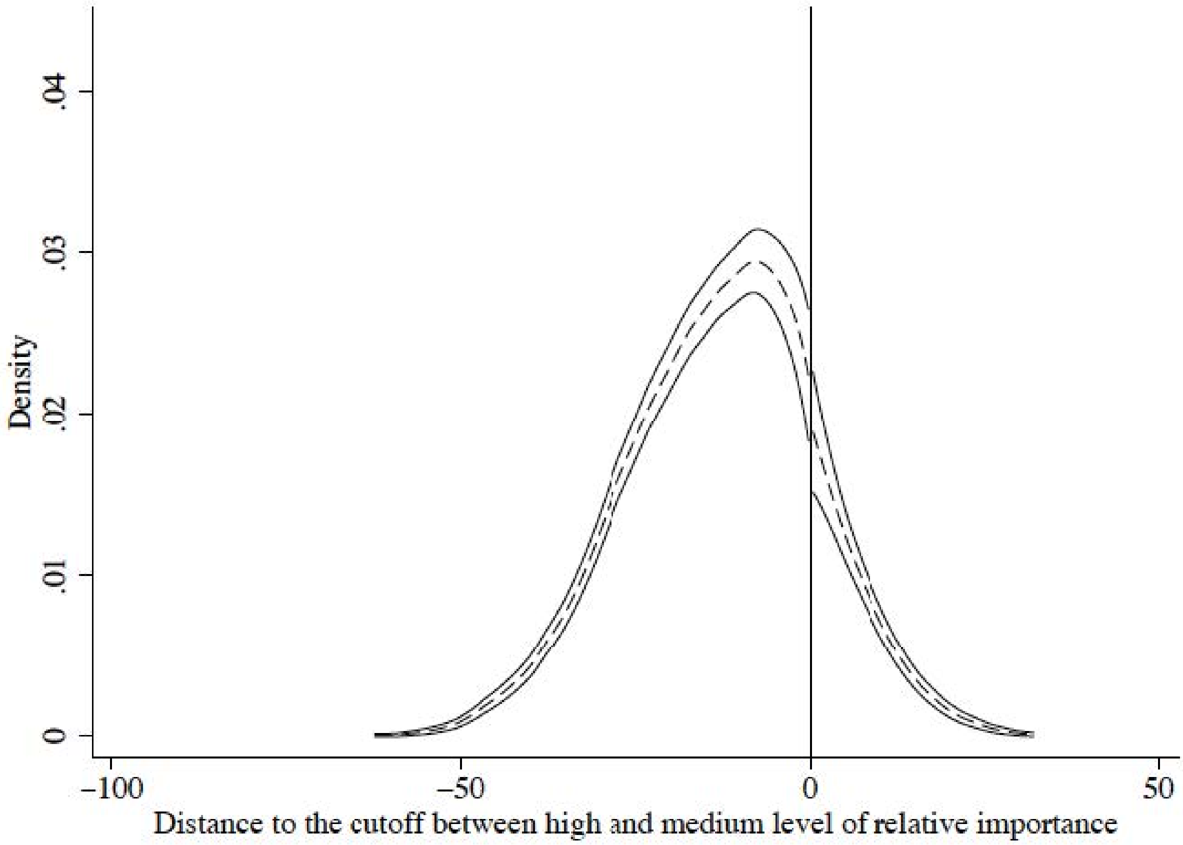
	Cutoff between high and medium relative importance, medium risk								
Dependent variable	Unique supplier			Emergency			Trust in suppliers		
$I[X \geq 0]$	-0.005	-0.010	-0.009	0.066***	0.039*	0.038*	0.020*	0.017	0.018*
	(0.019)	(0.014)	(0.012)	(0.024)	(0.021)	(0.022)	(0.012)	(0.011)	(0.010)
Comparison mean	0.035	0.04	0.039	0.014	0.012	0.013	0.008	0.009	0.009
Bandwidth	±4	±6	-8/+6	±4	±6	-8/+6	±4	±6	-8/+6
Linear spline	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	161	234	297	161	234	297	161	234	297
R-squared	0.007	0.015	0.016	0.0061	0.017	0.012	0.018	0.011	0.012
Dependent variable	Disproportionate cost			Cost less than 720 USD			Other		
$I[X \geq 0]$	0.001	-0.007	-0.006	-0.004	0.001	0.002	0.030	0.061	0.036
	(0.003)	(0.004)	(0.004)	(0.007)	(0.005)	(0.005)	(0.044)	(0.045)	(0.040)
Comparison mean	0.002	0.003	0.003	0.013	0.011	0.009	0.061	0.020	0.046
Bandwidth	±4	±6	-8/+6	±4	±6	-8/+6	±4	±6	-8/+6
Linear spline	✓	✓	✓	✓	✓	✓	✓	✓	✓
Observations	161	234	297	161	234	297	161	234	297
R-squared	0.016	0.029	0.021	0.007	0.002	0.002	0.020	0.028	0.016

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±10 range in 2012.

Appendix A3

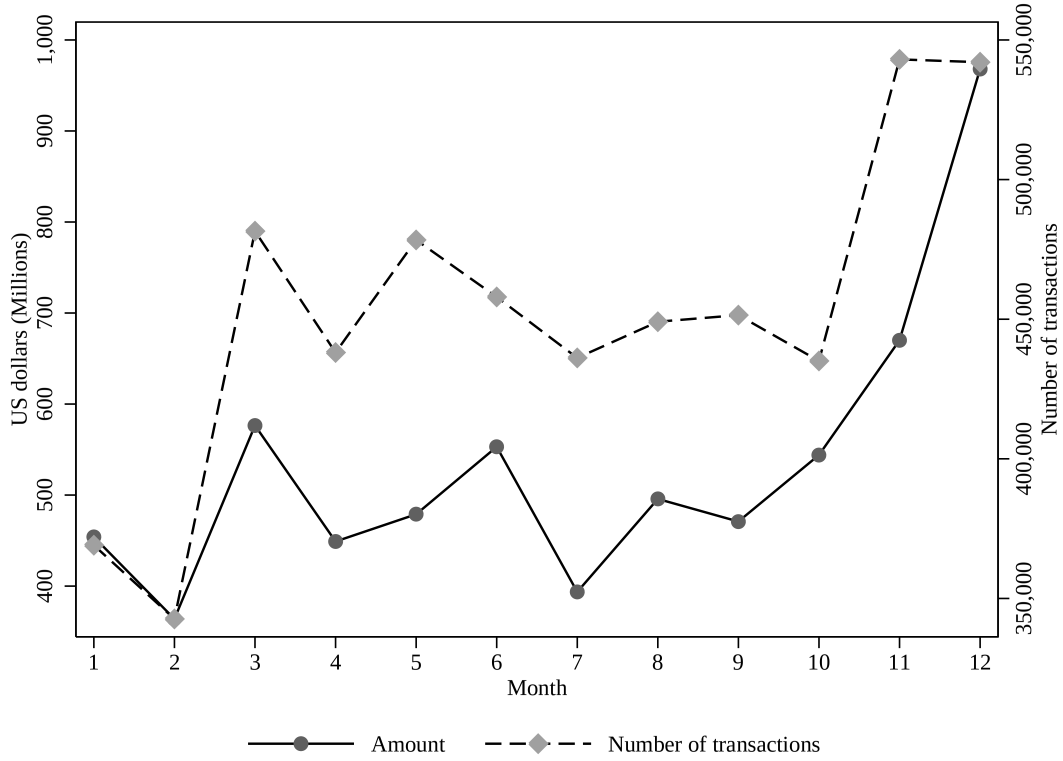
Figures and Tables

Figure A3.1 Density Test, Full Sample



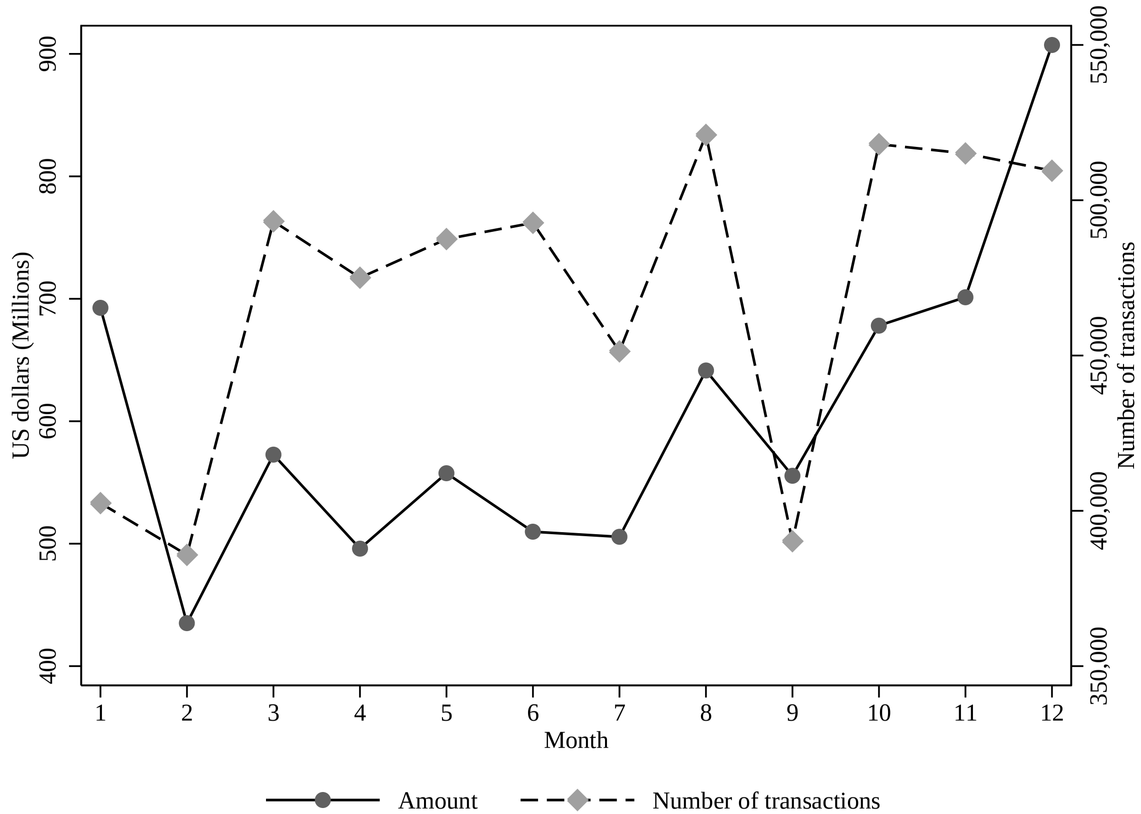
Notes: Dashed line indicates density estimate, and solid lines indicate 95% confidence interval. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of all entities in 2012.

Figure A3.2: Total Amount of Purchases and Number of Transactions by Month in 2011



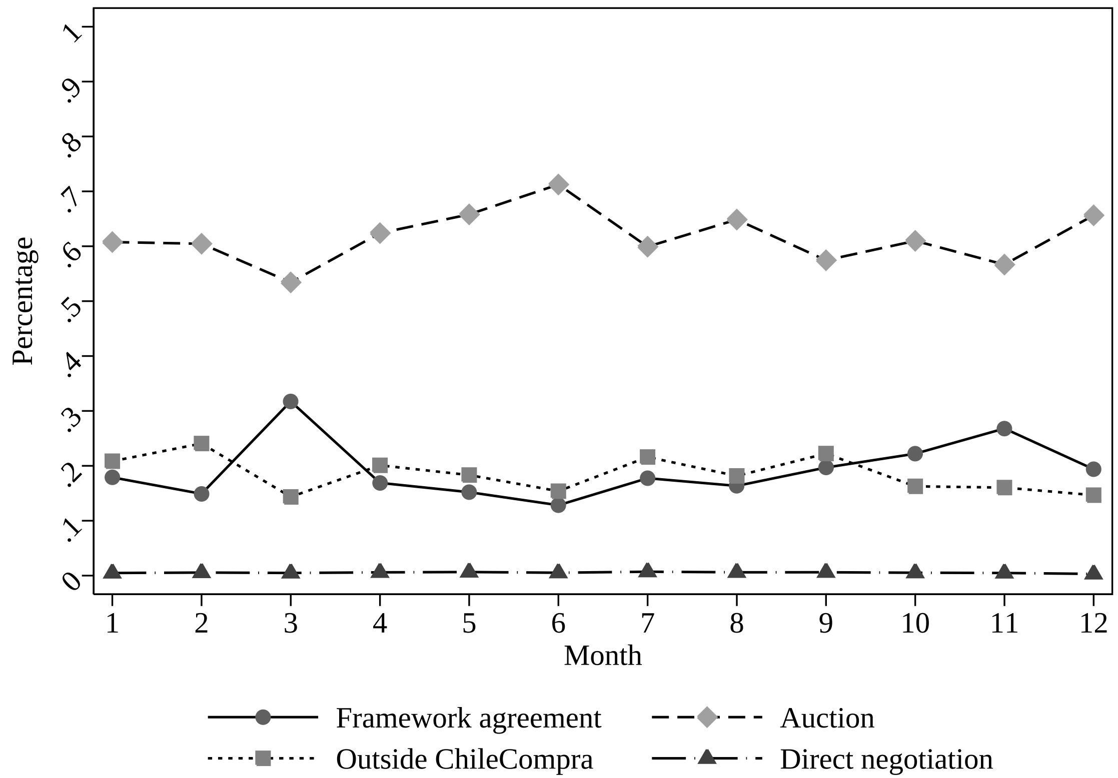
Notes: This figure is based on the universe of transactions done in 2011. Each transaction corresponds to a separate item or service.

Figure A3.3: Total Amount of Purchases and Number of Transactions by Month in 2012



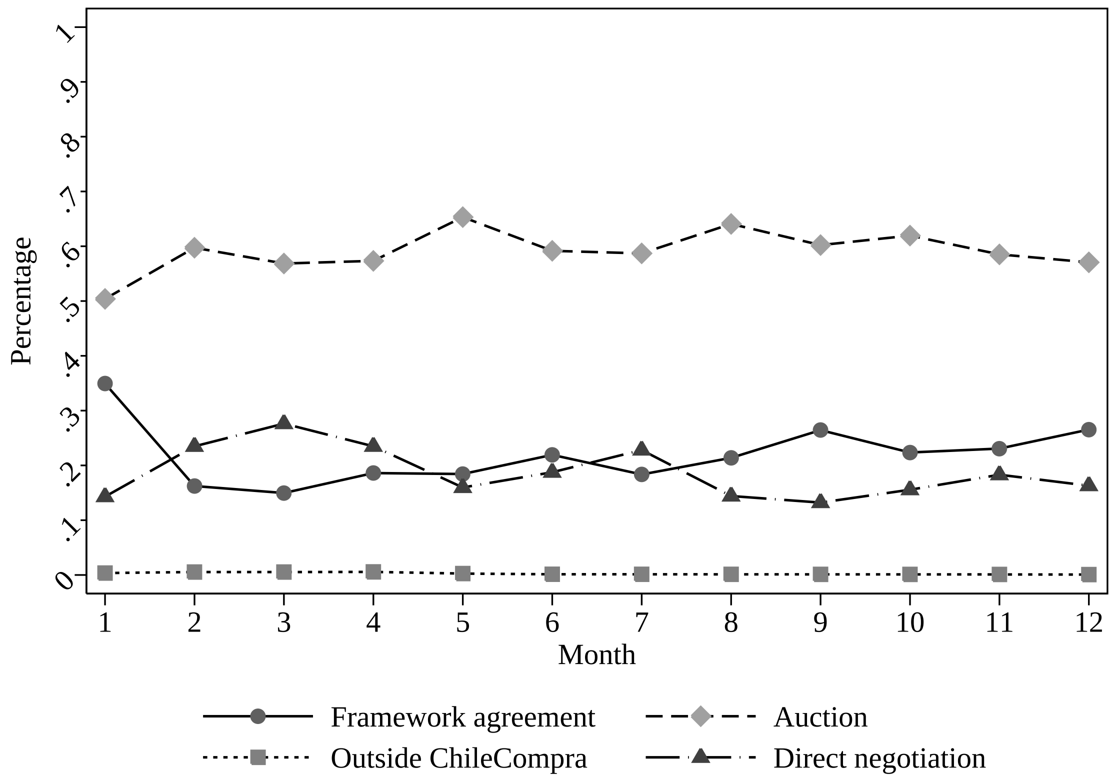
Notes: This figure is based on the universe of transactions done in 2012. Each transaction corresponds to a separate item or service.

Figure A3.4: Share of the Amount by Modality and Month in 2011



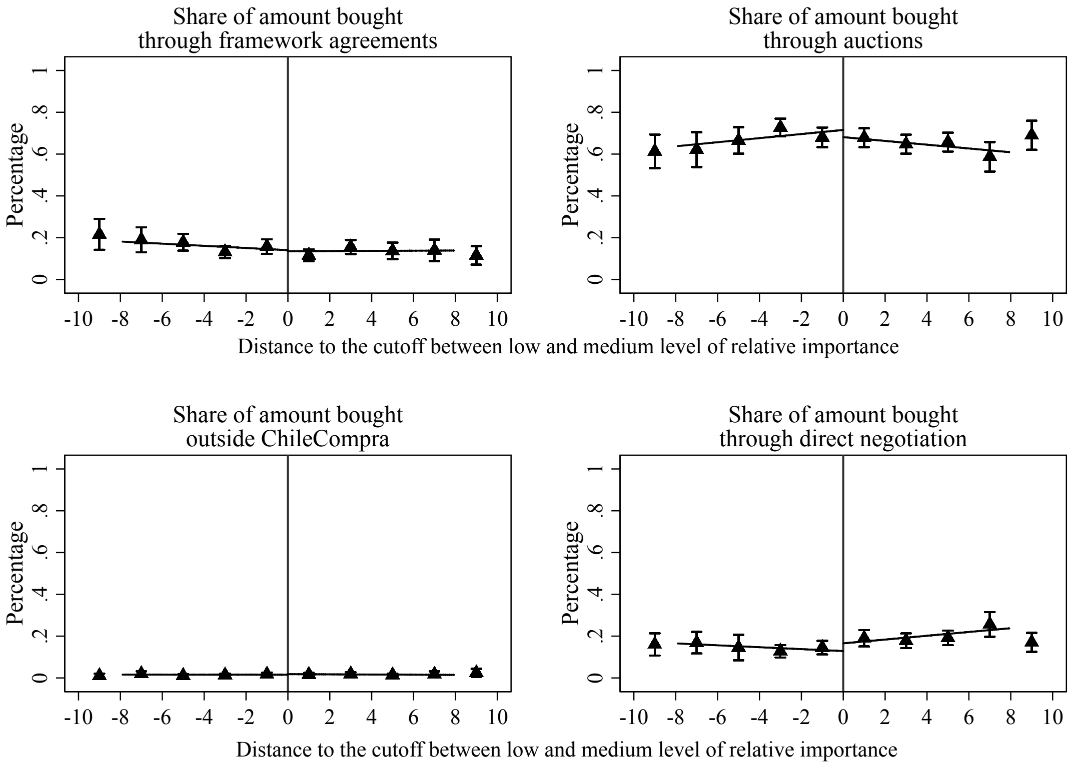
Notes: This figure is based on the universe of transactions done in 2011.

Figure A3.5: Share of the Amount by Modality and Month in 2012



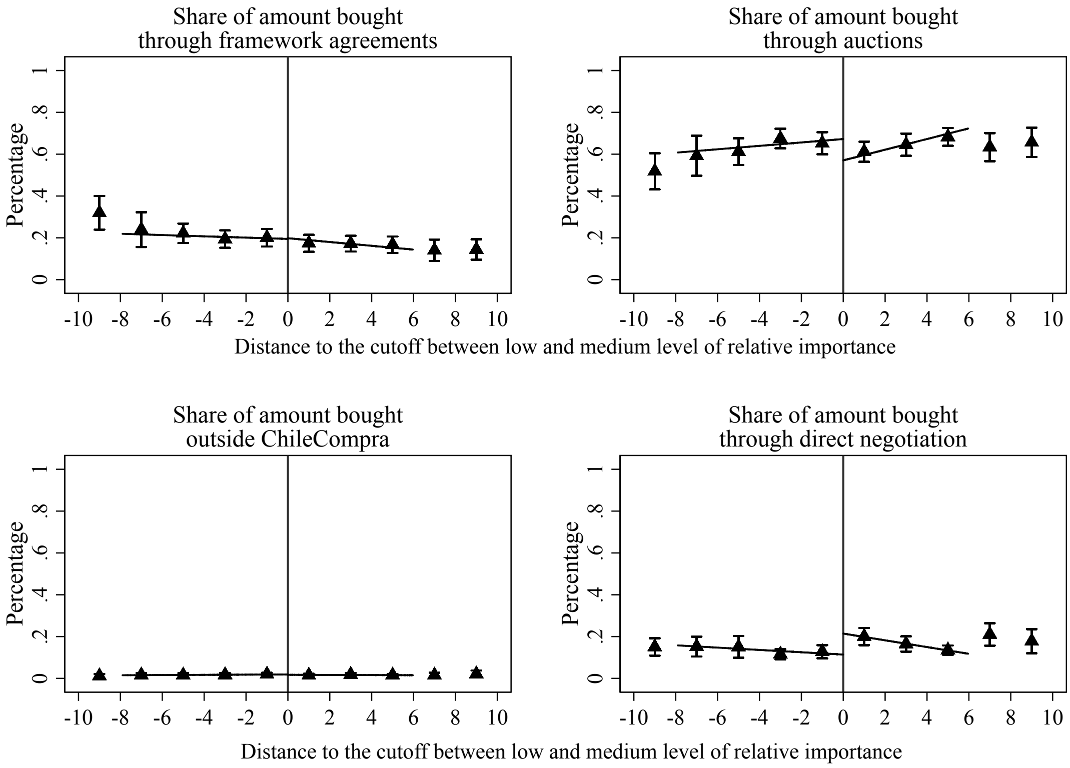
Notes: This figure is based on the universe of transactions done in 2012.

Figure A3.6: Impact on the Share of the Amount by Modality of Purchase in 2009-2010 (using Sample 2011)



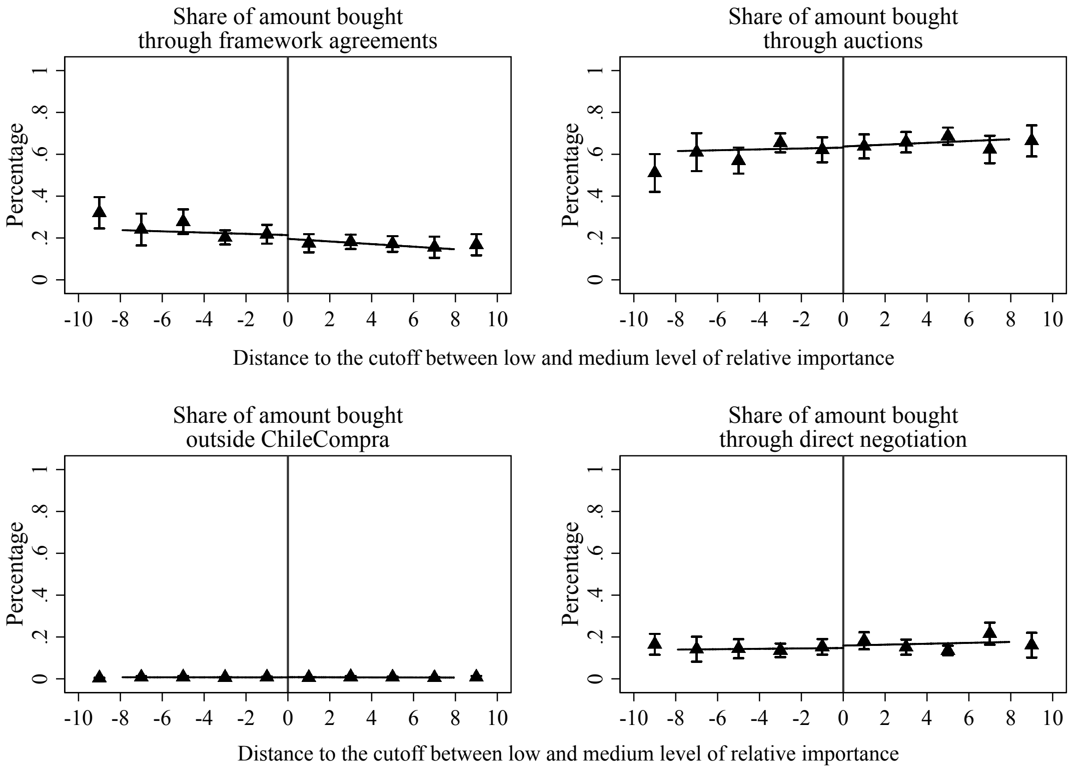
Notes: Solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2011.

Figure A3.7: Impact on the Share of the Amount by Modality of Purchase in 2011 (using Sample 2011)



Notes: Solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2011.

Figure A3.8: Impact on the Share of the Amount by Modality of Purchase in 2012 (using Sample 2011)



Notes: Solid lines show linear fit. Triangles represent share of the amount purchased through a given modality in a given bin, and bars represent 95% confidence intervals. The relative importance score for each entity is normalized by the respective year- and region-specific cutoff. Sample consists of entities with normalized importance scores within the ±10 range in 2011.

Table A3.1: Summary Statistics, Full Sample

Panel A: Auctions						
Type of Auction	Number of Auctions	Share of Auctions	Mean Number of Bids	Mean Number of Bidders	Mean Duration	
Open Auction, <7,200 USD	1,182,682	76.5%	5.94	5.80	40.13	
Open Auction, 7,200-72,000 USD	280,347	18.1%	4.74	4.60	52.24	
Open Auction, >72,000 USD	77,490	5.0%	3.88	3.60	84.01	
Restricted Auction	4,741	0.3%	1.54	1.48	15.95	
Other	85	0.0%	3.22	2.64	197.48	
Panel B: Purchase Orders						
Source of Purchase Order	Number of Orders	Share of Orders	Number of Items	Share of Items	Amount in Millions of USD	Share of Dollar Total
Framework Agreement	659,546	31.3%	2,088,030	35.7%	\$1,700	22.6%
Auction	925,635	44.0%	2,629,370	45.0%	\$4,440	59.0%
Direct Negotiation	377,744	17.9%	833,590	14.3%	\$1,3800	18.3%
Outside ChileCompra	142,150	6.8%	293,279	5.0%	\$18	0.2%
Panel C: Audits						
Public Procurement Focus	Number of Audits	Mean Duration	Median Duration			
High	380	87	56			
High-Medium	513	86	59			
All	633	85	57			

Notes: Sample consists of entire universe of public entities in 2012.

Table A3.2: Impact on the Share of the Amount by Modality of Purchase in 2009-2010 (using 2011 Sample)

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
$\mathrm{I}[X \geq 0]$	-0.055*	-0.018	0.005	0.013	-0.023	-0.034
	(0.033)	(0.029)	(0.026)	(0.047)	(0.039)	(0.036)
Comparison mean	0.169	0.153	0.140	0.657	0.686	0.715
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	320	444	536	320	444	536
R-squared	0.009	0.002	0.006	0.018	0.008	0.016
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
$\mathrm{I}[X \geq 0]$	0.000	-0.000	0.003	0.042	0.042	0.037
	(0.008)	(0.006)	(0.006)	(0.036)	(0.030)	(0.028)
Comparison mean	0.021	0.019	0.015	0.154	0.143	0.129
Bandwidth	±4	±6	±8	±4	±6	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	320	444	536	320	444	536
R-squared	0.005	0.004	0.001	0.026	0.023	0.036

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±8 range in 2011.

Table A3.3: Impact on the Share of the Amount by Modality of Purchase in 2011 (using 2011 Sample)

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
I[$X \geq 0$]	-0.016	0.001	0.003	-0.056	-0.090**	-0.102**
	(0.043)	(0.034)	(0.033)	(0.053)	(0.042)	(0.040)
Comparison mean	0.209	0.197	0.195	0.630	0.660	0.672
Bandwidth	±4	±6	-8/+6	±4	-6/+6	-8/+6
Linear spline	✓	✓	✓	✓	✓	✓
Observations	321	445	479	321	445	479
R-squared	0.007	0.011	0.012	0.018	0.021	0.021
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
I[$X \geq 0$]	-0.003	-0.002	-0.002	0.075**	0.091***	0.100***
	(0.008)	(0.006)	(0.006)	(0.035)	(0.028)	(0.026)
Comparison mean	0.022	0.020	0.019	0.139	0.123	0.114
Bandwidth	±4	±6	-8/+6	±4	-6/+6	-8/+6
Linear spline	✓	✓	✓	✓	✓	✓
Observations	321	445	479	321	445	479
R-squared	0.002	0.001	0.001	0.047	0.035	0.034

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±8 range in 2011.

Table A3.4: Impact on the Share of the Amount by Modality of Purchase in 2012 (using 2011 Sample)

	Cutoff between high and medium relative importance, medium risk					
Dependent variable	Share of the amount bought through framework agreements			Share of the amount bought through auctions		
$\mathrm{I}[X \geq 0]$	-0.034	-0.009	-0.018	0.000	0.000	-0.005
	(0.042)	(0.032)	(0.030)	(0.056)	(0.041)	(0.039)
Comparison mean	0.225	0.205	0.214	0.608	0.637	0.632
Bandwidth	±4	-6/+8	±8	±4	-6/+8	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	321	503	537	321	503	537
R-squared	0.010	0.024	0.023	0.007	0.005	0.005
Dependent variable	Share of the amount bought outside ChileCompra			Share of the amount bought through direct negotiation		
$\mathrm{I}[X \geq 0]$	-0.002	0.000	0.001	0.036	0.009	0.012
	(0.003)	(0.003)	(0.003)	(0.039)	(0.029)	(0.027)
Comparison mean	0.009	0.008	0.007	0.157	0.150	0.147
Bandwidth	±4	-6/+8	±8	±4	-6/+8	±8
Linear spline	✓	✓	✓	✓	✓	✓
Observations	321	503	537	321	503	537
R-squared	0.003	0.001	0.001	0.009	0.005	0.005

Notes: OLS estimations. The relative importance score X for each entity is normalized by the respective year- and region-specific cutoff. The largest sample consists of entities with normalized importance scores within the ±8 range in 2011.