ECONOMIC STUDIES

DEPARTMENT OF ECONOMICS
SCHOOL OF BUSINESS, ECONOMICS AND LAW
UNIVERSITY OF GOTHENBURG
246

The Economics of Coercive Institutions, Conflict, and Development

Melissa Rubio Ramos

UNIVERSITY OF GOTHENBURG

ISBN 978-91-88199-51-5 (printed)
ISBN 978-91-88199-52-2 (pdf)
ISSN 1651-4289 (printed)
ISSN 1651-4297 (online)

Printed in Sweden,
Gothenburg University 2020

To Santiago

Acknowledgments

I do not have enough words to express my gratitude to all who have encouraged me, challenged me, and supported me during this chapter of my life. Now it is time to continue learning and become a better researcher.

My deepest gratitude goes to my supervisors Randi Hjalmarsson and Ola Olsson for their support and guidance throughout these five years. Randi, this dissertation would not have been possible without your invaluable advise, feedback and challenge. Thanks for believing in me during hard times. I was very happy of having you as my mentor. Ola thanks for your continuous encouragement and valuable comments. Anna Bindler thank you for your rigorous and friendly discussions.

To my coauthor, William Maloney, thanks for always encouraging me follow through with my research and being my mentor for many years.

I am greatly indebted to all the colleagues at the Department of Economics. Thanks to the labor and development groups. Andreea Mitrut, Ariel Pihl, Nadine Ketel, Paul Muller and Måns Söderbom provided valuable comments to my papers. Thank you Mikael Lindahl for always taking time to discuss my research ideas. This dissertation benefited from my visit to George Washington University. I am particularly grateful to Bryan Stuart for hosting me. I would also like to thank Marion Dupoux, Inge van Den Bijgaart, Elisabet Olme, Li Chen, Gustav Kjellsson, Annika Lindskog and Joe Vecci for being supportive.

I was fortunate to share my PhD experience from the very beginning with great colleagues and friends. Simon Schürz thanks for always supporting me. Tamás, Maks, Debbie, Sebastian, Samson, Eyoual, Tewodros, Anh, Ida, Anna, Carolin, Josephine, Laura, Verena, Simon F, Marcela J, David, Magnus, Dominik, Vu, Jakob, Ruiji, Lina, Louise, Lisa, Andrea, and Po, thank you for all the good conversations. I am indebted to Åsa Löfgren, Selma Oliveira, Elizabeth Földi, Mona Jönefors, Ann-Christin Räätäri Nyström, Katarina Forsberg, Maria Siirak, Margareta Ransgård and Marie Andersson for helping me with administrative issues.

I am grateful to my friends in Gothenburg. Zoli, thanks for always for being there for me. Imelda, Lorena, Juliana, Gabriel, Juan Pablo, Diana, Anna K, and Walter thank you to all of you. I am grateful to the Uni Andes group of friends and economists that supported me and took good care of me during my visits to DC. To my friends and family back in Colombia, thanks for lifting my spirits up during my visits while it was winter in Sweden.

Finally, I would like to thank my father, Hector, and my siblings, Gabriela, Vanessa, and Hector for their visits to Sweden and their unconditional support and love. To my beloved Mother, Marilu, I would wish you could have seen me graduate.

I will always be grateful!

Melissa Rubio
Gothenburg, May 2020

Introduction

This dissertation is a compilation of three papers that put together my research interests on the effects of institutions on development outcomes with inequality as the connecting thread. These institutions include slavery in the United States and (and its abolition), and conflict in developing countries. These research questions are relevant both in modern day developing countries and, from an economic history perspective, developed countries. This dissertation is motivated by both the scientific interest in understanding how individuals are affected by institutions, and how the answers to these questions inform us about social injustices and inequality in the world both today and historically.

The first chapter concerns racial inequality. The relative social disadvantage of African Americans is one of the most profound and enduring characteristics of U.S. society. For any given dimension of socioeconomic well-being, one is quite likely to find relatively poor outcomes for Blacks (Raphael, 2002). Blacks are considerably less likely to participate in the labor force than whites; they earn less per hour than whites; and they suffer unemployment rates consistently twice the national average (Bayer & Charles, 2018). The justice system is no exception. African Americans are more likely to face longer prison terms than whites for the same crimes (Rehavi & Starr, 2014), and they are overrepresented in the prison population. Black males constitute 6.5% of the US population but account for 40% of the prison population (FJS, 2013). While empirical research focuses on explaining the contemporary disparities between racial groups in the criminal justice system, we know very little about the historical roots of this race-based gap.

Thus, the first chapter of my dissertation explores the role of a fundamental part of American history that could have shaped the large racial disparities in the justice system -the slave-based labor system that prevailed in the United States until 1865. Although, slavery has long attracted the attention of social scientists, this is the first empirical attempt to study the legacy of this institution on the origins of the race gap in incarceration. In doing that, I rely on historical datasets from US census records from 1860 to 1940. I document a substantial increase in black incarceration immediately after the abolition of slavery, with no comparable effects on whites, and that this black-white incarceration gap continued to grow. I have also transcribed novel historical data on prison work camps from the Department of Labor to provide evidence that the high levels of black incarceration in the US started, at least in part, due to labor scarcity in which convict labor was used to replace slave labor. This mechanism is supported in an analysis of three natural experiments that reduced the demand for labor.

The second chapter studies the institution of conflict. Most conflicts around the world take place in poor countries (Collier *et al.* , 2009), and as the literature shows, the conflicts are costly themselves. Their consequences are immediate and direct in terms of deaths, injuries, and the destruction of infrastructure (Collier & Hoeffler, 2004). However, in addition to the losses in human and physical capital, I show that conflict also affects one of the most important assets in developing countries -social capital. In countries with weak institutions, social capital not only provides support during adverse situations (Foster & Rosenzweig, 2001; Fafchamps & Lund, 2003), but it also guarantees a more efficient provision of public goods (Nannicini *et al.* , 2013; Glennerster *et al.* , 2013), and better outcomes in terms of fiscal capacity (Guiso *et al.* , 2004), governance (Aghion *et al.* , 2010), trade (Cassar *et al.* , 2013), and the diffusion of knowledge and technologies (Conley & Udry, 2010; Bandiera & Rasul, 2006; BenYishay & Mobarak, 2014). Thus, if such outcomes are relevant for economic development, understanding the possible linkages between conflict and social capital is necessary. To estimate causal effects, I study the case of the Colombian conflict, and exploiting changes in violence attributed to cross-border shocks on coca markets in neighboring countries, interacted with a novel index of suitability for coca cultivation. I find that conflict has a negative effect on social cohesion measures such as trust, participation in community organizations, and cooperation.

The final chapter (with William Maloney) overlaps my interests in development, labor economics and inequality. The motivating factor for writing this paper is that the distribution of income is often seen as the key variable in the economics of inequality. The World Bank produces different income indicators that allow us to compare welfare across countries. However, that is only part of the story. The income risk that individuals face during their lifetime should enter into the inequality discussions as well. We argue that not taking into account risk underestimates the traditional measures of inequality. Take, for instance, two countries: Honduras and the US. With traditional income measures, Honduras would appear more equal, just because its population is very young; but, once income dynamics are taken into account, inequality could be higher. However, to analyze income dynamics, panel data are required, and that is not available for most developing countries. Therefore, we propose a method that allows us to measure labor income risk from repeated cross-sections. We find that in poorer countries, workers face higher levels of risk. Finally, we map our measures of risk into an inequality measure -the Theil index-, and we show that if Latin American countries would have the risk levels of the U.S., inequality would decrease.

References

Aghion, Philippe, Algan, Yann, Cahuc, Pierre, & Shleifer, Andrei. 2010. Regulation and Distrust. *The Quarterly Journal of Economics*, **125**(3), 1015–1049.

Bandiera, Oriana, & Rasul, Imran. 2006. Social Networks and Technology Adoption in Northern Mozambique. *The Economic Journal*, **116**(514), 869–902.

Bayer, Patrick, & Charles, Kerwin Kofi. 2018. Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940. *The Quarterly Journal of Economics*, **133**(3), 1459–1501.

BenYishay, Ariel, & Mobarak, A. Mushfiq. 2014 (May). *Social Learning and Communication*. Working Paper 20139. National Bureau of Economic Research.

Cassar, Alessandra, Grosjean, Pauline, & Whitt, Sam. 2013. Legacies of Violence: Trust and Market Development. *Journal of Economic Growth*, **18**(3), 285–318.

Collier, Paul, & Hoeffler, Anke. 2004. Greed and grievance in civil war. *Oxford Economic Papers*, **56**(4), 563–595.

Collier, Paul, Hoeffler, Anke, & Rohner, Dominic. 2009. Beyond Greed and Grievance: Feasibility and Civil War. *Oxford Economic Papers*, **61**(1), 1.

Conley, Timothy G., & Udry, Christopher R. 2010. Learning about a New Technology: Pineapple in Ghana. *American Economic Review*, **100**(1), 35–69.

Fafchamps, Marcel, & Lund, Susan. 2003. Risk-Sharing Networks in Rural Philippines. *Journal of Development Economics*, **71**(2), 261 – 287.

FJS. 2013. Federal Justice Statistics - Statistical Tables. *International Centre for Prison Studies*.

Foster, Andrew D., & Rosenzweig, Mark R. 2001. Imperfect Commitment, Altruism, And The Family: Evidence From Transfer Behavior In Low-Income Rural Areas. *The Review of Economics and Statistics*, **83**(3), 389–407.

Glennerster, Rachel, Miguel, Edward, & Rothenberg, Alexander D. 2013. Collective Action in Diverse Sierra Leone Communities. *The Economic Journal*, **123**(568), 285–316.

Guiso, Luigi, Sapienza, Paola, & Zingales, Luigi. 2004. The Role of Social Capital in Financial Development. *American Economic Review*, **94**(3), 526–556.

Nannicini, Tommaso, Stella, Andrea, Tabellini, Guido, & Troiano, Ugo. 2013. Social Capital and Political Accountability. *American Economic Journal: Economic Policy*, **5**(2), 222–50.

Raphael, Steven. 2002. Anatomy of The Anatomy of Racial Inequality. *Journal of Economic Literature*, **40**(4), 1202–1214.

Rehavi, M. Marit, & Starr, Sonja B. 2014. Racial Disparity in Federal Criminal Sentences. *Journal of Political Economy*, **122**(6), 1320–1354.

Chapter 1: From Plantations to Prisons: The Legacy of Slavery on Black Incarceration in the US

Chapter 2: The Effect of Violence on Social Capital: Evidence from Exogenous Variation in an Illegal Market

Chapter 3: Comparing Labor Income Risk Across Countries

Chapter I

From Plantations to Prisons: The Legacy of Slavery on Black Incarceration in the US

Melissa Rubio*

Abstract

Black males constitute 6.5% of the US population but account for 40% of the prison population. The extent to which this disparity reflects differences in criminal conduct and socio-economic background, as opposed to differential treatment is a long-standing question. However, little is known about the roots of this disparity. This paper uses US decennial censuses for the period 1850 to 1940 to show that the race gap in incarceration can be traced back to the abolition of slavery in 1865. In particular, I exploit the variation in the prevalence of slavery across counties in southern states to estimate the short- and long-run impact of slavery on black incarceration rates. I document a substantial increase in black incarceration immediately after the abolition of slavery, with no comparable effect on whites, and that this black-white incarceration gap continued to grow. These baseline OLS results are not driven by omitted variables given their robustness to: (i) observable controls, which proxy for racial attitudes and socioeconomic and geographic characteristics, (ii) analyses of neighboring counties that are more likely to be comparable on unobservable dimensions, and (iii) an IV analysis that instruments for slavery intensity with a county's suitability for growing cotton. Using novel historical data on prison work camps from the Department of Labor, I provide evidence that the high levels of black incarceration in the US started, at least in part, due to labor scarcity in which convict labor was used to replace slave labor. This mechanism is further supported in analyses of three natural experiments – land grant allocations, Boll Weevil cotton pests, and the Mississippi River floods – that reduced the demand for labor; these reverse shocks are associated with lower black incarceration rates.

*PhD student in Economics, University of Gothenburg. Email: melissa.rubio@economics.gu.se. Webpage: www.melissarubio.com. I am extremely grateful to Randi Hjalmarsson for her invaluable advice and guidance; I also thank Ola Olsson for his guidance. Special thanks to Bryan Stuart for hosting me at the George Washington, and to Mikael Lindahl, Anna Bindler, Matthew Lindquist, Erik Plug, William Maloney, Juan F. Vargas, Nadine Ketel, Sebastian Braun, Erik Hornung, and Moritz Schularick for inspiring conversations. This paper benefited from seminars at George Washington University, Aarhus University, Uppsala University, Cologne University, Bonn University, Bayreuth University and University of Gothenburg. Anh Vu Tran and Maxim Brüls provided excellent research assistance. All errors are mine.

"Slaves went free; stood a brief moment in the sun; then moved back again toward slavery" (Du Bois & Lewis, 1999).

1 Introduction

Relative to whites, black males in the United States are six times more likely to be incarcerated. They constitute 6.5% of the US population but account for 40% of the prison population (FJS, 2013). This racial disparity has been of increasing interest to economists. While empirical research focuses on explaining the contemporary disparities between racial groups in the criminal justice system, we know very little about the historical roots of this race-based gap.[1] This article explores the role of a fundamental part of American history that could have shaped the large racial disparities in the justice system. In particular, it tests whether the highly disproportionate representation of African Americans in the penal system today is the legacy of the slave-based labor system that prevailed in the United States until 1865.

The four million enslaved people at the beginning of the Civil War were an inexpensive workforce that made Southern agriculture immensely lucrative (Wright, 2013). Slaves were valuable assets and represented a significant share of Southern wealth.[2] However, the end of slavery via the Civil War devastated this economy. Ager *et al.* (2019) document that the emancipation of slaves represented one of the largest ever destructions of wealth in the US. Though agriculture faced enormous difficulties, including the loss of livestock, fences, and barns, the largest concern to farmers was the lack of a system to ensure an adequate supply of labor. Most planters had great difficulty in satisfying their demand for labor.[3] How was this demand for labor met? To date, evidence on the answer to this question is anecdotal. Historian David Oshinsky argues that white southerners took advantage of the 13^{th} Amendment, which authorized "slavery" or "involuntary servitude" as punishment for crimes. Black men were convicted for petty offenses and sent to plantations as convict labor. As a result, incarceration rates increased for African Americans.

This paper is the first to investigate the legacy of slavery on the penal system, and empirically evaluates whether the black-white incarceration gap can, at least in part, be attributed to the use of the justice system to replace the loss of manual labor upon the abolition of

[1]For instance, Rehavi & Starr (2014) document that black males tend to face longer prison terms (9 percent higher) than whites arrested for the same crimes, even after controlling by case and defendant characteristics. Anwar *et al.* (2012) find evidence that jury pools convict black defendants significantly (16 percentage points) more often than white defendants. But when the jury pool includes at least one black, conviction rates are almost identical.

[2]Slaves were financial assets. They allowed planters mobility by maintaining credit relationships across distances, and the payment of debs because of their liquid character. González *et al.* (2017) documents the role of slave wealth in business formation in Maryland during the Civil War.

[3]Some former slaver owners, those "who had dealt honorably and humanely towards their slaves," were able to retain many of their former field hands (Alston & Ferrie, 2007).

slavery. My analysis relies on a series of data sources, which can be matched by county. First, using individual full count census records, I measure the share of blacks and whites that were in prison at each census from 1860 to 1940. This allows me to describe the relationship between incarceration for blacks and whites for counties with differential intensities of slavery measured in the 1860 census. Second, I digitize novel historical data from prison records from the Department of Commerce, which complements the census records by identifying the type of correctional facility in which inmates were held. Third, I digitize records from the Department of Labor, which document the type of activity and profitability of all convict camps in the US.

My analysis relies on variation across counties in the intensity of slavery in 1860 just prior to the Civil War. The first goal of the paper is to estimate the effect of slavery intensity on the black incarceration rate immediately after the abolition of slavery (1865) and in each subsequent census year. The second goal is to understand the mechanisms underlying these effects. Demand side mechanisms include the use of prisoners to replace slaves as a labor force while a supply side story could be one in which former slaves actually commit more crimes in the face of poor economic conditions. In particular, I present empirical evidence in support of a demand side story.

Motivated by the possibility that slave-reliant counties in the South were systematically different from other counties, I control for pre-existing differences in 1860 that might be related to the development of slavery.[4] My baseline set of covariates includes controls for county size (in acres), population, average farm value per acre of improved land, total acres of improved land, presence of railways, presence of waterways, the proportion of small farms, a measure for ruggedness, the proportion of county population reported to be "free black" on the 1860 census, a measure of land inequality, and the percentage of votes for the democratic party in 1860. I find that slavery had substantial effects on subsequent incarceration rates in the US: a one standard deviation increase in the intensity of slavery is associated with an increase of 2.5 in the black incarceration rate per 1,000 population, or a 15% higher proportion of African Americans in the prison population, immediately after slavery is abolished. This persisted until 1940. Moreover, I find no evidence that the abolition of slavery increased the share of whites in prison.

The main threat to the validity of these results is that the measure of slavery spuriously captures the latent negative effects of places more reliant on slavery. For instance, slavery could also be an indication of conservative racial attitudes towards African Americans, and these attitudes could directly affect incarceration rates. Thus, in addition to controlling for a rich set of socio-economic variables, three additional steps are taken. First, it remains

[4]For instance, it could be the case that the South's more rural and wealthy areas were more likely to develop slavery-based economies and it is the persistent wealth of these areas, not the legacy of slavery, that drives the white-black incarceration gap.

possible that the results are driven by differences between slaveholding and non-slaveholding areas that are not captured by a number of historical and geographic covariates. For instance, it could be that the "upland" regions of northern Alabama and Georgia differed systematically from the Black Belt (as suggested by Kousser (2010)). Therefore, I restrict the sample to the set of neighboring counties that border a county with differing proportions of slaves. This enables me to compare the effects of slavery across counties that are geographically and perhaps also politically, economically and culturally similar (as Banerjee & Iyer (2005) do with Indian districts). Second, if the main effects are genuinely attributable to the prevalence of slavery, then there should no be differences in incarceration outcomes between areas in the South that were largely non-slaveholding and areas in the North that also did not have slaves. Furthermore, I perform a similar exercise by comparing counties right at the south-north border to have a better counterfactual. Finally, I use a cotton growing suitability index to capture a potentially exogenous variation in slavery. The idea behind this instrument is that slavery grew with the importance of cotton. Cotton plantations required specific climatic conditions that are arguably exogenous to the treatment of African Americans in the justice system, after controlling for socio-economic and geographic county characteristics. Importantly, I can replicate the baseline OLS results with this IV specification.

The final part of the paper links back to the original hypothesis, and provides empirical evidence of a demand side mechanism, in which convict labor was used to replace slave labor. Using manually transcribed and geocoded data on the types of prison institutions, I study the relationship between slavery intensity and the nature of post-Civil war prison institutions. I find that places that relied more on slavery, i.e in which there was a greater shock to the labor force, were significantly more likely to have post-Civil war convict camps. Moreover, these relationships are even stronger for those types of prison institutions that provided the most intensive labor: chain gangs, lumber prison camps, and farm prison camps. In addition, I show that the introduction of convict labor in the United States increased incarceration rates, especially among African Americans. Following the literature of labor coercion (Acemoglu & Wolitzky, 2011), I exploit exogenous reverse shocks to the demand for black labor: the introduction of agricultural stations in the south (Kantor & Whalley, 2019)[5], the boll weevil cotton pest (Clay *et al.* , 2018)[6], and the Mississippi River flood (Hornbeck & Naidu, 2014)[7]. The results indicate that black incarceration rates were lower in counties where workers were replaced with labor-saving technologies, and again I do not find evidence of an effect for white incarceration rates. My interpretation of these results is that the legacy of slavery in the penal system started at least in part as a way to secure cheap labor after slavery was abolished.

[5]The introduction of agricultural serves as exogenous variation in the location of agricultural knowledge. They allowed the diffusion of advanced practices that increased agricultural productivity.

[6]This pest adversely affected cotton production.

[7]Flooded counties experienced a mechanization in agricultural practices.

A second possible mechanism is a supply side mechanism in which the widespread poverty and low education of African Americans immediately after the Civil War pushed them to disproportionally commit crime (à la Becker (1968)). This paper provides a wide range of empirical evidence in support of the demand side mechanism. Though I cannot empirically rule out the supply side, and indeed it may very well be that both mechanisms were important, this paper makes the first systematic evaluation of this question.[8]

This paper makes important contributions to several strands of the economic literature. First, it relates to the institutional origins of the black-white incarceration gap. Different studies have documented the continued impact of slavery on economic inequality. Bertocchi & Dimico (2014) show that the transmission channel from slavery to racial inequality is through human capital accumulation. Mitchener & McLean (2003) show that slavery had a strong and persistent effect on productivity levels, measured as income per worker across the US in the 1880-1980 period. Lagerlöf (2009) documents a negative relationship between slavery and current income at the county level for a sample of former slave states.[9] However, this is the first paper that relates slavery to the subsequent treatment of blacks in the judicial system.

Second, the estimates presented here are consistent with the view that coercion is more likely when labor is scarce. This idea is along the lines of Naidu & Yuchtman (2013), who show that criminal prosecutions for contract breaches and unemployment move in opposite directions across nineteenth-century Britain. Contrary to this, North (1971) argues that coercive relations started to decline when labor became scarce following the Black Death and other demographic shocks that reduced the population. Previous contributions with regard to the US slavery experience are based on sporadic anecdotal evidence. Here, I show empirically how the shortage in cheap labor led to incentives for incarcerating more African Americans.

Finally, this paper contributes to the literature on racial discrimination in the judicial system. There is an increasing number of papers studying the causes of the disproportionately high black-white incarceration gap. In contrast, the root of this phenomenon has been

[8]Using data at state level, I show that African Americans were more likely to be charged for "non-sense" crimes as vagrancy and crimes against the good morals (loud talking, being out at night). However, there are only 14 states in the South, so this does not allow me to test this possibility more rigorously.

[9]Other related evidence from outside the US studies the legacy of slavery. For instance, Nunn (2008) shows that those African countries that exported the most slaves are comparatively poorer today. Nunn & Wantchekon (2011) show a negative relationship between an individual's reported trust and the number of slaves taken from the individual's ethnic group during the slave trades. In Brazil, Valencia *et al.* (2011) and Soares *et al.* (2012) document a strong relationship between slavery and modern levels of inequality. In Peru, Dell (2010) shows how the Mita colonial system of forced mining in Peru and Bolivia continues to have negative impacts today. Buonanno & Vargas (2019) investigate the long-run effects of slavery on economic inequality and crime in Colombian municipalities. Finally, Markevich & Zhuravskaya (2018) look at the immediate effect of abolishing serfdom on the Russian Empire.

relatively understudied, and therefore this paper analyzes the racial gap from the time when a large pool of blacks was "entitled" to be incarcerated. A recent exception includes Eriksson (2019), who shows that Blacks that migrated to the north during the 1920s were more likely to end up in prison.

The remainder of the paper is organized as follows. Section 2 describes the historical institution of slavery and its abolition. Section 3 presents the data. Section 4 traces out the OLS estimated impact of slavery on the black-white incarceration gap. Section 5 provides extensive evidence regarding the robustness of the results to omitted variables, including an IV analysis. Section 6 turns to the mechanisms, using information on convict camps and reverse shocks to the demand for labor to provide empirical evidence in support of a labor demand mechanisml. Section 7 concludes.

2 Background

2.1 Slavery in the US

Slavery was introduced in the US in the 17^{th} century by the British, and served to recruit and regulate the unfree workforce forcibly imported from Africa and West Indies. However, slavery rapidly disappeared in the Northern states, while slaves were the main labor force in southern cotton plantations. This institution became so crucial that historian Barbara Fields has written, slavery was "the central organizing principle of society" in the South (Fields, 1982, p.143). Economics and politics were dominated by the southern elite – plantation owners with large land and slave holdings (Wright, 1978). Slaveholding was reserved for the top echelon of white households, with an even smaller minority owning a large plantation. In 1860, 21 percent of white southern households owned at least one slave and 0.5 percent owned 50 or more slaves (Soltow, 1975; Table 5.3). Larger plantations took advantage of economies of scale to achieve efficient production. Fogel & Engerman (1974, p.203) describe the slave workforce on large plantations as "rigidly organized as in a factory," with teams separated by task and following an "assembly line" structure from plowing to planting (Metzer, 1975; Fogel and Engerman, 1977; Toman, 2005). Slaves provided a low-skilled agricultural labor force, which made cotton growing so profitable that the number of slaves increased from 700,000 in 1790 to 4 million in 1860. They represented 13% of the US population and were distributed across 15 slave states, mostly in the South. By the same year, almost 90% of all blacks living in the US were slaves. (Wright, 2013, p.69) estimates that almost half of the total wealth held by whites were slaves. In addition, cotton accounted for half of the value of all American exports before the Civil War, and helped spur Northern industry (Davis, 2006, p.184). Slave prices increased steadily from 1850 to 1860, betraying no signs that market participants anticipated the coming emancipation.

2.2 The Abolition of Slavery (1865)

Enslaved people throughout the South were freed over the course of the Civil War. Outside of the District of Columbia, southerners were not compensated for the forfeiture of their slave wealth.[10] Therefore, the emancipation proclamation stripped slave owners of their slaves and the market value of these assets. The Confederacy's defeat in the Civil War and the formal abolition of slavery in 1865 led to one of the largest compressions of wealth inequality in human history. As one Georgia planter bemoaned in 1866, "by our defeat, we have lost [...] millions in the emancipation of our slaves, we have virtually lost [everything]" (Bryant, 1996, p. 113). Although few southerners had their lands confiscated, land holdings also substantially declined in value, particularly in cotton-growing areas that had been dependent on slave agriculture. Taken together, the wealth held by white southerners fell by 38 percent at the median and by 75 percent at the 95th percentile from 1860 to 1870 (Ager *et al.* , 2019). Much of this loss came from plantation farms (Wright, 2013).[11]

As a result, the need to secure cheap labor from previously enslaved blacks was most dire for plantation owners living in areas that had high slave concentrations. This demand for cheap labor now had to be negotiated with freed men. However, after the Civil War and in the absence of cash or an independent credit system, sharecropping and tenant farming emerged quickly as an alternative system (by 1870, hiring wage workers was very rare). Sharecropping was a system where the landlord/planter allows a tenant to use the land in exchange for a share of the crop. Nevertheless, there were liability problems, high interest rates, unpredictable harvests, and unscrupulous landlords that kept black tenants severely indebted. Laws favoring landowners made it difficult, or even illegal, for sharecroppers to sell their crops to others besides their landlord, or prevented sharecroppers from moving if they were indebted to their landlord. In addition, most of the lynchings were directed towards African Americans looking to purchase land, which was seen by many at this time as being important for economic independence (Acharya *et al.* , 2016).

With the ratification of the Thirteenth Amendment in 1865, which abolished slavery throughout the nation, the penal laws of southern states became applicable to all offenders regardless of race. The 13^{th} amendment explicitly authorized "slavery" or "involuntary servitude" as a punishment for crime, leaving the "white elite class" free to reintroduce forms

[10]The cost of national emancipation through compensation, rather than through war, would have been very high; the estimated value of all slave wealth was $2.7 billion in 1860, more than 50 percent of the annual GDP (Goldin, 1973). In other parts of the Americas, the abolition of slavery compensated slave owners with cash or labor time. In other cases, the abolition was gradual, so slave owners did not face a dramatic shock to their wealth (Acemoglu & Robinson, 2008).

[11]Income per capita of the South fell to about 50% of the U.S. Income per capita remained at about half the average until the 1940s when it finally began slowly to converge (Wright, 1986), pg 70. While the North developed large manufacturing sectors, the South remained primarily agricultural. The South had very low rates of urbanization (around 9% as opposed to 35% in the Northeast) and relatively little investment in infrastructure. For example, the density of railroads (miles of track divided by land area) was three times higher in Northern than Southern states.

of slavery. Historians suggest that, as a response, white elites established local laws and institutions with the purpose of securing cheap labor to sustain the cotton economy –the so-called "Black Codes". Former slave states enforced these codes, many of which were enacted shortly after abolition in 1866. They were designed to control the mobilityof free black menand restrict economic opportunities of freed black men. One subset of these laws criminalized vagrancy, which made it illegal to loiter or appear out of work without written evidence of it. The failure to have such "lawful employment" was punishable by arrest and imprisonment. These codes also allowed prisons to lease out their inmates as low-cost laborers to local farms (Naidu, 2010). Furthermore, blacks were excluded from juries, and endured extreme punishment for small crimes (Acharya *et al.* , 2016). These sentences often included hefty sums that blacks simply could not pay. Anecdotes suggest that many African Americans were randomly captured by rural whites, who falsely accused them of falling to pay their debts. They then used the court system to extract labor under a system called "peonage", or debt bondage, in which prisoners were "leased out" by the state to private farmers or companies guaranteeing, in this sense, the provision of black labor. It is this anecdotal evidence that I empirically test in the mechanism section.

The US Commissioner of Labor (1885-1905) claims that prison labor was by far less expensive than other sorts of labor (Department of Labor, 1887, 1906, 1925). Poyker (2019) estimates that the cost of prison labor was just 19% of the cost of free labor. This is consistent with the theoretical framework of Robinson & Acemoglu (2008), who argue that the Southern landed elites exercised de facto political power to compensate for the loss of their de jure political power, and therefore they invested in alternative mechanisms to maintain control.[12]

2.3 Imprisonment

Imprisonment was not a suitable punishment for blacks in the antebellum South because it would have deprived the owner plantation of the labor of his slave (Sellin, 1976). Rather, the antebellum penitentiaries of slave states were meant to confine criminals from the master class. For instance, in 1850, in the Alabama penitentiary, there were 167 white males, 1 white female, and 4 free colored persons. Local jails were a place for pre-trial slave detention, or to house runaway slaves until their owners could be located. Instead, slave-owners legitimized their domestic disciplinary violence and protected their property rights. Because slaves owned no property and had no ability to pay fines, for instance, corporal punishment (whipping) was the most common penalty (Sellin, 1976).

[12]Ager (2013) finds that the southern elite used the facto power (as proxied by pre-war relative wealth) to maintain their economic and political status after the Civil War.

2.4 After the Abolition of Slavery

The end of the Civil War saw an increase in correctional facilities and the prison population. The southern penal system consisted of three types of institutions: state prison buildings, which resembled those in the North, the county chain gang, and the state prison farm. They were built within plantations, near coal mines and pine forests where turpentine was extracted, and close to rail-roads. In addition, punishment in the post-Civil War era also included a county system of hiring out vagrants and petty offenders to local farmers.

Together with this, convict labor was introduced during the Reconstruction period (1865-1877) when the US government was trying to revive the economy of the former Confederate states. Convict labor also spread to the Northern states.[13] By the end of the presidential term of Rutherford Hayes (1877-1881), this system was introduced in almost all states (Wines, 1871) and was very profitable. The larggest prison-farms were located in Texas, Arkansas, Louisiana, and Mississippi. Nearly 250,000 acres of land in the United States were under cultivation by convicts. Texas alone had 83,407 acres farmed by prisoners, raising products that were valued in 1927 at $1,362,958 . Louisiana had an income from its prison system of $1,557,715. This income from the forced labor of prisoners helped to keep down tax rates on the big plantations (Wilson, 1933). There was also a great deal of construction work done by convicts for government institutions (Garret, 1929).

Systems of employing prisoners

The most important systems are the "contract", the "state account", "state use", "public work", and "lease". The "contract" system was one of the oldest systems to be introduced. As early as 1867, prison contractors were flourishing in all prisons. Under this system, a private business man or a firm contracted with the state for the use of a certain number of convicts. The private contractor then set up machinery in the prison and provided tools and materials. The state fed, sheltersed and guarded the prisoners for the contractor, who sold the products made by the convicts in the market. In the "state account" system, the state went into this practice on its own. There is anecdotal evidence of the state settting up dummy companies to market goods for them (Wilson, 1933, p 39). Under the "state use" system, convict-made goods were not sold in open markets but consumed in the state's institutions. Under the "public work" system, convicts were used in construction or repair work, such as roads. Finally, the "lease" was the most used convict system in the US. It worked by renting or hiring convicts out entirely to the custody of a private business or company.[14] The contractor had complete authority to guard, feed, discipline and exploit convicts. This system grew after the Civil War. Prior to that time, convicts in the South were white workers. But after the abolition of slavery, the prison population rapidly became

[13]In prisons in southern states, working times were between 12 and 16 hours, whereas in northern states and in other parts of the country, the day's work was frequently eight or nine hours (Wilson, 1933).

[14]A detailed list of companies and business that engaged in hiring convict labor can be found at the Convict Labor Records from the US Bureau of Labor Statistics.

black. African Americans were convicted of minor crimes and hired out to contractors. As a result some historians claim that this system was a move by the ruling class to secure forced labor on a large scale as a partial substitute for chattel slavery (Wilson, 1933, p 40).

The chain gang, one of the penal institutions of the South, was the most brutal type of convict force labor in the United States. Historians argue that the chain gang was one of the devices consciously developed by the former slaveholders to put the newly "freed" African Americans back into bondage. Consistent with this, one of the qualifications to get a job as a guard was to know how to handle "Negroes" (Wilson, 1933, p 72). In addition to this, the convict system provided monetary incentives to the police and judicial system (Sharkey & Patterson, 1933). There is anecdotal evidence that police would "round up idle blacks in times of labor scarcity" and that sheriffs were directly asked to arrest more people before the cotton harvest season (Oshinsky, 1997; Cohen, 1976).[15]

The majority of the convict population was black, about 85 to 90 percent. Convict labor peaked around 1880, as it was used to supply labor to farming, railroads, mining and the timber industry. By 1886, 70% of the prisoners were working as convict-laborers (45,277 of the nation's 64,349).[16] Convict leasing persisted in various forms until it was abolished by Franklin Roosevelt in 1941.[17]

3 Data

3.1 Data Sources and Sample

The main analysis sample includes all counties that belonged to the 14 former Confederate states. I focus on the Southern states because slavery was not allowed in the Northern states by 1860.[18] There are approximately 1,000 counties included in the sample overall; however, there is some variation in this number across censuses because some counties divided over

[15] The sheriff and court officials in many states were pay per arrest and conviction. For instance, in 1929, the sheriff of Bolivar county, Mississippi received $24,350.70, which was a cotton producer county, while other sheriff received $20,000 a year (Wilson, 1933). The *The New York Times* for example, wrote in September 26, 1931:
LITTLE ROCK, ARK.—Police action to force unemployed men to help pick this year bounteous cotton crop to-day had extended from Helena, in Eastern Arkansas, to Bowie County, Texas, on the southwestern border. Helena and Phillips County officers already have started a drive to get cotton pickers to the fields by threats of vagrancy charges and Bowie officials to-day said a similar campaign would start the next Monday. Cotton planters in various sections of the State have complained that they were unable to obtain an adequate number of pickers, despite an unusually large number of unemployed persons. They attributed the situation to the prevailing low rate of 30 to 40 cents per hundred pounds being paid to pcikers, but said a higher price could not be paid because of the low price of cotton. Several truckloads of Negroes were captured and sent out to the cotton fields. The sheriff and other officers followed to see that none escaped.

[16] 15,100 were engaged in prison duties, and 3,972 were sick or idle.

[17] The state of Virginia never imposed a convict leasing system. Tennessee was the first state to officially abandon it in 1893 while Alabama was the last in 1928 (Poyker, 2019).

[18] One of the reasons to focus on the Southern states is that slavery was not allowed in the Northern states. The proportion of black males in the population in the South was 29%, whereas in the North it was 2,5%.

time.

To study the effect of slavery on black incarceration, I combine data from several sources. First, I use official decennial Census records from the *Integrated Public Use Microdata Series* (IPUMS) spanning the period 1850-1940 to calculate the number of prisoners by race in each county.[19]

This information is complemented with historical prison archives on the location (state and county), number of prisoners, race of the prisoners and type of correctional facility (prison, jail, workhouse, or chain gang). These come from the *Department of Commerce's* "Crime, Pauperism, and Benevolence" report for the years 1880, 1890, 1904, and 1910. Additional official data are taken from the *Department of Labor.* As competition between convict labor and free labor was a widely discussed topic at that time, the Bureau of Labor decided to inspect all penitentiary facilities to determine the level of competition between goods produced under convict labor and goods produced by free workers.[20] The data include all prisons, houses of correction, and convict labor camps, as well as juvenile reformatories and industrial schools, and allows me to identify the presence of convict labor in a correctional facility. I use all of the available reports for the following years: 1886, 1895, 1905, and 1923. Then I matched all prisons and convict labor camps by name and location to their corresponding county in 1860. Thus I can establish the relationship between slavery and the presence of these convict camps. I do this by assigning GPS coordinates and then county FIPS codes for each of them.[21] The coordinates allocation was performed with Google APIs Maps.[22] By using a Python program, Google Maps automatically looks for every address and assigns coordinates at the county level; in 3% of the cases, Google found more than two results for a place. The main reason is that those places do not exist anymore, so coordinates were allocated manually. Overall the dataset contains 460 different correctional facilities for every year for which data is available. Appendix Figures A1 and A2 include excerpts of these data sources. The next subsections describe in detail the construction of the main variables.

3.2 Construction of Variables

Imprisonment data. I use the full universe of prisoners from the 1860 to the 1940 Census from the Integrated Public Use Microdata Series (IPUMS-USA) database. Following Eriksson (2019) and Lochner & Moretti (2004), prisoners are identified using two variables. First, I use the group quarters type of residence in the Census, which indicates if the individual is in a correctional facility. Second, I only count individuals reporting a relationship

[19]IPUMS collects, preserves and harmonizes U.S. census micro-data. Data can be requested here. The completed census forms for 1890 were lost in a fire thus data is unavailable for this census year

[20]The data was collected by the Bureau of Labor employees who traveled directly to prisons and filled out the surveys according to the accounting books provided by prisons. Only the data for the 1895 report was obtained not in person but through mail: prison warders filled out the survey themselves.

[21]County FIPS codes is five-digit code which uniquely identifies counties in the United States.

[22]Google APIS is a set of application programming interfaces (APIs).

to the household head as "Prisoner" or "Inmate". I rule out guards by using a variable on occupation.[23] I create a dummy variable equal to 1 if the respondent is a prisoner in a correctional institution. I aggregate the data to the race and county level to construct the share of African Americans in the prison population in county c and census year t relative to the number of blacks total in county c and census year t:

$$Black\ Incarceration_{ct} = \frac{Blacks\ incarcerated_{ct}}{Black\ population_{ct}}$$

For robustness purposes, I also calculate alternative measures for black incarceration including the share of African Americans in the prison population, and the share of African Americans in the prison population relative to their share in the total population.

Slavery. To measure slavery, I use the proportion of each county's 1860 population that was enslaved, measured by the 1860 Census. This measure represents the last record before slavery was abolished in 1865.Figure 1 shows that there is considerable variation in the intensity of slavery. Darker shaded counties were more reliant on slavery. Slavery spread from Virginia to Mississippi, in what scholars call the Black Belt, and alongside the Mississippi river. In the average Southern county, 36.7% of the population was enslaved in 1860, with a minimum value of 2%, and a maximum value of 92%. There was also substantial variation within states in the prevalence of slavery. For instance, in Benton County, in the northwest corner of Arkansas, 4.1% of the population was enslaved, whereas in Chicot County, in the southeast corner of Arkansas, 81.4 % of the population was enslaved. By the 1860 Census, there were approximately 4 million slaves. In particular, I use the number of slaves in the 1860 Census, and I divide it by the total population in that county:

$$ShareSlaves_c{}^{1860} = \frac{Number\ of\ Slaves_c^{1860}}{Total\ population_c^{1860}}$$

Controls. All county-level data controls are taken from the Inter-University Consortium for Political and Social Research (ICPSR), specifically from the Historical, Demographic, Economic and Social Data (ICPSR 2896)[24]. I use the following variables as controls: county size (in acres), population, average farm value per acre of improved land, total acres of improved land, presence of railways, presence of waterways, the proportion of small farms, a measure for ruggedness, the proportion of county population reported to be "free black" on 1860 census, a measure of land inequality, and the percentage of votes for the Democratic party in 1860.

3.3 Descriptive Statistics

Table 1 gives an overview of the key variables. Panel A shows that 34% of the Southern population was enslaved by 1860, and 1.1% of the black population was free. For the entire

[23]The average prisoner to staff ratio was 11, with Arkansas and Louisiana having the highest ratios of 40 and 35, respectively.

[24]These data have been used by others in well-known publications in Economics. To see the complete list of papers using this dataset please go here.

studied period (1860-1940), African Americans were overrepresented in the prison population. They were 26% of the population in the US south, but accounted for 43% of the total inmate population. In the average county, incarceration for African Americans was 5.79 per 1,000 population, while the same ratio for whites was around two. The average county reported 62 prisoners; the range was from 0 to 5763 (which I scale by population in the empirical analysis). Panel B shows that in about of 7,6% of the counties, there was prison; 64,2% had a jail; 27% had a chain gang, 1,1% a mining prison; and 6,7% a railroad prison. Panel C shows that in about 25% of counties, there was a railroad, and 34% were located next to a river.

Figure 2 demonstrates the evolution of the black and white incarceration rates from 1850 to 1940. The blue solid line represents the black incarceration rate, while the green dashed line represents the white incarceration rate. African Americans were incarcerated at a rate of 1.10 per 1,000 population, while for whites the rate was 0.49 in 1880, and this gap expanded to 5.8 and 3.2, respectively in 1940.

Furthermore, Figure 3 shows that there was not only a racial gap at the extensive margin, but also at the intensive margin. Using an 1890 census component of "Crime and Pauperism", I show that African Americans received longer sentences for the same type of crimes compared to whites. Unfortunately, I do not have micro-data to see how this gap evolves over time, nor how it is related to the prevalence of slavery, as the data is only available at the state, and not the county level. Ignoring the intensive margin implies that the interpretation of the results presented in this paper may actually be a lower bound of the effect of slavery on the race gap in incarceration.

4 Slavery and Post-abolition Incarceration Gap

In this section, I first present the empirical strategy for tracing out the impact of slavery on the black and white incarceration rates after slavery was abolished. I then discuss the main results, and show that the results are robust to alternative measures of black incarceration.

4.1 Empirical Strategy

I start by documenting the correlation between slavery measured in 1860 and black incarceration after the abolition of slavery. To do that, I estimate the following equation for every census year[25]:

$$Y_{cs} = \beta ShareSlaves_{cs}^{1860} + \mathbf{X}_{cs}^{1860}\gamma + \psi_s + \epsilon_{cs} \tag{1}$$

Y_{cs} represents the various measures for black incarceration in the US at county level c

[25] An alternative specification includes the pooled censuses with census fixed effects.

and state s . $ShareSlaves_{cs}^{1860}$ denotes the share of slaves in a county c and state s in 1860 (5 years before the slavery abolition). I am using the intensity of slavery in 1860 in each country regardless of whether the county split into smaller counties in later year, a robustness check I estimate my results by using counties that did not split over time. $\mathbf{X}_{cs}^{1860}$ is a vector of control variables measured in 1860. In particular, I control for factors that may correlate with slave intensity in 1860. First, since wealthier and larger counties may have relied differently on slave labor, I control for economic indicators. These controls include county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets, and therefore ithey could have influenced the slave force used in different counties.

To account for the possibility that counties may have had different norms about race, I use different proxies for antebellum attitudes of whites towards slavery. Since comprehensive data on racial views are not available in the antebellum period, I instead look for measures that might be consequences of such attitudes. The first is the percentage of votes for the Democratic Party in 1860. At this time, the Democratic Party was the racially conservative party while the Republican Party was the racially progressive party, which could have affected the subsequent treatment of African Americans in the justice system.[26] Second, I include a measure for the relative mortality of slaves to whites. In particular, I use the natural log of the ratio of the slave mortality rate to the white mortality rate. Negative racial attitudes could have led white planters and farmers to increase the mortality of slaves, either directly through violence or indirectly through overwork, undernourishment, and poor medical care. Third, I use the average occupant size of slave quarters in farms as a proxy for slave treatment. Across the South, the average slave quarters housed around five individuals, though this number varied considerably. The idea is that planters with more extreme negative racial attitudes might provide less housing for their slaves, which would be measured as a higer occupancy in the average slave dwelling. I also control for the proportion of the free black population before slavery was abolished. Finally, I control for state fixed effects, which capture differences in treatment to African Americans across states that can be attributed to different laws in the justice system.[27] Standard errors are clustered at the state level.[28]

[26] Slavery was abolished under the first Republican President of the US - Abraham Lincoln. By around 1950, the Democratic party moved towards a civil rights platform (Acharya *et al.* , 2016).

[27] For instance, the Black Codes in the Southern States were restrictive laws designated to limit the freedom of African Americans. Some states required blacks to sign yearly labor contracts; if they refused, they risked being arrested. Mississipi and South Caroline were the frist states to enact the first black codes. Mississippi's law required blacks to have written evidence of employment for the coming year each January. In South Carolina, a law prohibited blacks from holding any occupation other than farmer or servant unless they paid an annual tax of \$10 to \$100. In both states, blacks were given heavy penalties for vagrancy, including forced plantation labor in some cases (Du Bois & Lewis, 1999).

[28] There are only 14 states in my analysis. For this reason I also present bootstrapped standard errors in the next section.

The main assumption in my identification strategy is that after including a series of control variables, I am able to take into account the possible unobserved variables that might be related both to the intensity of slavery and black incarceration rates and that could lead to biased estimators. However, it remains possible that low and high slavery areas differ in unobservable characteristics that affect both the prevalence of slavery and the treatment of African Americans in the justice system. I address this concern in Section 5 through a number of robustness checks that make the 'treated' and 'control' counties more comparable. In addition, I use an Instrumental Variable approach that exploits exogenous variation in slavery intensity driven by the county's suitability for growing cotton.

4.2 Baseline OLS Results

Figure 4 shows that after the abolition of slavery, there was an increased and persistent effect on black incarceration rates. The figure displays the OLS estimates of equation (1), where I estimate the effect of slavery separately for each census. Estimates are indistinguishable from zero for 1870, indicating that slavery effects on black incarceration rates did not appear five years after the abolition of slavery. This is consistent with the Reconstruction period (1865-1877), in which attempts were made to redress the inequities of slavery and its political, social, and economic legacy.[29] However, effects from slavery on black incarceration rates emerge in 1880, with these effects being statistically significant and persistent until 1940.

The effect is quantitatively important, as the point estimate in 1880 implies that going from a county with zero slavery to a county where the entire population was enslaved increases the black incarceration rate by 11.5. In other words, the increase in one standard deviation in the prevalence of slavery results in 2.5 more blacks incarcerated per 1,000 population, with a mean of 1.5.[30] By 1940, the point estimates imply that an increase in one standard deviation in slavery increases the black incarceration rate by 8.7. Standard errors are clustered at state level. However, to account for the fact that there are only 14 states, and that this might be a concern for reliable inference, I follow Angrist & Pischke (2009) to estimate the main specifications with bootstrapped standard errors. The results remain significant (see Appendix Table A1).

Importantly, the estimates for whites are indistinguishable from zero for all years, except 1870, which provides evidence that not all individuals from counties that were highly reliant on slavery were sent to prison, and that there was selective enforcement towards African Americans. In 1870, we see an increase for white incarceration which is driven by

[29]For instance, The Freedmen's Bureau was created in 1865 to provide aid to 4,000,000 newly freed African Americans in their transition from slavery to freedom.

[30]Going from the county at the mean, Richmond (Virginia) to Madinson (Alabama), one stardard deviation above increases the black incarceration rates by 2.5.)

Confederate prisoners of the Civil War that supported slavery.[31] Furthermore, looking at the the point estimates for blacks and whites at the same time provides a piece of evidence to understand the type of unobservable characteristics that might be related to slavery and incarceration rates. For instance, if general unobserved economic conditions were driving the results, we would expect to see similar effects for blacks and whites, and this is not the case.

Now, looking at these results more closely, Table 2 presents the estimates underlying Figure 4 for black and white incarceration rates for each census year in a different panel. Column 1 reports the estimates for blacks without controls, while columns 2 and 3 include economic controls and proxies for attitudes towards blacks, respectively. The coefficient of interest is stable across specifications even after additional covariates are included, suggesting a small amount of selection on observables. In addition, it alleviates the concern that slavery might simply be proxying for geographical, economic, or political factors that continue affecting black incarceration.[32] In particular, the effect of slavery does not disappear while controlling for measures proxying for white attitudes towards African Americans, which suggests that antebellum racism is not driving the results. Columns 4-6 of Table 2 show that slavery did not have an effect on white incarceration rates. This indicates that there was not a generalized increase in incarceration in counties that were more reliant on slavery, but that the increase occurred only for African Americans.

In addition, I address the potential issue of spatial autocorrelation in the residuals that is presented by Morgan (2019). The main concern here is that neighboring places tend to have similar values of residuals, and this raises the question of whether the explanatory power of some persistence regressions might be a consequence of fitting spatial noise that reflects deep structural characteristics of slavery. In other words, pro-slavery counties are likely to be surrounded by other pro-slavery counties, and when looking at the legacy of slavery on another variable, it is likely that again neighbor will resemble neighbor, leading to correlations. The Moran statistic for the main specification is $z = 1.23$; thus it is not possible to reject the null hypothesis that there is zero spatial correlation.[33]

Finally, one natural concern regarding this historical period is migration. Does migration to the north confound these estimates? Migration of African Americans to the North started around 1915 prompted by the confluence of rising labor demand in northern factories

[31]More detailed information can be found in the "Records of the War Department Relating to Confederate Prisoners of War, 1861-1865" (NARA M598 at the National Archives), which is a collection consisting of 427 volumes. The records are of Confederate prisoners of war and political prisoners confined in Union prisons. They consist mainly of registers and lists of captured soldiers and civilians. The records contain information such as names, rank, unit or residence, dates of capture, deaths, and prisoners released.

[32]Section 5 presents a discussion on omitted variables by using the Altonji (2005)'s and Oster (2016)'s approaches. In addition, I present a series of exercises in which I aim to make treated and control counties more comparable.

[33]To calculate the Moran statistic I used 5 neighbors as in Morgan (2019).

during World War I and a temporary freeze on immigration from Europe, which encouraged northern employers to consider alternative sources of labor supply (Abramitzky & Boustan, 2017). Abramitzky & Boustan (2017) show that fathers employed in both low and high-skill positions were more likely to have sons who migrated to the north, as compared with fathers in mid-skill occupations. Unskilled workers had the strongest incentive to leave the South, where pay for low-skilled work was low. High-skilled black migrants may have been particularly motivated by the political and social freedoms available in the North. All of these together suggest that, if anything, the positive sample selection will attenuate the possibility that there was more crime because of economic conditions, and instead the results present a conservative estimation of the effect of slavery on incarceration rates. However, even if the Great Migration might be an issue for my analysis, it is worth emphasizing that the effects of slavery are seen in 1880 and persist over the next 50 years – i.e. even before migration. Moreover, a robustness check that replicates the main results using an alternative dependent variable – number blacks in prison relative to the size of the prison population -– confirms these findings. Migration would affect these dependent variables in different ways. To the extent that the size of the prison population is determined by the demand for labor – a mechanism I return to later – the alternative dependent variable should be less sensitive to migration. The baseline, however, would be more sensitive in the later years to the extent that the denominator is getting smaller. Figure 5 plots estimates where the dependent variable is the share of blacks and whites in the prison population. A similar pattern is observed: there is an increase in the share of blacks sent to prison relative to whites prisoners, but the effect starts to decrease after the Great Migration in 1915. Therefore, migration could therefore explain the increasing pattern in magnitudes in the post-migration periods. As a result, I am more confident about the estimates presented from 1880 to 1910 because during that period there was not an important out-migration of African Americans from Southern states to Northern states.[34]

5 Strategies to Deal with Potential Omitted Variable Bias

Even if I have included controls for economics characteristics and attitudes towards African Americans, it still remains possible that there are unobservable variables that are related both with the presence of slavery and incarceration rates that would bias the main results. For instance, counties that were more reliant on slavery could have had different institutions by which law enforcement affected differently incarceration rates. This subsection first provides evidence that the selection into unobservables is not a concern for the estimates presented in Section 4.2. Second, it presents an Instrumental Variable (IV) strategy, which instruments

[34]There is however evidence of migration within the South. I have estimated the main regressions with individuals that at the moment of the census still live in the county in which they were born. In doing this I lose about 30% of the observations. Results are still statistically significant but standard errors increase.

for 1860 slavery intensity with cotton suitability.

5.1 Counties that are Likely to be Similar on Unobservables

Comparing neighboring counties (in the South)

I restrict the sample to the set of neighboring counties that border a county in which the proportion of slaves differs by more than 20 to 50 percentage points. For instance, in column 1 of Table 3, I keep the sample of neighbors bordering counties with 20 percentage points difference in the intensity of slavery. Column 2 of Table 3 keeps counties for which the neighbors differ by more than 30 percentage points, and so on. This allows me to compare the effects of slavery across counties that are geographically and perhaps also politically, economically, and culturally similar given their proximity to each other. Estimates from Table 3 show that the results are robust to restricting the analysis to only these neighboring counties, even though this removes more than half of the counties in the original sample. For instance, when looking at column 1 of Table 3, the effect of slavery on black incarceration for counties that neighbor another county that differers in its intensity of slavery by 20% increases by 11% with respect to the baseline specification of Table 2.[35] Thus, even within fairly geographically concentrated areas, there are strong and statistically significant differences between counties with higher and lower concentrations of slaves.

Comparing neighboring counties along the North-South border

I next use exogenous variation in slavery driven by the "discontinuity" in the adoption of slavery along the "Mason-Dixon line" that divided slave and non-slave states. Slavery changed discretely at the boundary: on the northern side, blacks were free while on the southern side, blacks were slaves. The "Mason-Dixon" line was marked in 1767 by English surveyors to settle a dispute between The Penn and Calvert families for colonial property.[36] This discontinuity can be used to evaluate the effects of slavery in a Regression Discontinuity style design (RD). The RD approach requires an identifying assumption that all relevant characteristics besides treatment (i.e. slavery) must vary smoothly at the Mason-Dixon boundary. Table A2 assesses the plausibility of this assumption by showing county characteristics measured just outside the southern-northern division. Panel A shows that most of the geographic characteristics are statistically identical across the boundary, as are farm values and crops. Counties in the In contrast, the difference in democratic vote shares are statistically significant; hence it is important that they are controlled for. Furthermore, Table A2 , Panel B presents county characteristics measured in 1790, the first year for which demographic characteristics at the county level are available, and slavery was allowed in northern states. Again there were no discontinuities in observed characteristics along the border. Using the RD approach (i.e. restricting the sample to those counties on either side of the border), I

[35] The effects increase when looking at counties bordering counties that differ by more than 50% in the intensity of slavery. The reason is that there is a sharper discontinuity in the prevalence of slavery.

[36] See more about the "Mason-Dixon" line here.

estimate in Table 4 that slavery increases the proportion of African Americans in prison by 9% relative to the mean.

Falsification Test: Low Slave Southern Counties versus Northern Counties

If the effects that are estimated in Table 2 can genuinely be attributed to the local prevalence of slavery, then we should see no difference in the outcomes between areas in the South that were largely non-slaveholding and areas in other parts of the country that did not have slaves, such as counties in the North. In addition, if no such differences exist, then that would provide evidence against the story that it is the institutional legacy of slaveholding, rather than the local prevalence of slavery, that is driving the results. Making these comparisons with the North also enables the creation of a more appropriate counterfactual. I examine the differences between Southern counties with very few slaves and non-Southern counties with no slaves. To do this, I restrict the data to counties in slave states where fewer than 5% of the county population was enslaved.[37] Table 5 shows that there are no differences between the Southern counties and non-Southern counties as coefficients are not statistically significant. This provides evidence that the local prevalence of slavery, rather than state laws permitting the ownership of slaves, drives the results.[38]

Discussion about selection into observable variables

Table 2 showed that the results are barely affected when including control variables, and the main message is unchanged: slavery had an impact on the subsequent incarceration gap between blacks and whites. The stability in the coefficients across specifications suggests a relatively small amount of selection on observables. However, it is not impossible that a small amount of selection on unobservables could explain the whole effect. I explore this possibility following Altonji (2005)'s omitted variable approach. Roughly speaking, the smaller the difference between the coefficients with and without controls, the less the estimate is affected by selection on observables, and so the larger the selection on unobservables needs to be in order to explain away the entire effect of the variables of interest. This approach uses the degree of selection on observables as a guide to the degree of selection on the unobservables.[39] The value of the ratio indicates that selection on unobservables would need to be 9 times stronger than the selection on observables for the entire analysis period, which seems highly unlikely.[40]

[37] Results are robust to restricting the sample from 1% to 20% of the enslaved population.

[38] In northern states the percentage of the African American males was

[39] The Altonji ratio is calculated as $\beta_f/(\beta_r - \beta_f)$, where β_r corresponds to the coefficient without controls, and β_f is obtained when the full set of observable characteristics is controlled for. In my case, $\beta_f = 18.047$ and $\beta_r = 16.230$.

[40] More formally, the shift in the distribution of unobservables would have to be 9 times as large as the shift in observables to explain away the entire effect of slavery.

5.2 Instrumental Variable Approach: Cotton suitability

The previous sections demonstrated a close association between subsequent black incarceration and the prevalence of slavery. Whether or not this relationship should be interpreted as causal depends on whether there are unobserved characteristics of southern counties that affect both the prevalence of slavery and the treatment of African Americans in the justice system.[41] Section 5.1 provided a series of tests that demonstrate that such selection on unobservables is unlikely to be driving the results. This section takes an instrumental variable approach to address the same concern.

5.2.1 IV Estimation Strategy

I instrument for the prevalence of slavery in each county in 1860 with the suitability for growing cotton. Historical sources provide evidence that the evolution of slavery grew with the importance of cotton. For instance, slaves arrived first to Virginia, but were rapidly moved to more suitable climates for cotton production. Similarly, one of the reasons to acquire the Louisiana territory in 1803 was to get access to fertile land for growing cotton. Therefore, I measure soil suitability for cotton with the cross-county Food and Agriculture Organization (FAO)'s potential yield for this crop.[42] The first stage equation is as follows:

$$ShareSlaves_{cs}^{1860} = \alpha CottonSuitability_{cs} + \mathbf{X}_{cs}^{1860}\gamma + \psi_s + \epsilon_{cs} \quad (2)$$

where $CottonSuitability_{cs}$ is an index from 0 to 1 that indicates how good a county is for growing cotton, where 1 corresponds to counties that are very suitable for cotton production.

5.2.2 IV Assumptions

In this subsection, I discuss the identifying assumptions of the instrumental variable design, which uses cross-county variation in the suitability for growing cotton.

Instrument relevance: first stage

I start by documenting the relationship between suitability for growing cotton and slavery prevalence. Figure 6, is a flexible analog to the first stage in equation (2), plotting estimates from a local linear regression with a 95 percent confidence interval. Counties that were more suitable for cotton cultivation had higher levels of slavery, in particular, the prediction in the intensity of slavery is is monotonically increasing in the suitability for growing cotton.

[41] Simultaneity is not a concern here because slavery is measured in 1860, and incarceration rates start to be measured in 1870.

[42] The estimates are based on climate averages from 1961 to 1990. I omit suitability for other crops, such as tobacco, because they have no relationship with slavery conditional on cotton suitability. While these measures use data from the contemporary period. Most of the changes to the suitability between 1860 and 1960 were either uniform shifts across the entire region due to worldwide climate change or be unrelated to attitudes towards African Americans. Technological change in the production of cotton is not a worry as what changed was the yield per hectare and not the suitability for growing cotton.

Table 6 confirms that the relationship between cotton suitability and prevalence of slavery is strong. Each panel of the table corresponds to a different census, and has a slightly different number of observations due to the number of counties changing over time. The estimates are, however, similar. The fact that the results are robust to the inclusion of controls indicates the high quality of the instrument. First, it suggests that the first stage is not explained by other variables different than the instrument, which would bias the IV estimates towards the OLS estimates. Second, the inclusion of controls could pick up a small amount of the endogenous variation in slavery (as R^2 goes to 1), making the exclusion restriction invalid. Third, the instrument is not correlated with the controls, which would invalidate the conditional independence assumption.

In all cases, the coefficients have the expected sign and are significant. The coefficient in the second column, after including controls, indicates that counties that are more suitable for cotton production had a higher enslaved population. In particular, one standard deviation above the cotton suitability mean increases the share of slaves by 0.14, over a mean of 0.31. Robust (Montiel-Plueger) F-statistics, accounting for clustered residuals at the state level, are above the conventional threshold for weak instruments.[43]

Instrument validity

Although the exclusion restriction is not testable, I discuss its plausibility. This condition is violated if there are unobservable factors correlated with cotton suitability and the main outcome. An example might be that counties that are more suitable for growing cotton have different labor markets and this could directly affect incarceration rates. I start by performing a placebo test to provide evidence for the exclusion restriction. This tests aims to show that the only reason for which cotton suitability affected incarceration rates was through the intensity of slavery. Therefore, I separate counties with a positive intensity of slavery and counties with no slavery and I estimate a reduced form of the effect of cotton suitability on black incarceration rates for all censuses. The idea behind this test is that there should not be a correlation between cotton suitability and black incarceration rates for counties that did not have slavery. Such correlation would invalidate the exclusion restriction. Figure 7 plots cotton suitability (horizontal index) against the average of the black incarceration rate filtered by a set of observed characteristics as in equation (1) and state and time effects (vertical axis). The left panel shows counties characterized by a positive share of slaves in the population, while the right panel displays counties where none of the population was enslaved. The relationship is positive and highly significant across counties that experienced slavery, but is insignificant across those with zero slavery. Though not a formal test for the exclusion restriction, this falsification analysis suggests that suitability for cotton production has an effect on black incarceration only through the channel of slavery.

[43] The standard Stock-Yogo critical values for weak instruments are only valid under *i.i.d* assumptions on the residuals (Montiel-Olea & Pflueger (2013); Kleibergen & Paap (2006)).

Monotonicity

In this setting, the monotonicity assumption requires that counties with a low suitability for growing cotton that had slavery, would also have slave-holdings if they had a high suitability for growing cotton, and vice versa for non-slave counties. This assumption ensures that the 2SLS identifies the local average treatment effect (LATE), i.e. the average causal effect among the subgroup of counties who could have had a different slavery intensity if the need for slave-labor would have been different because of their conditions for growing cotton.[44]

One testable implication of this assumption is that the first stage estimates should be non-negative for any subsample. Table 7 shows that the first stage estimates are positive and significant for a wide range of subsamples, characterized by: geographic features, slave mortality rates, political preferences, and railroad presence. This is consistent with the monotonicity assumption.

Reduced form

Table 8 presents the reduced form estimates of the effect of the instrument, cotton suitability, on black and white incarceration rates; odd numbered columns have no controls and even columns include controls. The coefficients show that cotton suitability positively affects the black incarceration rate, but not that for whites. Estimates are roughly unchanged when including controls indicating that the effect is not driven by county characteristics correlated with slavery. Overall, it is reassuring to find an effect in the reduced form, as the intuition behind the instrumental variables is that differences in black incarceration are assumed to be accounted by differences in slavery given the cotton suitability.

5.2.3 2SLS Results

Figure 8 presents the IV estimates for blacks and whites for the entire period. The point estimates for African Americans are statistically significant after 1880, and they continue to grow until 1940. Again, one does not see an effect of slavery on white incarceration rates. Figure 9 shows the analogous impacts of slavery on incarceration for blacks from a 2SLS specification as well as the OLS estimates. Coefficients follow a similar pattern as those in Figure 4, and are approximately the same size, providing evidence that the main effect on incarceration rates come from counties more reliant on slavery.

Table 9 presents the 2SLS results for black incarceration after including economic and geographic controls in Column 2. Again, there is no evidence that slavery affected white incarceration rates (columns 3-4). The point estimates for whites are indistinguishable from zero at conventional levels, whereas for blacks they are statistically significant and persistent over time.

[44]Following Imbens & Angrist (1994), this assumption is also known as the "no defiers" assumption, and it assumes that the instrument affects the treatment in the same direction for the entire sample.

2SLS coefficients are larger than OLS. One possible explanation for this pattern is that OLS is downward biased because counties that relied more on slavery tried to maintain control over the newly freed African American population. Anecdotal evidence indicates that the abolition of slavery and the subsequent disenfranchisement of blacks threatened whites who controlled local politics which. As a result, this created incentives in former slaveholding counties to promote an environment of violence and intimidation against the new freedmen. However, 2SLS and OLS coefficients are not not significantly differrent.

A robustness test on the estimated 2SLS coefficient
I perform an additional placebo test in the form of randomization inference following Barrett & Paul (2017). This test rests on the principle that introducing randomness into the endogenous explanatory variable of interest (slavery prevalence) while holding constant the instrument, should eliminate, or at least, substantially attenuate, the estimated causal relationship if indeed an exogenous variation in the endogenous explanatory variable (slavery) drives the main outcome (incarceration rates). Therefore, I randomly assign (without replacement) the share of slaves in the population among counties in the sample. This "new dataset" preserves the source of endogeneity that Barrett & Paul (2017) worry about - selection into the treatment- but sweeps out the source of variation by randomizing slavery among counties. This way, incarceration rates can remain spuriously related to the instrument because neither the incarceration rates nor the instrument are altered, but the causal mechanism has been rendered non operational by randomization.

If it is true that the causal relationship between slavery and black incarceration rates is positive and the identification is unaffected by selection bias, the distribution of coefficients would shift to the left relative to the original point estimate, and if the share of counties in which slavery causes incarceration rates is small relative to a large enough sample, would center around zero. This is because the randomization of slavery would attenuate the estimated relationship between slavery and incarceration. Figure 10 plots the results of this exercise. The distribution of parameter estimates shifts to the left of the IV coefficient estimate. I obtain a non statistically significant mean coefficient of -70 from this randomization process. This implies that there is not a direct correlation between cotton suitability and black incarceration rates, which would invalidate the exclusion restriction.

6 Labor Demand as a Mechanism

My results show that slavery had a persistent effect on black incarceration. In this section, I turn to two competing mechanisms underlying the relationship between slavery and black incarceration. The first is a demand side mechanism, in which convict labor was used to replace slave labor. The second is a supply side mechanism, in which poor economic conditions after the abolition of slavery pushed blacks to commit more crime. This paper focuses on providing evidence in support of the former, which is a mechanism often highlighted by historians and even the popular press, but which has not been formally empirically tested. Of course, this does not exclude the possibility that a supply side mechanism plays a role too, but testing for it it is beyond the scope of the available data. More datasets on the types of crimes committed and sentences need to be collected and digitalized.[45]

Historians have argued that the end of slavery affected the Southern economic landscape not only by the wealth shock that it represented to whites but also by affecting the labor market. Cotton production was a labor-intensive process, but with the abolition of slavery, land was abundant and labor supply the limiting factor (Ransom & Sutch, 2001). The southern stagnation after the Civil War made it difficult for planters to pay wages to the recently freed black population (Higgs, 2008).[46] Furthermore, emancipation brought blacks some freedom over the amount of labor they supplied, and many ex-slaves chose to work for themselves rather than for white planters (Ransom & Sutch, 2001). This both reduced the labor supply and increased labor costs sharply, threatening the Southern plantation economy (Alston & Ferrie, 2007; Ransom & Sutch, 2001). As a result, whites could have had incentives to establish new forms of labor coercion that would replace slavery. Taken together with the fact that slavery was abolished except as a punishment for crimes, this might explained why African Americans were more likely to end up in prison.

Before turning to the empirical analysis, it is worth describing the features of the penal labor system in the South. It was introduced during the Reconstruction period (1865-1877) when the government of the US was trying to revive the economy of the former Confederate states. Historical accounts document that it was intended to replace the labor force once slaves had been freed (Wilson, 1993; Alston & Ferrie, 2007; Ransom & Sutch, 2001). On average, profit made from convicts was four times higher the cost of prison administration. The main labor systems were convict leasing and chain gangs. The former included monetary incentives to the police and judicial system.[47] Blacks were charged with vagrancy and minor crimes, and then leased on first-bid auctions (Cohen, 1976).[48] Anecdotal evidence suggests that sheriffs were directly asked to arrest more people before the cotton harvest season (Blackmon, 2008; Oshinsky, 1997).

[45] The demand side story is consistent to a very simple variation of a Becker (1971) type of discrimination model. In this setting one can think of the following assumptions: i) assume that convict labor is cheaper than free labor (according to the Department of Labor, convict labor was 20% the price of free labor), ii) there is some disutility from using convict labor (e.g. moral cost), iii) owner plantations have incentives to use black convicts when the cost of doing so it is smaller than hiring free workers. Alternatively, one can think of the supply story as the Becker (1968) traditional model of crime, in which individuals assess the "economic gains" from legal activity against the "economic gains" from committing crime.

[46] Figure A3A illustrates the geographic distribution of wages.

[47] Department of Labor. Laws relating to prison labor, 1933.

[48] While a high share of the auction money went to the state, some portion of that price was paid to the sheriff and judge who were involved in the criminal case (Cohen, 1976).

6.1 Types of Correctional Institutions that Emerged Post-slavery

I start testing this conjecture by collecting and digitalizing records from the Department of Labor on convict labor camps. Overall, the dataset contains 464 correctional facilities in southern counties. Figure 11 maps the number of prisoners employed in forced labor after controlling for population; most of the prison camps were concentrated in what is known as the black belt in the south.[49] The records from the Department of Labor also document in detail the name of the company that employed prisoners, as well as the cost of maintaining inmates and the profits from goods produced by them.

After the Civil War, different types of correctional institutions emerged, including jails, prisons and convict camps. This can be seen in Table 10, which presents the results of regressing dummy variables indicating which types of prison institutions existed in a given county and year on the pre-Civil war intensity of slavery for all counties in the South for the entire period of analysis. All specifications control for state and year fixed effects. Specifically, the results show that the introduction of convict camps was more likely to occur in counties with high concentrations of slaves. The point estimates indicate that a county with a 10 percentage point higher prevalence of slavery is associated with a 2.4 percentage point higher probability of having at least one chain gang. The same is true for penitentiaries, lumber prison camps, and farm prison camps. But the effects are largest for those institutions that provide the most labor: chain gangs, lumber camps and farm camps, and not statistical effect is found for institutions such as jails, which existed in 64% of the counties. The same is true for reformatory and military institutions.

6.2 Reverse Shocks to the Demand of Black Labor

A clear implication of the shortage of labor after the abolition of slavery is that once the demand for black labor drops due to exogenous shocks, the incentives for whites to interfere in the labor market should lessen, and thus the effect of slavery on black incarceration should also diminish. This implication is testable given that much of the Southern economy was agricultural, and its main cash crop – cotton– was heavily labor intensive until about the 1930s when Southern agriculture started to mechanize and tractors began to replace labor. To test this, I exploit three exogenous reverse shocks to the demand for black labor by using a difference in difference approach with state and year fixed effects.

6.2.1 Proximity to Agricultural Stations Established in 1880

The establishment of federal agricultural experiment stations in the late nineteenth century serves as a source of exogenous variation in the location of agricultural knowledge production. These stations positively affected land productivity 20 years after they were opened, and allowed the diffusion of advanced farming practices (Kantor & Whalley, 2019). These

[49] The year of 1886 corresponds to the first time in which convict camps are legally created in the US.

stations were opened at predetermined land-grant universities in response to nationwide concerns about agriculture. They created a positive shock to research independent of local economic conditions (Kantor & Whalley, 2019). As a result, one could expect that a more mechanized production reduced the demand for black labor. I calculate the linear distance in kilometers between the counties' most central points and the closest agricultural stations.

Table 11 presents a triple difference in difference estimation of the effect of slavery in counties that were further away from agricultural stations after they were implemented in 1880 for black and white incarceration rates on columns 1 and 2. The estimates indicate that the effects of slavery on black incarceration are stronger for counties that were further from the agricultural stations. An interpretation of this result is that counties closer to agricultural stations quickly replaced manual labor in agriculture, and therefore, there was lower demand for the labor force provided by African Americans.

6.2.2 Exposure to Boll Weevil Cotton Pest

The Boll Weevil was an agricultural plague that adversely affected cotton production in the American South. The boll weevil feeds almost exclusively on cotton, and its arrival caused large declines in cotton yields. During 1909-1935 the average reduction in cotton production was 11%, ranging from 0.8% in Missouri to 17.8% in Louisiana (Ager *et al.* , 2015). Importantly, the spread of the boll weevil was determined by climatic and geographic conditions, in particular, temperature and wind directions (Hunter & Bert, 1923; Lange *et al.* , 2009a). Farmers and local authorities could do little to prevent it, implying that its arrival was largely exogenous to counties' economic conditions (Lange *et al.* , 2009b). As a result, this pest could have shifted the demand for labor in cotton plantations.

Table 12 shows the effect of slavery on black and white incarceration rates for counties that were affected by the cotton pest in a difference in difference setting. Results show that counties that were affected by the cotton pest experience both lower black and white incarceration rates. Although the effect for African Americans is about three times higher. One possible interpretation of this result is that black labor was not demanded in counties where cotton crops were affected. In addition, this supports the story that the abolition of slavery affected black incarceration through the labor market mechanism, and not through lower economic conditions that decreased the opportunity cost of committing crime. One could imagine that if it is true that African Americans were involved in more criminal activities, then counties that were also more adversely affected by the pest should have seen an increase in black incarceration, since cotton production was central to the Southern economy. This is not the case.[50]

[50]One caveat from this exercise is that I only have information on whether a county has been affected by the cotton pest during the entire period of analysis. More detailed data on the exact year in which the pest arrived at the county would allow me to identify better the effect of slavery. This information is available but it needs to be digitalized from the USDA archives.

6.2.3 Exposure to Mississippi River Floods

Hornbeck & Naidu (2014) show that flooded counties in the Great Mississippi Flood of 1927 experienced an immediate and persistent out-migration of black population. As a result, over time, landowners modernized agricultural production and increased capital intensity relative to landowners in nearby similar non-flooded counties. Therefore, I interact the proportion of slaves in 1860 with the proportion of land destroyed after this natural event in a triple difference in difference specification.[51] As Table 13 shows, the effect of slavery is weaker for counties where the Mississippi flooded.[52]

7 Discussion

When doing a back of the envelope calculation of the number of prisoners required to meet the labor demand in the US, it is possible to see that inmates were a relatively small share of the working population. According to census records, by 1880 there were 36.761.607 workers in the US, and 60,140 inmates, which correspond to 0,1% of the working population. Therefore, even if I find that one of the reasons for which there is an increase in black incarceration rates in counties that relied more slavery, is the need to replace labor after the abolition of slavery; one could think that another way to interpret the main results presented in this paper is that with the labor shortage after the abolition of slavery, the prison system was used as a threat to keep African Americans under coercive working relationships.[53]

For instance, in addition to the use of convicts, peonage was another system of forced labor that emerged after the abolition of slavery in th US. Peonage was a tenant farming system in which a black workers were held in involuntary servitude. This system was prevalent in the Southern states where planters and employers developed it as a substitute for the kind of slavery that was abolished with the Civil War. The Supreme Court called it "a status or condition of compulsory service, based upon the indebtedness of the peon to the master". One of the ways of recruiting workers under this system was for an employers to go a center where unemployed workers were seeking jobs. He promised them good wages and conditions. After arriving to the job, workers would find that they are in debt for transportation. Then they go into debt for food and clothes. After this, they were legally bound to pay off their debts before leaving the job, and the punishment was to pay time in prison (Wilson, 1933).

[51] Data come from Hornbeck & Naidu (2014).

[52] In this exercise it is important to mention that after the Mississippi River floods there was a decrease in the black population, and this in turn could have mechanically reduce black incarceration rates. However, the triple difference in difference specification allows me to compare counties that were flooded but that had different intensity of slavery. Therefore, I can learn what happens whenever there is a shortage of labor in terms to incarceration rates. The effect that I find in the Missisippi River floods can be comparable to the effect that I find in counties exposured to the Cotton Pest. In both cases there was a reduction in black incarceration rates.

[53] In future research, there is a need to investigate more about the Black Codes that regulated working conditions for African Americans, as a way to understand whether these "laws" were more prevalent in places that required more workers.

As a result, in future work, it might be worth to look at the racial composition of households to provide empirical evidence of the existence or not of this labor institution.

8 Conclusion

This paper contributes to understanding the origins of the substantial race gap in incarceration in the US. In particular, I show the role of the institution of slavery as one of the factors shaping this incarceration gap. Specifically, I document a substantial increase in black incarceration immediately after the abolition of slavery, with no comparable effects on whites, and that this black-white incarceration gap continues to grow. Importantly, these results are not driven by omitted variables. Furthermore, I use novel historical data from the Department of Labor to provide evidence of the use of prison labor after abolishing slavery, which suggests that the scarcity of labor supply could have driven the effects. This mechanism is supported when looking at different exogenous shocks that reduced the demand for labor. This, of course, does not rule out that other mechanisms, such as increased criminality due to poor living conditions for instance, were also important.

The findings in this paper provide motivation for future work in several directions. First, one key question not addressed here is how the effects of slavery are mediated by the intergenerational mobility consequences of slavery.[54] Specifically, one can ask whether some of the persistence of the race gap in incarceration, and even its growth over time, is partly attributed to spill-over effects of parental incarceration on their children. Therefore, I plan to match individuals across censuses to compare incarceration outcomes of former slaves, and their children and grandchildren.

Second, one limitation of my work to date is that it stops in 1940 because of confounding factors, like the Great Migration. But, events like this as well as other social and economic events in the 20^{th} century, e.g. the Civil Rights movement in the 1960s, could also have played a role in shaping the race gap in incarceration today. Further analyses are needed to understand the potential impacts, and how they compare in size to the large role played by the abolition of slavery.

[54] There is a growing literature that finds that when parents were in prison, children have a higher chance to end up in prison as well (Hjalmarsson & Lindquist, 2012).

References

Abramitzky, Ran, & Boustan, Leah. 2017. Immigration in American Economic History. *Journal of Economic Literature*, **55**(4), 1311–45.

Acemoglu, Daron, & Wolitzky, Alexander. 2011. The Economics of Labor Coercion. *Econometrica*, **79**(2), 555–600.

Acharya, Avidit, Blackwell, Matthew, & Sen, Maya. 2016. The Political Legacy of American Slavery. *The Journal of Politics*, **78**(3), 621–641.

Ager, Philipp, Brückner, Markus, & Herz, Benedikt. 2015. Effects of Agricultural Productivity Shocks on Female Labor Supply: Evidence from the Boll Weevil Plague in the US South. *Working Paper*.

Ager, Philipp, Eriksson, Katherine, & Boustan, Leah. 2019. The Intergenerational Effects of a Large Wealth Shock: White Southerners after the Civil War. *NBER Working Paper*, March.

Alston, Lee, & Ferrie, Joseph. 2007. *Southern Paternalism and the American Welfare State: Economics, Politics, and Institutions in the South, 1865-1965*. Political Economy of Institutions and Decisions. Cambridge University Press.

Anwar, Shamena, Bayer, Patrick, & Hjalmarsson, Randi. 2012. The Impact of Jury Race in Criminal Trials*. *The Quarterly Journal of Economics*, **127**(2), 1017–1055.

Banerjee, Abhijit, & Iyer, Lakshmi. 2005. History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India. *The American Economic Review*, **95**(4), 1190–1213.

Barrett, Christopher, & Paul, Christian. 2017. Revisiting the effect of food aid on conflict: a methodological caution. *Policy Research Working Paper; no. WPS 8171. Washington, D.C. : World Bank Group.*

Becker, Gary. 1971. *The Economics of Discrimination*. 2 edn. University of Chicago Press.

Becker, Gary S. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy*, **76**(2), 169–217.

Bertocchi, Graziella, & Dimico, Arcangelo. 2014. Slavery, education, and inequality. *European Economic Review*, **70**, 197 – 209.

Blackmon, D.A. 2008. *Slavery by Another Name: The Re-enslavement of Black People in America from the Civil War to World War II*. Doubleday.

Buonanno, Paolo, & Vargas, Juan F. 2019. Inequality, crime, and the long run legacy of slavery. *Journal of Economic Behavior and Organization*, **159**, 539 – 552.

Clay, Karen, Schmick, Ethan, & Troesken, Werner. 2018. How the Boll Weevil Left Southerners Richer, Taller, and Better Fed. *Working Paper*.

Cohen, William. 1976. Negro Involuntary Servitude in the South, 1865-1940: A Preliminary Analysis. *The Journal of Southern History*, **42**(1), 31–60.

Davis, D.B. 2006. *Inhuman Bondage: The Rise and Fall of Slavery in the New World*. Oxford University Press.

Dell, Melissa. 2010. The Persistent Effects of Peru's Mining Mita. *Econometrica*, **78**(6), 1863–1903.

Du Bois, W.E.B., & Lewis, D.L. 1999. *Black Reconstruction in America 1860-1880.* Free Press.

Eriksson, Katherine. 2019. Moving North and into jail? The great migration and black incarceration. *Journal of Economic Behavior and Organization*, **159**, 526 – 538.

Fields, Barbara. 1982. *Slavery, Race and Ideology in the United States of America.* Oxford University Press.

FJS. 2013. Federal Justice Statistics - Statistical Tables. *International Centre for Prison Studies.*

Fogel, William, & Engerman, Stanley. 1974. Time on the Cross: The Economics of American Negro Slavery. *History: Reviews of New Books*, **2**(8), 199–199.

Garret, Paul, MacCormick-Austin. 1929. Handbook of American Prisons and Reformatories. *National Society of Penal Information.*

González, Felipe, Marshall, Guillermo, & Naidu, Suresh. 2017. Start-up Nation? Slave Wealth and Entrepreneurship in Civil War Maryland. *The Journal of Economic History*, **77**(2), 373–405.

Higgs, R. 2008. *Competition and Coercion: Blacks in the American Economy 1865-1914.* Hoover press publications. Cambridge University Press.

Hjalmarsson, Randi, & Lindquist, Matthew J. 2012. Like Godfather, Like Son: Exploring the Intergenerational Nature of Crime. *Journal of Human Resources*, **47**(2), 550–582.

Hornbeck, Richard, & Naidu, Suresh. 2014. When the Levee Breaks: Black Migration and Economic Development in the American South. *American Economic Review*, **104**(3), 963–90.

Hunter, Walter, & Bert, Coad. 1923. The boll-weevil problem. *USDA*.

Imbens, Guido W., & Angrist, Joshua D. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, **62**(2), 467–475.

Kantor, Shawn, & Whalley, Alexander. 2019. Research Proximity and Productivity: Long-Term Evidence from Agriculture. *Journal of Political Economy*, **127**(2), 819–854.

Kleibergen, Frank, & Paap, Richard. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, **133**(1), 97–126.

Kousser, J. Morgan. 2010. The Immutability of Categories and the Reshaping of Southern Politics. *Annual Review of Political Science*, **13**(1), 365–383.

Lagerlöf, Nils-Petter. 2009. Slavery and Other Property Rights1. *The Review of Economic Studies*, **76**(1), 319–342.

Lange, Fabian, Olmstead, Alan L., & Rhode, Paul W. 2009a. The Impact of the Boll Weevil, 1892–1932. *The Journal of Economic History*, **69**(3), 685–718.

Lange, Fabian, Olmstead, Alan L., & Rhode, Paul W. 2009b. The Impact of the Boll Weevil, 1892–1932. *The Journal of Economic History*, **69**(3), 685–718.

Lochner, Lance, & Moretti, Enrico. 2004. The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports. *The American Economic Review*, **94**(1), 155–189.

Markevich, Andrei, & Zhuravskaya, Ekaterina. 2018. The Economic Effects of the Abolition of Serfdom: Evidence from the Russian Empire. *American Economic Review*, **108**(4-5), 1074–1117.

Mitchener, Kris James, & McLean, Ian W. 2003. The Productivity of US States since 1880. *Journal of Economic Growth*, **8**(1), 73–114.

Montiel-Olea, José-Luis, & Pflueger, Carolin. 2013. A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*, **31**(3), 358–369.

Morgan, Kelly. 2019. The standard errors of persistence. *Working Paper Series, No. 19/13, University College Dublin, UCD Centre for Economic Research.*

Naidu, Suresh. 2010. Recruitment Restrictions and Labor Markets: Evidence from the Postbellum U.S. South. *Journal of Labor Economics*, **28**(2), 413–445.

Naidu, Suresh, & Yuchtman, Noam. 2013. Coercive Contract Enforcement: Law and the Labor Market in Nineteenth Century Industrial Britain. *American Economic Review*, **103**(1), 107–44.

North, Douglass C. 1971. Institutional Change and Economic Growth. *The Journal of Economic History*, **31**(1), 118–125.

Nunn, Nathan. 2008. The Long-term Effects of Africa's Slave Trades*. *The Quarterly Journal of Economics*, **123**(1), 139–176.

Nunn, Nathan, & Wantchekon, Leonard. 2011. The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review*, **101**(7), 3221–52.

Oshinsky, D.M. 1997. *Worse Than Slavery*. Free Press.

Poyker, Michael. 2019. Economic Consequences of the U.S. Convict Labor System. July.

Ransom, R.L., & Sutch, R. 2001. *One Kind of Freedom: The Economic Consequences of Emancipation*. Cambridge University Press.

Rehavi, M. Marit, & Starr, Sonja B. 2014. Racial Disparity in Federal Criminal Sentences. *Journal of Political Economy*, **122**(6), 1320–1354.

Robinson, James A., & Acemoglu, Daron. 2008. The Persistence and Change of Institutions in the Americas. *Southern Economic Journal*, **75**(2), 282–299.

Sellin, J. Thorsten. 1976. *Slavery and the Penal System*. New York: Elsevier.

Sharkey, Charles F., & Patterson, George. 1933. *Laws Related to Prison Labor in the United States as of July 1, 1933*. Washington: United States Government Printing Office.

Soares, Rodrigo R., Assunção, Juliano J., & Goulart, Tomás F. 2012. A note on slavery and the roots of inequality. *Journal of Comparative Economics*, **40**(4), 565 – 580. Slavery, Colonialism and Institutions Around the World.

Valencia, Felipe, Fujiwara, Thomas, & Laudares, Humberto. 2011. Tordesillas, Slavery and the Origins of Brazilian Inequalit. *American Economic Review*, **101**(7), 3221–52.

Wilson, Walter. 1933. Forced labor in the United States. *Books on American Labor. International Publishers New York*.

Wilson, Walter. 1993. *Forced labor in the United States*. Free Press.

Wines, E.C. 1871. Transactions of the National Congress of Penitentiary and Reformatory Discipline held at Cincinat. *Albany: The Argus company, printers*.

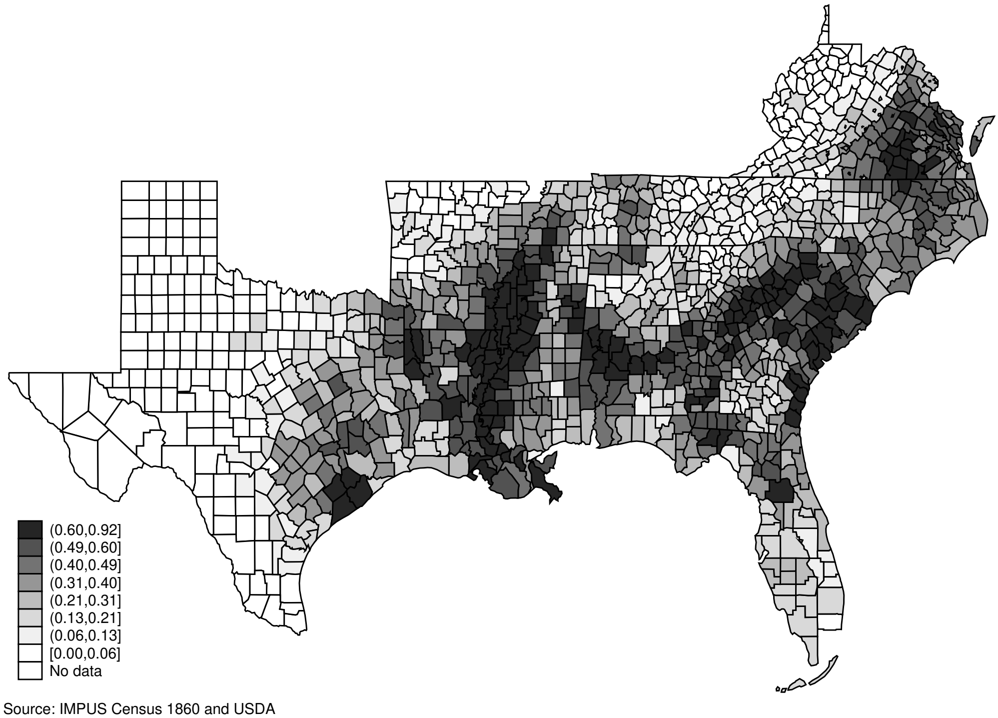
Wright, G. 1978. *The Political Economy of the Cotton South: Households, Markets, and Wealth in the Nineteenth Century*. Norton.

Wright, G. 1986. *Old South, New South: Revolutions in the Southern Economy Since the Civil War*. Basic Books.

Wright, G. 2013. *Slavery and American Economic Development*. Walter Lynwood Fleming Lectures in Southern History. Louisiana State University Press.

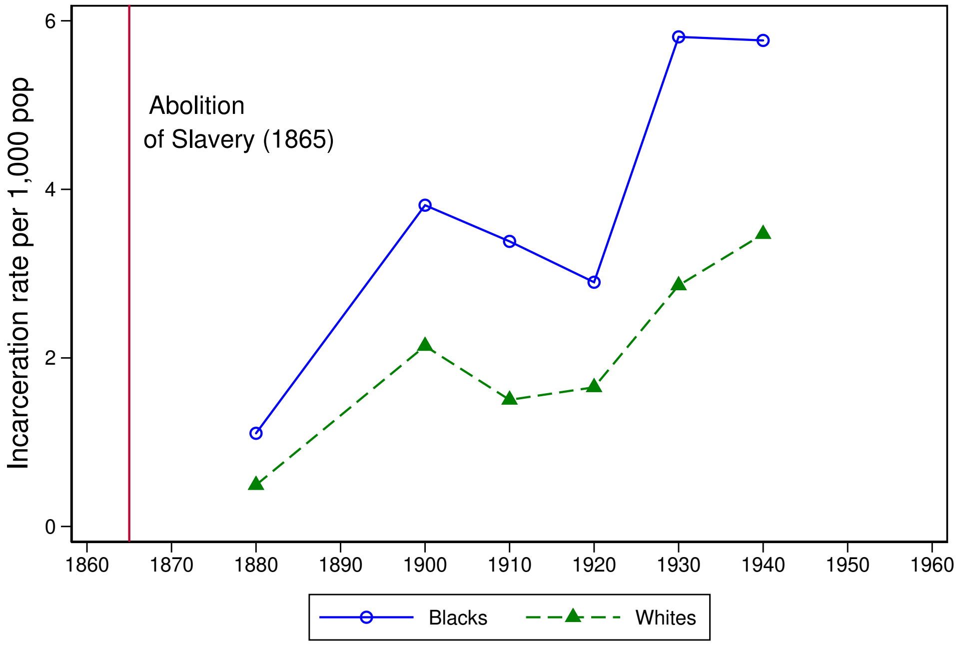
Figures

Figure 1: Geography of Slavery: slaves as share of population in 1860 counties



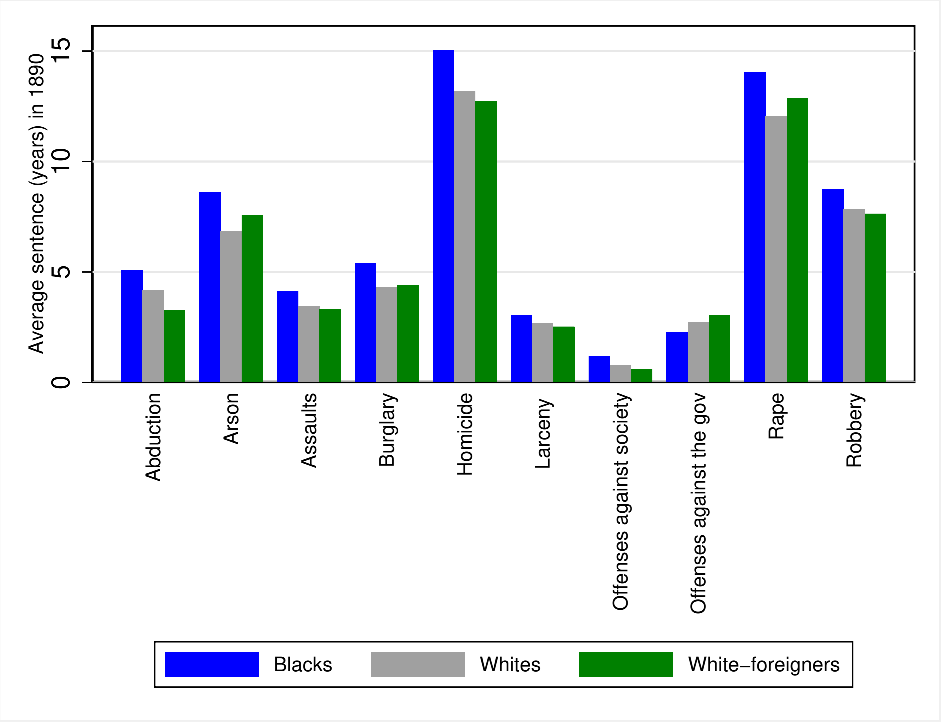
Note: This figure plots the explanatory variable - the intensity of slavery. It corresponds to the proportion of each county's 1860 population that was enslaved, as measured by the 1860 US Census. Darker counties are counties with higher prevalence of slavery. Source: US Census IPUMS.

Figure 2: Black-white incarceration rate 1850-1940 in the whole US



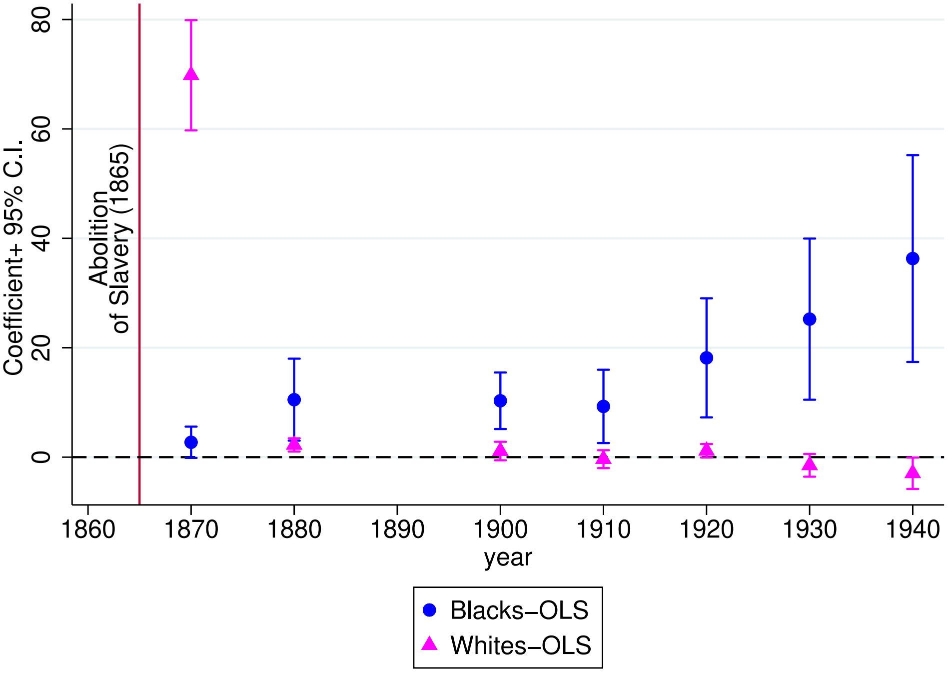
Note 1: Prisoners are counted as the number of males (non-wardens) living in group quarters corresponding to correctional institutions by the time the decennial census was taken place. The incarceration rates are calculated by dividing the total inmate population for each race by their corresponding population at county level for the Southern states. The graph shows the average per year. *Note 2:* The vertical line indicates the 1865 slavery abolition. *Note 3:* Southern states include: Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, Texas, Virginia, West Virginia. Source: US Census IPUMS.

Figure 3: Average length of sentences by race (1890-1910)



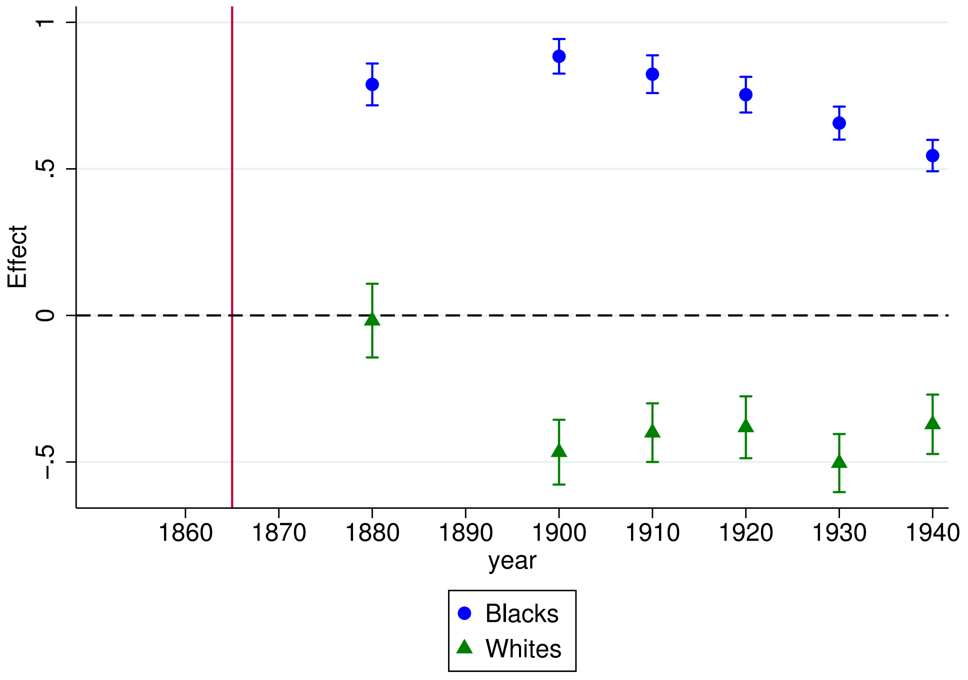
Note 1: The figure shows the average length of sentences for prisoners disaggregated by race for the period 1890-1910. Black prisoners were charged with longer sentences for the same crimes compared to other prisoners. *Note 2:* Source: "Crime and Pauperism, Census records." Tables 105-131. Average sentences by crime and race, males Pg 89

Figure 4: Main Results: Effect of slavery on incarceration rates



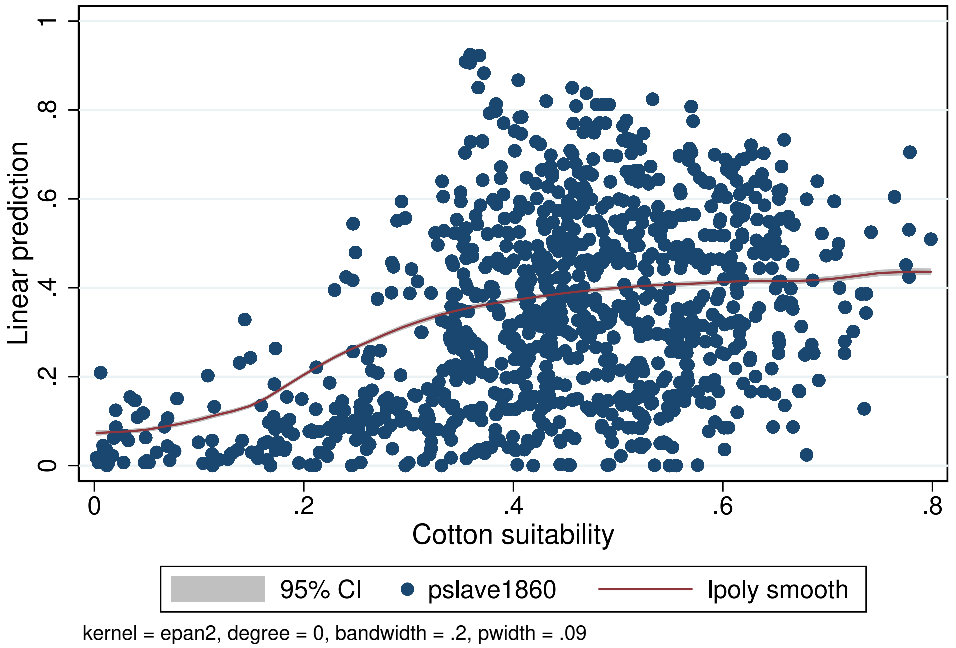
Note: The figure shows the coefficient on slavery (shares of slaves in the total population) estimated separately for each census year as in Equation 1. Dots represent the point estimates for African Americans, while triangles for whites. 95% confidence intervals are plotted as well. In particular, for each census year, I estimate an OLS regression that relates incarceration rates by race against the intensity of slavery, after including county characteristics measured in 1860, and state fixed effects. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment.

Figure 5: Results on alternative dependent variable: share of blacks and whites in the prison population



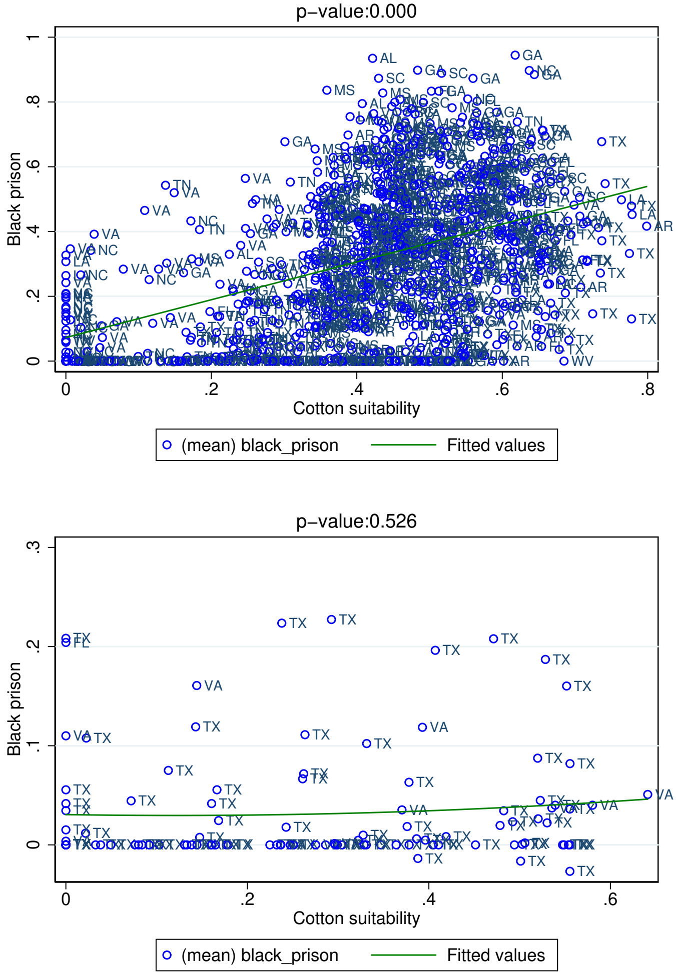
Note: The figure shows the coefficient on slavery (shares of slaves in the total population) estimated for each census year on the share of blacks and whites in the prison population. Dots represent the point estimates for African Americans, while triangles are for whites. 95% confidence intervals are plotted as well. In particular, for each census year, I estimate an OLS regression that relates incarceration rates by race against the intensity of slavery, after including county characteristics measured in 1860, and state fixed effects. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment.

Figure 6: First stage: slavery on cotton suitability (local linear regression)



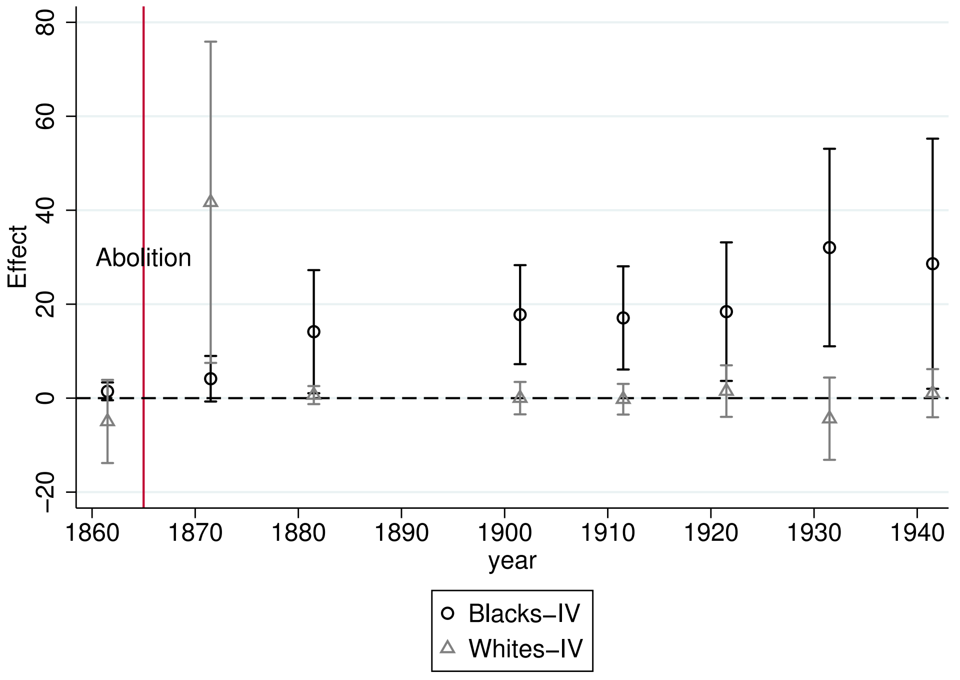
Note: This figure plots the prediction in the intensity of slavery (Y-axis) against the suitability for growing cotton at county level (X-axis) from a local linear regression. 95% confidence intervals.

Figure 7: Placebo test for exclusion restriction of the instrument



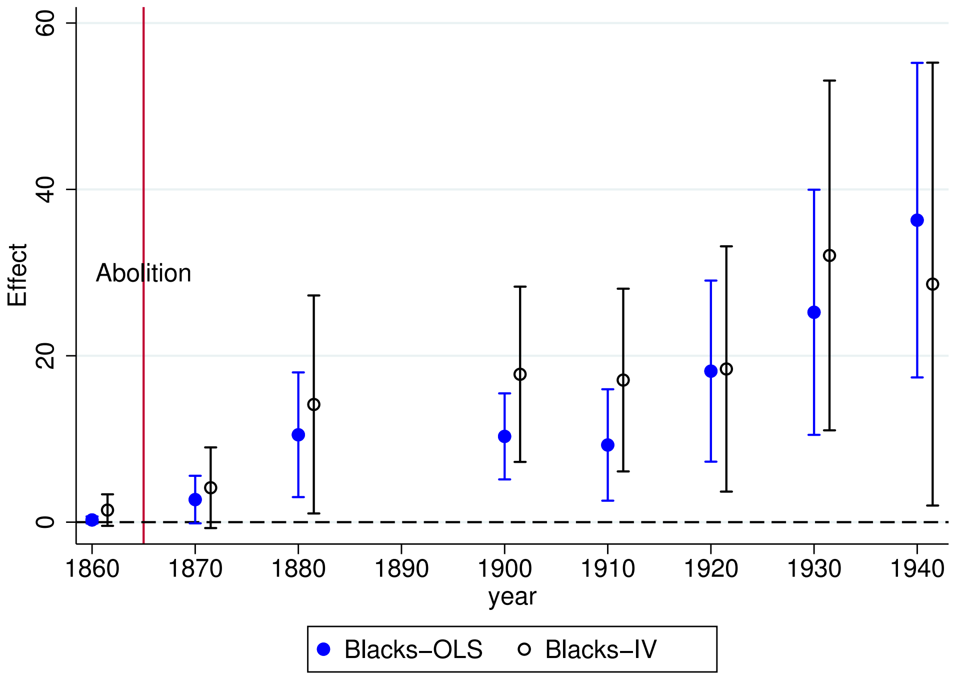
Note: The figures show the average black incarceration filtered by a set of state and time effects. The top figure presents the counties that have positive intensity of slavery. The other figure displays counties without slavery. There is only a positive relationship between the instrument-cotton suitability index- and the conditional measure of black incarceration, suggesting that the only way in which the instrument affects incarceration rates is through the levels of slavery.

Figure 8: IV Results: black and white incarceration rates on slavery (Instrument: cotton suitability)



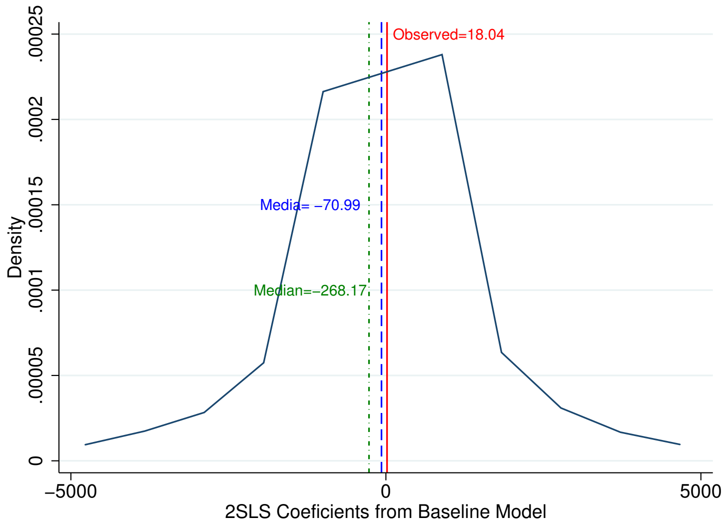
Note: The figure shows the coefficient on slavery (shares of slaves in the total population) estimated for each census year. Slavery is instrumented with the suitability for growing cotton. Dots represent the effect for African Americans, while triangles are for whites. 95% confidence intervals are plotted as well. In particular, for each census year, I estimate an IV regression that relates incarceration rates by race against the intensity of slavery, after including county characteristics measured in 1860, and state fixed effects. Slavery is instrumented with the suitability for growing cotton. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment.

Figure 9: OLS vs. IV Results: black incarceration rates on slavery



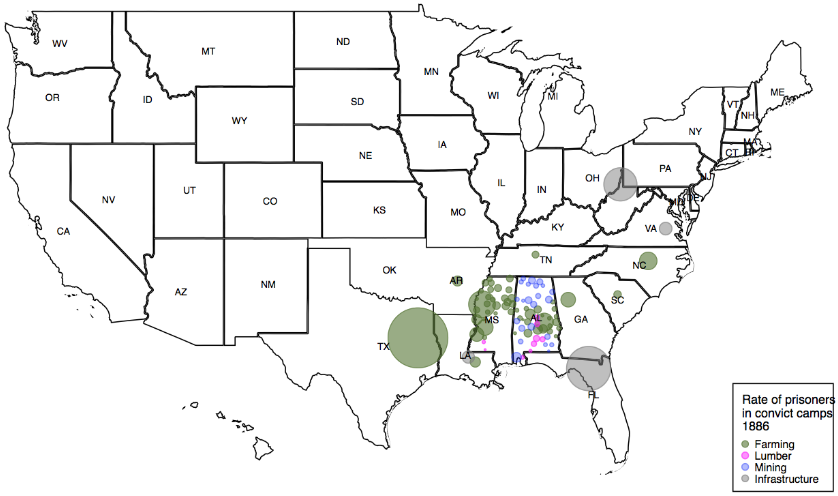
Note: The figure shows the coefficient on slavery (shares of slaves in the total population) estimated for each census year. Slavery is instrumented with the suitability for growing cotton. Dots represent the effect for African Americans, while triangles are for whites. 95% confidence intervals are plotted as well. In particular, for each census year, I estimate an OLS regression that relates incarceration rates by race against the intensity of slavery, after including county characteristics measured in 1860, and state fixed effects. Furthermore, IV estimates where slavery is instrumented with cotton suitability are presented as well. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment.

Figure 10: Distribution of 2SLS Coefficient Estimates Using Randomized Slavery Allocations



Note: The density plot depicts the distribution of 2SLS coefficient estimates using the set of baseline controls with 10,000 draws of randomized allocation of slavery among counties. This figure shows that cotton suitability does not affect incarceration rates in a different way than slavery.

Figure 11: Prisoners employed in convict camps



Note: The map shows the distribution of labor convict camps in 1886 after controlling by population and disaggregated by industry. Bigger circles represent more prisoners employed in these camps. Source: Department of Labor Archival Records.

Tables

Table 1: Descriptive Statistics

Variables employed in the main specifications	(1) N	(2) mean	(3) sd	(4) min	(5) max
Panel A: *Prison and population variables*					
Prop. of slaves in pop. 1860	6587	0.343	0.217	0	0.924
Share of free black pop. 1860	7180	0.011	0.0243	0	0.246
Incarcerated pop.	7234	61.93	260.7	0	5763
Black incarcerated pop.	7234	26.2	116.2	0	2107
White incarcerated pop.	7197	35.85	172.7	0	3801
Share of blacks in pop.	7234	0.262	0.231	0	0.935
Share of whites in pop.	7234	0.737	0.231	0.064	1
Share of blacks in prison	7234	0.430	0.353	0	1
Share of whites in prison	7234	0.567	0.353	0	1
Black incarceration rate	6879	5.79	23.58	0	640.63
White incarceration rate	7197	2.03	5.23	0	118.17
Panel B: *Type of correctional facility*					
				avg N. in 1880	avg N. in 1940
Prison (=1)	1180	0.0763	0.266	3	20
Jail (=1)	1180	0.642	0.48	111	570
Penitentiary (=1)	1180	0.0254	0.157	18	180
Chain gang (=1)	1180	0.272	0.445	18	180
Lumber (=1)	1180	0.0331	0.179	1	29
Mining (=1)	1180	0.0119	0.108	2	16
Railroad (=1)	1180	0.0669	0.25	2	64
Panel C: *County characteristics measured in 1860*					
Land inequality 1860	7180	0.485	0.0844	0	0.789
Farm value 1860	7180	184.3	123.2	0	966
Slave mortality rate	7180	1611	529.6	0	6667
% Democratic vote	7180	6.727	9.733	0	70.2
Rail presence (=1)	7180	0.255	0.405	0	1
River presence (=1)	7180	0.344	0.432	0	1

Note: This table shows the summary statistics at county level. Panel A presents the measures of incarcerated population and total males for the period 1860-1940. All the variables are presented as the average across years for this period at county level. Panel B displays the socio-economic characteristics measured in 1860. Source: IPUMS Census

Table 2: Main results: The effect of slavery on black and white incarceration rates

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	*Black incarceration rate*			*White incarceration rate*		
	OLS	OLS	OLS	OLS	OLS	OLS
Panel A: *All Sample*	(N=6,719; Mean dep. Var (1)-(3)=5.79, Mean dep. Var (4)-(6)=2.03)					
Proportion Slaves	16.230***	16.872***	18.047***	-1.138	-1.138	-1.741**
	(3.830)	(4.257)	(4.436)	(0.737)	(0.737)	(0.810)
R-squared	0.045	0.045	0.052	0.144	0.144	0.145
Panel B: *Census 1870*	(N=822; Mean dep. Var (1)-(3)=2.26, Mean dep. Var (4)-(6)=0.64)					
Proportion Slaves	2.714*	2.323	-0.921	69.814***	69.814***	69.937***
	(1.459)	(1.442)	(1.637)	(5.138)	(5.138)	(5.245)
R-squared	0.013	0.014	0.048	0.001	0.001	0.001
Panel C: *Census 1880*	(N=872; Mean dep. Var (1)-(3)=1.53, Mean dep. Var (4)-(6)=0.35)					
Proportion Slaves	10.504***	10.464**	11.511**	1.323***	1.323***	1.282***
	(3.823)	(4.162)	(4.537)	(0.432)	(0.432)	(0.452)
R-squared	0.024	0.024	0.032	0.041	0.041	0.041
Panel D: *Census 1900*	(N=941; Mean dep. Var (1)-(3)=4.75, Mean dep. Var (4)-(6)=2.42)					
Proportion Slaves	10.307***	10.335***	12.034***	0.114	0.114	-0.433
	(2.639)	(2.903)	(2.958)	(0.745)	(0.745)	(0.784)
R-squared	0.039	0.039	0.081	0.016	0.016	0.022
Panel E: *Census 1910*	(N=978; Mean dep. Var (1)-(3)=4.25, Mean dep. Var (4)-(6)=1.58)					
Proportion Slaves	9.275***	8.996**	9.590***	-0.742	-0.742	-0.981
	(3.416)	(3.605)	(3.714)	(0.679)	(0.679)	(0.727)
R-squared	0.028	0.028	0.033	0.019	0.019	0.02
Panel F: *Census 1920*	(N=1032; Mean dep. Var (1)-(3)=4.33, Mean dep. Var (4)-(6)=1.75)					
Proportion Slaves	18.154***	18.227***	19.903***	0.764	0.764	0.318
	(5.553)	(6.040)	(6.211)	(0.578)	(0.578)	(0.622)
R-squared	0.037	0.037	0.048	0.013	0.013	0.016
Panel G: *Census 1930*	(N=1032; Mean dep. Var (1)-(3)=8.77, Mean dep. Var (4)-(6)=3.06)					
Proportion Slaves	25.227***	27.251***	28.599***	-1.788*	-1.788*	-2.517*
	(7.517)	(8.633)	(8.801)	(1.029)	(1.029)	(1.475)
R-squared	0.047	0.049	0.055	0.019	0.019	0.026
Panel H: *Census 1940*	(N=1042; Mean dep. Var (1)-(3)=9.84, Mean dep. Var (4)-(6)=3.68)					
Proportion Slaves	36.301***	38.747***	40.049***	-3.375**	-3.375**	-4.678**
	(9.646)	(10.885)	(11.011)	(1.433)	(1.433)	(1.816)
R-squared	0.051	0.053	0.059	0.069	0.069	0.084
Inst. controls	No	Yes	Yes	No	Yes	Yes
Econ. controls	No	No	Yes	No	No	Yes

Note: Column (1) presents the OLS results without controls, Column (2) includes socio-economic controls. Column (3) presents the 2SLS results, and Column (4) adds controls. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. The level of observation is the county. Standard errors in parenthesis (robust clustered at state level). All regressions include state fixed effects. Column 5-8 have the same structure but for the white incarceration rate. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Table 3: Effect of slavery on black incarceration: restricting the sample to neighboring counties with different levels of slavery

	(1)	(2)	(3)	(4)
Panel A:	Neighboring counties that differ by more than:			
Black incarceration rate	20%	30%	40%	50%
Proportion Slaves	18.350***	18.731***	27.548***	44.951***
	(2.171)	(2.783)	(4.188)	(6.989)
Mean dep. var.=5.79				
Observations	4,181	2,354	1,457	1,176
Controls	Yes	Yes	Yes	Yes
Panel B:	Neighboring counties that differ by more than:			
Share of blacks in prison	20%	30%	40%	50%
Proportion Slaves	0.641***	0.505***	0.494***	0.569***
	(0.043)	(0.051)	(0.050)	(0.049)
Mean dep. var.=2.03				
Observations	4,181	2,354	1,457	1,176
Controls	Yes	Yes	Yes	Yes

Note: The table shows the results from the main estimation after restring the sample for those counties that border a county in which the proportion of slave differs. Columns presents the estimation when counties differ in more than 20%, Column 2, 3 and 4, in 30%, 40% and 50%, respectively. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. Standard error in parenthesis (robust clustered at the state level). The dependent variables is the black incarceration rate in Panel A, and the share of prisoners that are black in Panel B. All regressions include state and year fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 4: Effect of slavery: southern-northern border counties

	(1)	(2)	(3)	(4)
Dependent variable	**Panel A**: *Black Incarceration rate*		**Panel B**: *Share of blacks in prison*	
Sample	All counties	`Border counties`	All counties	`Border counties`
Proportion Slavery	18.047***	12.453**	0.629***	0.156**
	(4.436)	(0.602)	(0.013)	(0.079)
Mean dep. var.	5.79		2.03	
Observations	7,197	670	7,197	670
Controls	Yes	Yes	Yes	Yes

Note: The table shows the results from the main estimation after restricting the sample for those counties located in southern-northern border. Column 1 and 2 presents the results for black incarceration rates, and Column 3 and 4 for the share of blacks in prison. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. Standard error in parenthesis (robust clustered at the state level). All regressions include state and year fixed effects. When estimating Column (2) and Column (4) without the controls roughness of the terrain and the percentage of the democratic vote the points estimates change to 12.5** and 0.159* respectively.
$*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 5: Effect of slavery: northern and southern counties with low levels of slavery

	(1)	(2)	(3)
Panel A:	Slavery intensity Southern & Northern counties		
Black incarceration rate	<1%	<5%	<10%
Proportion Slaves	-47.608		
	(89.494)	(8.202)	(5.313)
Mean dep. var.=5.79			
Observations	10,553	10,898	11,452
Controls	Yes	Yes	Yes
Panel B:	Slavery intensity Southern & Northern counties		
Share of blacks in prison	<1%	<5%	<10%
Proportion Slaves	-0.010	-0.002	-0.001
	(0.482)	(0.039)	(0.016)
Mean dep. var.=2.03			
Observations	10,553	10,898	11,452
Controls	Yes	Yes	Yes

Note: The table shows the results from the main estimation after restring the sample for those counties in the South that have low levels of slavery, and counties in the north. Columns presents the estimation when slave population is less that different percentages. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. Standard error in parenthesis (robust clustered at the state level). The dependent variables is the black incarceration rate. All regressions include state and year fixed effects $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 6: First stage: cotton suitability and slavery

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Proportion of slaves	**All censuses**		**Census 1870**		**Census 1880**		**Census 1900**	
Cotton suitability	0.6511***	0.5969***	0.6646***	0.5014***	0.6492***	0.6072***	0.6610***	0.6303***
	(0.0328)	(0.0334)	(0.0363)	(0.0363)	(0.0352)	(0.0370)	(0.0341)	(0.0357)
Constant	0.1294***	0.0703	0.1226***	0.0758***	0.1275***	0.0629	0.1382***	0.0817**
	(0.0287)	(0.0488)	(0.0289)	(0.0291)	(0.0295)	(0.0479)	(0.0297)	(0.0319)
F-test	139	58.8	123.5	51.89	121.2	51.7	128.5	51.9
Observations	6,719	6,719	822	822	872	872	941	941
R-squared	0.4315	0.4767	0.4321	0.5710	0.4356	0.4729	0.4389	0.4813
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Dependent variable	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Proportion of slaves	**Census 1910**		**Census 1920**		**Census 1930**		**Census 1940**	
Cotton suitability	0.6380***	0.5960***	0.6414***	0.5881***	0.6166***	0.5651***	0.6272***	0.5819***
	(0.0343)	(0.0362)	(0.0338)	(0.0352)	(0.0345)	(0.0360)	(0.0347)	(0.0362)
Constant	0.1370***	0.0569*	0.1455***	0.0911***	0.1542***	0.0819**	0.1590***	0.0934***
	(0.0294)	(0.0307)	(0.0296)	(0.0314)	(0.0305)	(0.0324)	(0.0310)	(0.0329)
F-test	115	47.69	124.3	46.44	114.9	51.27	112.2	54.46
Observations	978	978	1032	1032	1032	1032	1042	1042
R-squared	0.4237	0.4620	0.4378	0.4759	0.4314	0.4680	0.4294	0.4617
Controls	No	Yes	No	Yes	No	Yes	No	Yes

Note 1: The table shows the first stage for the whole sample and separately by censuses. The first column for each year presents the effect of cotton suitability on the proportion of slaves, the second column adds socio-economic controls. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. The level of observation is the county. All regressions include state fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Table 7: Testing monotonicity assumption

Dependent variable Slavery	(1) Low rugged	(2) High rugged	(3) Low slave mortality	(4) High slave mortality	(5) Low dem vote vote	(6) High democratic vote	(7) No river	(8) River	(9) No rail	(10) Rail
Cotton suitability	0.396***	0.448***	0.573***	0.666***	0.588***	0.768***	0.728***	0.304***	0.668***	0.584***
	(0.013)	(0.020)	(0.013)	(0.030)	(0.012)	(0.063)	(0.016)	(0.032)	(0.017)	(0.032)
Constant	0.138***	0.295***	0.146***	0.209***	0.175***	0.044	0.016	0.377***	0.075***	0.261***
	(0.011)	(0.016)	(0.012)	(0.023)	(0.011)	(0.036)	(0.014)	(0.020)	(0.014)	(0.020)
Observations	2,476	4,738	5,739	1,475	6,842	372	3,699	1,972	4,260	1,530
R-squared	0.519	0.353	0.507	0.450	0.510	0.404	0.507	0.435	0.459	0.363

Note: The table shows the results from the first stage of an instrumental variable regression for the different subsamples. Standard error in parenthesis (robust clustered at the state level). The dependent variables is the intensity of slaverty. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. All regressions include county and year fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 8: Reduced form: cotton suitability vs. the different measures of incarceration

Dependent variable	(1)	(2)	(3)	(4)
	Black incarceration rate		*White incarceration rate*	
Panel A: *All Sample*				
Cotton suitability	10.8953***	10.4994***	(1.2279)	(1.6698)
	(2.6797)	(2.9760)	(1.8332)	(1.7502)
Observations	6719	6719	6719	6719
R-squared	0.0402	0.0443	0.1572	0.2272
Panel B: *Census 1880*				
Cotton suitability	9.8824**	9.0035*	0.7392	0.3514
	(4.7705)	(4.9554)	(0.5200)	(0.5481)
Observations	872	872	872	872
R-squared	0.0139	0.0203	0.0148	0.0448
Panel C: *Census 1900*				
Cotton suitability	12.2771***	10.6492***	0.1364	-0.9962
	(2.9198)	(3.2282)	(1.2308)	(1.3053)
Observations	941	941	941	941
R-squared	0.0366	0.083	0.0114	0.0567
Panel D: *Census 1910*				
Cotton suitability	11.0082***	10.4898***	0.6305	0.3854
	(2.7600)	(2.9880)	(0.9074)	(0.9479)
Observations	978	978	978	978
R-squared	0.0275	0.0304	0.0137	0.0182
Panel E: *Census 1920*				
Cotton suitability	12.4105***	11.8612**	1.7258	1.3895
	(4.7700)	(4.9924)	(1.6202)	(1.6844)
Observations	1032	1032	1032	1032
R-squared	0.0157	0.0223	0.0125	0.0378
Panel F: *Census 1930*				
Cotton suitability	18.6230***	19.3859***	-0.8393	-1.4743
	(5.5732)	(6.4614)	(2.1356)	(2.2157)
Observations	1032	1032	1032	1032
R-squared	0.035	0.0384	0.0225	0.0352
Panel G: *Census 1940*				
Cotton suitability	18.0442**	18.1539**	1.0077	0.0653
	(7.0047)	(7.8198)	(1.5553)	(1.7005)
Observations	1042	1042	1042	1042
R-squared	0.029	0.032	0.0628	0.077
Controls	No	Yes	No	Yes

Note 1: This table shows the reduced form specifications for each census year. For all the specifications, the first column does not include controls. The second column includes controls. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. The level of observation is the county. Standard errors in parenthesis (robust clustered at state level). All regressions include state fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Table 9: IV results: The effect of slavery on the black and white incarceration rate

Dependent variable	(1)	(2)	(3)	(4)
	Black incarceration rate		*White incarceration rate*	
		IV Results		
Panel A: *All Sample*	(N=6,719; Mean dep. Var (1)-(2)=5.19, Mean dep. Var (3)-(4)=4.78)			
Proportion Slaves	19.184***	19.865***	-1.947	-3.166
	(5.619)	(6.912)	(2.407)	(2.794)
Panel B: *Census 1870*	(N=822; Mean dep. Var (1)-(4)=2.26, Mean dep. Var (5)-(8)=0.64)			
Proportion Slaves	4.135*	-0.435	-0.000	-0.000
	(2.474)	(4.170)	(0.000)	(0.000)
Panel C: *Census 1880*	(N=872; Mean dep. Var (1)-(2)=2.26, Mean dep. Var (3)-(4)=0.64)			
Proportion Slaves	14.148**	13.190*	0.102	-0.176
	(6.685)	(7.344)	(0.745)	(0.905)
Panel D: *Census 1900*	(N=941; Mean dep. Var (1)-(2)=4.75, Mean dep. Var (3)-(4)=2.42)			
Proportion Slaves	17.775***	17.214***	0.461	-0.488
	(5.375)	(6.106)	(1.690)	(1.887)
Panel E: *Census 1910*	(N=978; Mean dep. Var (1)-(2)=4.25, Mean dep. Var (3)-(4)=1.58)			
Proportion Slaves	17.079***	16.896**	0.050	-0.315
	(5.603)	(6.570)	(1.643)	(1.856)
Panel F: *Census 1920*	(N=1032; Mean dep. Var (1)-(2)=4.33, Mean dep. Var (3)-(4)=1.75)			
Proportion Slaves	18.413**	18.583**	1.479	0.783
	(7.522)	(8.603)	(2.784)	(3.202)
Panel G: *Census 1930*	(N=1032; Mean dep. Var (1)-(2)=8.77, Mean dep. Var (3)-(4)=3.06)			
Proportion Slaves	32.058***	35.900***	-2.768	-4.530
	(10.725)	(13.611)	(4.377)	(5.050)
Panel H: *Census 1940*	(N=1042; Mean dep. Var (1)-(2)=9.84, Mean dep. Var (3)-(4)=3.68)			
Proportion Slaves	28.615**	30.064*	0.881	-0.936
	(13.581)	(16.473)	(2.616)	(3.200)
Controls	No	Yes	No	Yes

Note 1: Column (1) presents the IV results without controls, Column (2) add socio-economic and proxies for racial attitudes. Columns 3 and 4 follow the same pattern for white incarceration rates. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. The level of observation is the county. Standard errors in parenthesis (robust clustered at state level). All regressions include state fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$

Table 10: Relationship between slavery and type of correctional institutions

Dependent variable (=1)	(1) *Chain gang*	(2) *Penitentiary*	(3) *Lumber*	(4) *Farm*	(5) *Jail*	(6) *Reformatory*	(7) *Military*
Proportion Slaves	0.242***	0.151***	0.707*	0.310***	0.169	0.193	0.110
	(0.021)	(0.055)	(0.390)	(0.075)	(0.158)	(0.171)	(0.644)
Observations	6125	6125	6125	6125	6125	6125	6125
R-squared	0.022	0.004	0.008	0.058	0.060	0.004	0.001
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: The table presents the results of regressing dummy variables indicating which types of prison institutions existed in a given county and year on the pre-Civil war intensity of slavery. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. Standard error in parenthesis (robust clustered at the state level). Results are presented for observations pooled for all years. All regressions include state and year fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 11: Proximity to Agricultural Stations Established in 1880 (DDD)

Dependent variable *Incarceration rates by race*	(1) *Black rate*	(2) *White rate*
Proportion Slaves	-2.063	3.627
	(3.580)	(5.154)
T=1 ($\geqslant$ 1 1880 After the introduction of Agri. Stations)	8.648***	7.604***
	(1.925)	(2.771)
Distance Agri. Stations	-0.002	0.001
	(0.002)	(0.003)
Proportion Slaves * Distance Agri. Stations *T$\geqslant$1880	0.015***	0.011
	(0.005)	(0.008)
Observations	6,719	6,719
R-squared	0.065	0.022
Controls	Yes	Yes

Note: The table presents a difference-in-difference-in-difference (DDD) of the effect of the distance to agricultutal stations for counties that were more reliant on slavery. Interactions between all the three terms of the DDD are included as well. Standard error in parenthesis (robust clustered at the state level). The dependent variables is the black and white incarceration rates. The Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. Results are presented for observations pooled for all years. All regressions include state and year fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 12: Exposure to Boll Weevil Cotton Pest

Dependent variable *Incarceration rates by race*	(2) *Black rate*	(4) *White rate*
Proportion Slaves	19.424***	0.215
	(1.866)	(0.986)
Pest (=1)	1.78	1.407**
	(1.174)	(0.615)
Proportion Slaves * Pest (=1)	-7.823***	-2.054*
	(2.753)	(1.084)
R-squared	0.047	0.121
Mean dep. var.=0.30		
Observations	6,719	6,719
Controls	Yes	Yes

Note: The cotton pest occured during the period 1909-1935. In this exercise I record with a dummy variable equal to one if a county has been affected by the cotton pest. The table shows the differential effect of slavery depending on whether counties were affected by the cotton pest. Standard error in parenthesis (robust clustered at the state level). The dependent variables is the black incarceration rate in Panel A, and the share of prisoners that are black in Panel B. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. All regressions include state and year fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 13: Exposure to Mississippi River Floods (DDD)

Dependent variable	(1)	(2)
Incarceration rates by race	*Black rate*	*White rate*
Proportion Slaves	11.214***	8.135***
	(1.863)	(2.054)
T=1 ($\geqslant$ 1927 Mississippi River Floods)	2.905*	1.278
	(1.512)	(1.667)
Flood (=1)	-0.410	0.695
	(4.092)	(4.512)
Proportion Slaves * Flood (=1) *T $\geqslant$ 1927	-24.218**	-16.449
	(11.932)	(13.155)
Observations	6,719	6,719
R-squared	0.052	0.031
Controls	Yes	Yes

Note: Missisippi River floods occured in 1927. The dummy variable $Flood = 1$ takes the value of one a county experienced floods. The table presents a difference-in-difference-in-difference (DDD) of the effect of the Mississippi River floods for counties that were more reliant on slavery. Interactions between all the three terms of the DDD are included as well. Standard error in parenthesis (robust clustered at the state level). The dependent variables are black and white incarceration rates. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. Results are presented for observations pooled for all years. All regressions include state and year fixed effects. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Appendix

Table A1: Main results: The effect of slavery on black and white incarceration rates. - Bootstrapping standard errors

Dependent variable *Black incarceration rates*	(1) All Sample	(2) Census 1870	(3) Census 1880	(4) Census 1900	(5) Census 1910	(6) Census 1920	(7) Census 1930	(8) Census 1940
				OLS estimations				
Proportion Slaves	16.230***	2.714**	10.504***	10.307***	9.275***	18.154***	25.227***	36.301***
	(2.353)	(1.358)	(4.003)	(2.569)	(3.167)	(5.140)	(7.257)	(10.180)
Observations	6719	822	872	941	978	1032	1032	1042
Dependent variable *Black incarceration rates*				IV estimations				
Proportion Slaves	19.184***	4.135*	14.148**	17.775***	17.079***	18.413***	32.058***	28.615**
	(3.494)	(2.396)	(6.123)	(5.024)	(5.014)	(6.881)	(9.075)	(13.145)
Observations	6719	822	872	941	978	1032	1032	1042
Dependent variable *White incarceration rates*				OLS estimations				
Proportion Slaves	-1.138	69.814*	1.323***	1.114	-0.742	0.764	-1.788	-3.375**
	(0.732)	(41.384)	(0.662)	(0.840)	(0.689)	(0.605)	(1.071)	(1.444)
Observations	6719	822	872	941	978	1032	1032	1042
Dependent variable *White incarceration rates*				IV estimations				
Proportion Slaves	-1.947	0	0.102	0.461	0.05	1.479	-2.768	0.881
	(1.464)	(0.000)	(1.037)	(2.093)	(1.523)	(3.148)	(4.100)	(2.281)
Observations	6719	822	872	941	978	1032	1032	1042

Note: Column (1) presents the OLS results without controls, Column (2) includes socio-economic controls. Column (3) presents the 2SLS results, and Column (4) adds controls. Controls include: county size (in acres), average farm value, the proportion of small farms, and a measure of land inequality. These variables proxy for the degree of workforce required for agriculture. In addition, I control for characteristics related to trade and commerce, including indicators for whether the county had access to rail and steamboat-navigable rivers or canals, and the ruggedness of the county terrain, which were crucial for agricultural markets. Finally I include proxies for racist attitudes as the percentage of votes for the democratic party in 1860, the relative mortality of slaves to whites, and the average occupant size of slave quarters in farms as a proxy for slave treatment. The level of observation is the county. Standard errors in parenthesis (bootstrapped). All regressions include state fixed effects. Column 5-8 have the same structure but for the white incarceration rate. $*p < 0.1; **p < 0.05; ***p < 0.01$

Table A2: Characteristics for southern-norther border counties before the abolition of slavery

Sample	Northern counties (N=59)	Southern counties (N=53)	t-stat
Panel A: County characteristics 1860			
Land inequality	0.441	0.449	-1.35
Value of farms	233.24	227.94	0.6
% Democratric vote	35.9	6.7	18.12
River access =1	0.4	0.36	1.45
Rail =1	0.39	0.33	1.54
Rugged terrain	73.43	44.2	5.02
Panel B: County characteristics 1790			
Total population	37906.35	34654.41	1.56
White population	15943.02	14843.87	1.39

Note: The table shows the mean for county characteristics for counties locates in the southern and northern border. Columns 3 displays the t-statistics. .

Figure A1: Appendix: Example of Prison Records from the Department of Commerce

Note: The table corresponds to the year 1910. These records provide information on all correctional facilities in the US, including the number of inmates divided by race.

INSTITUTION AND STATE.	In institutions January 1, 1910.							Committed in 1910.						
	Total.			White.			Col-ored.	Total.			White.			Col-ored.
	Total.	Male.	Fe-male.	Total.[1]	Na-tive.	For-eign born.		Total.	Male.	Fe-male.	Total.[1]	Na-tive.	For-eign born.	
GEORGIA—Continued.														
County jails, workhouses, and chain gangs—Contd.														
Brooks County Chain Gang, Quitman	18	18	......	1	1	......	17	26	25	1	1	1	......	25
Bryan County Chain Gang, Ellabell	13	11	2	......	......	......	13	13	13	......	2	2	......	11
Bulloch County Jail and Chain Gang, Statesboro.[2]	......	......	......	......	......	......	......	......	......	......	......	......	......	......
Burke County Chain Gang, Waynesboro	41	41	......	......	......	......	41	73	72	1	2	2	......	71
Camden County Chain Gang, Kingsland	6	6	......	......	......	......	6	10	10	......	......	......	......	10
Calhoun County Jail or Chain Gang, Morgan	12	12	......	......	......	......	12	2	2	......	......	......	......	2
Carroll County Jail, Carrollton[2]	......	......	......	......	......	......	......	......	......	......	......	......	......	......
Charlton County Jail, Folkston	15	15	......	3	3	......	12	37	37	......	20	20	......	17
Chatham County Chain Gang and Farm, Savannah.	248	209	39	14	14	......	234	974	807	167	58	57	1	916
Chatham County Jail, Savannah	51	43	8	15	14	1	36	449	339	110	276	235	41	173
Cherokee County Convict Camp, Canton	5	5	......	......	......	......	5	8	8	......	5	5	......	3
Clarke County Chain Gang, Athens[2]	......	......	......	......	......	......	......	......	......	......	......	......	......	......
Clay County Chain Gang, Fort Gaines	10	10	......	1	1	......	9	9	9	......	......	......	......	9
Clayton County Chain Gang, Jonesboro	24	23	1	4	......	......	20	36	36	......	16	16	......	20
Clinch County Convict Camp, Homerville	23	23	......	3	3	......	20	32	32	......	1	......	......	31
Clinch County Jail, Homerville	......	......	......	......	......	......	......	1	1	......	1	1	......	......

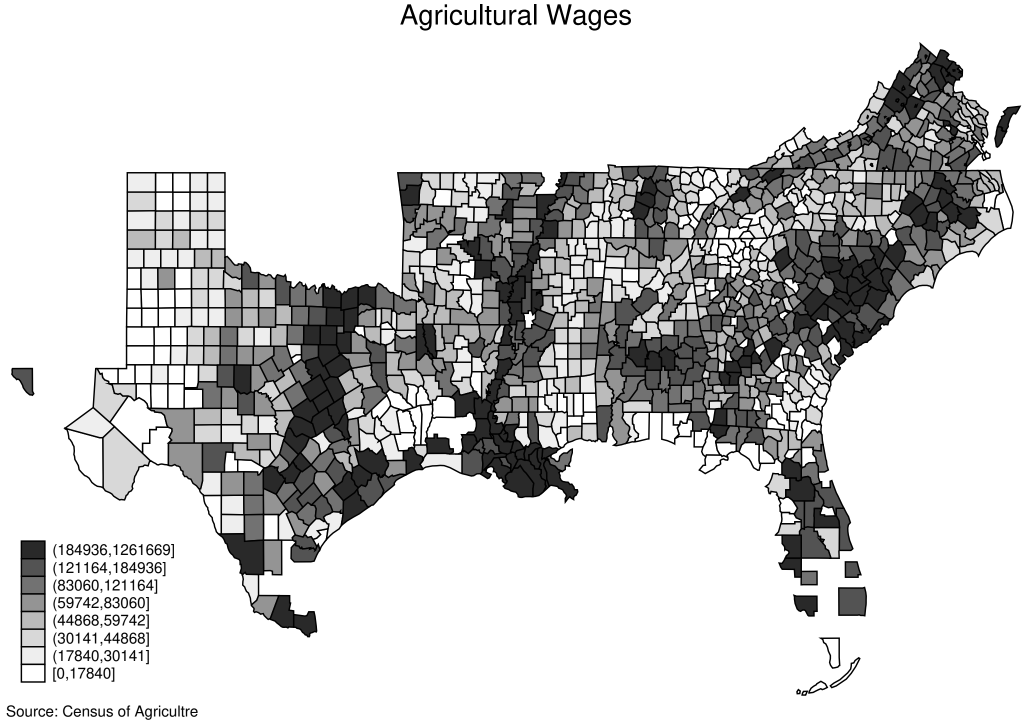
Figure A2: Appendix: Example of Convict Camps Records from the Department of Labor

TABLE I.—CONVICTS BY STATES AND TERRITORIES.

	Institution.	Location.	Official control.	System of work.	Industry.	Contractors or lessees.
	ALABAMA.					
1	State Penitentiary	Wetumpka ...	State	Lease.....	All industries	5
					Farming........................	....
					Mining, coal	....
					Stone, broken.................	....
2	Autauga Co. Jail..	Prattville.....	County ...	Lease.....	Lumber	1
3	Baldwin Co. Jail ..	Daphne.......	County ...	Lease.....	All industries	2
					Lumber	....
					Mining, coal	....
4	Barbour Co. Jail ..	Clayton.......	County ...	Lease.....	All industries	2
					Farming........................	....
					Mining, coal	....
5	Bibb Co. Jail	Centreville ...	County ...	Lease.....	Farming........................	1
6	Blount Co. Jail ...	Blountsville ..	County ...	Lease.....	Mining, coal	1
7	Bullock Co. Jail ..	Union Springs	County ...	Lease.....	All industries	2
					Farming........................	....
					Mining, coal	....
8	Butler Co. Jail....	Greenville	County ...	Lease.....	Lumber	1
9	Calhoun Co. Jail..	Jacksonville..	County ...	Lease.....	Mining, coal	1
10	Chambers Co. Jail	La Fayette. ..	County ...	Lease.....	Farming........................	1
11	Cherokee Co. Jail	Centre........	County ...	Lease.....	Mining, coal	1
12	Chilton Co. Jail...	Clanton.......	County ...	Lease.....	Lumber	1
13	Choctaw Co. Jail	Butler	County ...	Lease.....	All industries	2
					Farming........................	....
					Mining, coal	....
14	Clarke Co. Jail....	Grove Hill....	County ...	Lease.....	Mining, coal	1
15	Cleburne Co. Jail.	Edwardsville .	County ...	Lease.....	All industries	2
					Farming........................	....

Note: The table corresponds to the year 1886. These records provide information on all convict camps in the US.

Figure A3: Appendix: Agriculture wages



Note: The maps plots the agricultural wages in 1910, and shows that higher wages were concentrated in previous slavery areas. Source: Census of Agriculture.

Table A3: Agricultural wages and slavery

Dependent variable Log of agricultural wages	(1)	(2)
Slavery	0.025***	0.025***
	(0.001)	(0.001)
Observations	7058	7058
R-squared	0.197	0.201
Controls	No	Yes

Note: The table shows the impact of slavery on agricultural wages.

Chapter II

The Effect of Violence on Social Capital: Evidence from Exogenous Variation in an Illegal Market

Melissa Rubio*

Abstract

This paper studies the effects of violence on social capital using individual and municipal level data in Colombia. To estimate causal effects, I exploit changes in violence attributed to cross-border shocks on coca markets in neighboring countries, interacted with a novel index of suitability for coca cultivation; this resulted in greater violence in municipalities that are more suitable for coca production. I find that violence has a negative effect on social capital measures such as trust, participation in community organizations, and cooperation.

*University of Gothenburg (melissa.rubio@economics.gu.se). I am extremely grateful to Randi Hjalmarsson for her invaluable advice and guidance. Special thanks to Anna Bindler, Mikael Lindahl, Ola Olsson, William Maloney, Maria Lombardi, Juan F. Vargas, Alessandro Tarozzi, Laura Mayoral, Leopoldo Fergusson, Adriana Camacho, and seminar participants at the University of Gothenburg, University of los Andes, Berkeley, Stockholm, Göttingen, and Naples for useful comments and discussions. All error are mine.

"And how sad, because we were afraid of our friends. No one knew whether someone else was a crook. That screwed us up... I am still very afraid of people". [1]

1 Introduction

Conflicts have devastating effects on economic development (Collier *et al.* , 2009). However, they not only impose direct costs on society through the destruction of physical and human capital; conflicts can also lead to social and political disintegration (Collier *et al.* , 2009, Rohner *et al.* , 2013b). This paper investigates the causal effects of conflict-related violence on social capital outcomes such as trust, participation in community organizations, and cooperation, which have been proven crucial in sustaining economic outcomes (Algan & Cahuc, 2010, Rohner *et al.* , 2013a, Guiso *et al.* , 2006).[2]

Colombia offers an ideal setting to study this question for two reasons. First, most of the currently available evidence is based on post-conflict settings, and this limits our knowledge with respect to the impact of ongoing confrontations on local institutions. However, Colombia collected data on social capital measures during times of conflict, and therefore the results of this study are useful for understanding the effects of conflict in other developing countries. Second, there is wide variation in local state capacity (Acemoglu *et al.* , 2015), and as a result, some areas are very reliant on their communities. For instance, 59% of the households facing adverse shocks receive help primarily from networks of friends and relatives (Cadena & Zuluaga, 2012), providing an idea of the relevance of social capital when formal institutions work poorly. Finally, violence varies substantially over time and among municipalities across the country.

The key econometric challenge to my analysis is that violence might be endogenous. On the one hand, social capital levels could affect the intensity of conflict, leading to reverse causality concerns. This is the case if well-organized communities protect themselves from violent attacks through neighborhood watch schemes as was the case in some Colombian municipalities (Kaplan *et al.* , 2010), or if armed groups target communities with low social capital to take control over territories more easily. On the other hand, the correlation between conflict and social capital can be driven by omitted variables (i.e. the large variation on institutional quality), which have a confounding effect, rather than reflecting a causal impact. In either case, the estimates of conflict on social capital will likely be biased.

[1] Testimony of a victim of conflict in Segovia, Antioquia (Colombia). (de Memoria Historica, 2013), Pg 274.

[2] For instance, social networks not only provide support during adverse situations (Foster & Rosenzweig, 2001; Fafchamps & Lund, 2003), but also guarantee a more efficient provision of public goods (Nannicini *et al.* , 2013; Glennerster *et al.* , 2013), and better outcomes in terms of fiscal capacity (Guiso *et al.* , 2004), governance (Aghion *et al.* , 2010), trade (Cassar *et al.* , 2013), and diffusion of knowledge and technologies (Conley & Udry, 2010; Bandiera & Rasul, 2006; BenYishay & Mobarak, 2014).

To overcome these empirical challenges, I exploit two different sources of variation in violence. The first is the suitability for growing coca in Colombian municipalities, as armed groups finance their fighting with coca production. The second is external shocks to coca markets in Peru and Bolivia. Together with Colombia, these two countries are the main coca leaf producers in the world (UNODC, 2009). Thus, changes in the availability of coca leaves in neighboring countries affect the demand for coca in Colombia, and in turn the intensity of violence. Mejia & Restrepo (2015) argue that coca plantations lead to violence because property rights need to be protected from other armed groups, and because violence is required to control vast territories and communities living in coca areas.[3] Since it is expected that these shocks affect coca producer municipalities differently than non-producers, I use satellite data on ecological conditions to construct an exogenous index for coca cultivation at the municipal level. Using the interaction between cross-border shocks on coca markets and the coca index as an instrument has two main advantages: *i)* it provides random variation in violence over time, and *ii)* it allows for different trends of violence across municipalities.

The validity of the instrument hinges on one critical assumption: cross-border shocks to the coca market did not affect municipalities that were more suitable for coca production in other unobservable dimensions. I address this concern by showing that the results are robust to the inclusion of controls for employment, pupils registered at school, tax collection, and forced displaced population. In addition, I present several checks that suggest that coca production has an effect on social capital only through the channel of violence.

The data consist of three separate datasets. The first is sub-national panel data from the research center *CEDE* at Universidad de los Andes, which is a collection of official data on violent events perpetrated by non-state armed groups. The second corresponds to survey data from *LAPOP* (Latin American and Public Opinion Project) for the 2004-2011 period, a repeated cross-section including information on various measures of social capital and socioeconomic characteristics at the individual level. Finally, the data used to construct the instrument come from *UNODC* (United Nations Office on Drugs and Crime), which reports information on coca cultivation for Latin America. These data are combined with ecological information to estimate the determinants of each municipality's suitability for coca cultivation.

[3]Similar examples can be found in the literature. Angrist & Kugler (2008) show that aerial interdiction campaigns in Peru and Bolivia led to an increase in the demand for coca cultivation in Colombia. Traditional coca growing regions experienced an increase in coca cultivation and subsequently became more violent by increasing the resources available to insurgent groups. Dube & Vargas (2013) exploit exogenous price shocks in international commodity markets. They find that a sharp fall in coffee prices during the 1990s lowered wages and increased violence differentially in municipalities cultivating more coffee, whereas a rise in oil prices increased violence differentially in the oil region by increasing revenues to be appropriated by rebel groups. Rohner *et al.* (2013a) use an external political shock when the US declared the main rebel movements of Uganda to be terrorist organizations, affecting the intensity of fighting along the Sudanese border.

I find that violence has a negative and statistically significant effect on different measures of social capital. The estimated effect is quantitatively large and robust to alternative explanations of migration, spillover effects, and economic activity. A one-standard deviation increase in the rate of violence is associated with a 38%, 22%, and 23% standard deviation decrease in trust, participation in community organizations, and contribution to solving problems, respectively. The magnitude of the effect is relatively large; taking trust as an example, it corresponds to the difference in trust levels between Germany and Colombia.[4] These results seem to be driven by different factors. First, as in other non-related conflicts, in Colombia it is not possible to apply basic cues to identify friend from foe, and people can be especially wary about whom they can trust. Therefore, results are stronger for more heterogeneous municipalities. Second, and consistent with the story of warring factions not being readily identifiable, I present evidence that people are afraid to interact with their communities. They report fear to run in local elections, participate in community organizations, and vote in elections.

This work is related to a growing body of research that documents the legacy of violence on social capital, which has been described as the "the most important but least understood of all war impacts" (Bauer *et al.* , 2016) (pg 42). One group of studies points to negative consequences of conflict on social capital. Rohner *et al.* (2013a) show that exposure to conflict decreased trust in Uganda. Cassar *et al.* (2013) document low trust and willingness to engage in impersonal exchanges after the Tajik war.[5] Yet, there are some instances in which individuals exposed to war-related violence tend to increase their social participation by joining local groups or by taking on more leadership roles in their communities (Bellows & Miguel (2009), Voors *et al.* (2012)).[6] My paper contributes to this debate by investigating the *causal* effects of conflict on social capita, whereas most of the previous studies mainly show correlations.[7] In addition I provide empirical evidence of positive and negative effects, depending on the intensity of violence. These findings help to reconcile the apparently contrasting legacies of conflict presented in the literature.

The remainder of the paper is organized as follows. The next section summarizes the context of the Colombian conflict. Section 3 provides an overview of the conceptual framework. Section 4 describes the data. Section 5 presents the identification strategy. Section 6 discusses the results, the robustness analysis, heterogeneous effects, and an extension on the

[4]World Value Survey (2009).

[5]In a historical setting, Besley & Reynal-Querol (2012) find that precolonial conflict in Africa is negatively related to current levels of trust. Nunn & Wantchekon (2011) show that contemporary differences in trust levels within Africa can be traced back to slave trade. The reason is that slaves were captured primarily through state organized raids and warfare, but as trade progressed, the environment of ubiquitous insecurity caused individuals to turn on others—including friends and family members—and to kidnap, trick, and sell each other into slavery.

[6]See Bauer *et al.* (2016) for a complete survey.

[7]The exception is Rohner *et al.* (2013a), they provide causal evidence of the effect of conflict on social capital measures in Uganda.

potential consequences of low social capital on political participation. Section 7 concludes.

2 Background

2.1 The Colombian Conflict

The Colombian conflict dates back to the 1960s. This conflict involves three illegal armed groups that have competed for the control of villages, natural resources, and strategic corridors of illegal markets. These non-state armed groups reigned over most of the territory and though estimates vary, may have had around 50,000 men and women under arms at the beginning of the 21st Century (Acemoglu *et al.* , 2013). Two groups are left-wing guerrillas: the Armed Revolutionary Forces of Colombia (FARC) and the National Liberation Army (ELN). Both have fought with the stated aim of overthrowing the democratic government and claim to represent the rural poor (Dube & Vargas, 2013). Both guerrilla groups have been involved in the illegal cocaine business. After the demise of the Medellin and Cali cartels by the mid-nineties, armed groups started to participate in the production and commercialization of cocaine for financing insurgent activities against each other and against the Colombian government (Suarez, 2000). Even though revenues from underground economies are extremely difficult to measure, estimates of the guerrillas' income are about 1 billion dollars per year (Otis, 2014). The third actor is a right-wing paramilitary group known as United Self-Defense of Colombia (AUC), which emerged as an anti-insurgent self-defense group organized by rural landowners and drug barons in response to guerrilla extortions.

In the late 1990s and early 2000s, much of the fighting between guerrillas and paramilitary forces was for control over coca plantations and trafficking routes for cocaine (Bagley, 2012). However, armed groups also participate in the production chain. In particular, they process the leaf into cocaine in laboratories near plantations and sell it to international traffickers, obtaining profits from its trade.[8]

2.2 Armed Groups and Their Relationship with the Civil Population

Besides engaging in drug trafficking, all the armed groups have been accused of human rights violations. The multi-party nature of the conflict and its intensity has resulted in civilians being victims of cross-fire, threats, kidnappings, massacres, bombings, forced recruitment of minors, and extortion. In terms of bellicose activity, the most common guerrilla actions are the disruption of the economic infrastructure (e.g, attacks to oil pipelines), attacks to government military positions, bombings and roadblocks (Vargas, 2009), as well as kidnappings, and extortions. In contrast, paramilitary forces assaulted civilians through selective

[8] See Rico (2010) for a thorough description of the process of coca cultivation and cocaine production in Colombia, and Mejía and Posada (2008) for a description of wholesale cocaine markets.

killings, kidnappings of community leaders and threats to peasant organizations whom they presumed to support rival groups. The paramilitary groups publicly claimed that at least two-thirds of the guerrilla members were civilian supporters rather than proper combatants, so their priority was to block the "non-uniformed guerrilla" (Molina & Castaño, 2001). In that sense, civilians were targeted and persecuted as a deliberate strategy of war, often accused of aiding enemy groups (Morales, 2018).[9] As a result, because of rumors, false denouncements and finger-pointing, people have developed mechanisms of protection such as silence, distrust and isolation from their community. People opted to take up a "low profile" strategy (de Memoria Historica, 2011, 2013).[10]

Similarly, prominent community members have been victimized for being the spokespersons for collective claims and, over the last two decades, 1227 local leaders, 1496 politicians, 1287 public servants, and 74 human rights defenders have been killed, while 8000 leaders have been threatened (Department of Justice, 2017). In 2004, on the Caribbean coast, women leaders were systematically attacked. Similarly, the guerrillas declared any political state representative a military target, and many candidates were forced to renounce or govern their towns from other cities (de Memoria Historica, 2013).

3 Conceptual Framework

Social capital has been defined in the literature as the ability of a society to foster trust and cooperation among its members.[11] In this section, I discuss different reasons to expect violence to affect social capital. The first is rooted in the neoclassical theory, by which greater cooperation arises from a higher value of social insurance. Violence destroys household assets and therefore victims become more dependent on local systems of risk-sharing. During wartime, investments in physical and human capital can be too risky such that pro-social behavior becomes the optimal choice as a form of self-protection mechanism (Bauer *et al.* , 2016).[12] Jennings & Sanchez-Pages (2017) formalize this idea in a theoretical model in which communities that face an external threat use social capital to protect themselves. Case studies from political science provide evidence on the plausibility of this argument, as social cohesion enables civilians to overcome fear, implement collective strategies for protection, and denounce aggressions (Kaplan *et al.* , 2010).[13]

[9]Kaplan (2013) discusses the possibility of civilian collaborators in the Colombian conflict. However, unfortunately data on suspects aiding armed groups are only data available for one municipality in Colombia. This information was recovered based on interviews because there were no written records, so it is likely that there is measurement error. The interviews suggested 67 cases of civilians helping armed groups in a village of 5,000 inhabitants.

[10]The Centro de Memoria Histórica is an official institution belonging to the Colombian government that has the purpose to document testimonies related with the armed conflict.

[11]For example, see Homans (1958), Coleman (1994), Putnam *et al.* (1993); Putnam (2020).

[12]It has to be noted that there is no empirical study supporting this theory.

[13]For instance, organized communities in The Philippines stayed out of the conflict between the military and rebels. In Guatemala Communities of Populations in Resistance against conflict were created (Hancock & Mitchell, 2007). In Peru, the Peasant Rounds were originally formed as a protection force against theft

A second possibility cited by the psychology literature is related to a phenomenon called post-traumatic growth. The idea is that victims of traumatic experiences report greater value to personal relationships. Nevertheless, other studies have documented that violence is linked to depression and distress, including a lack of desire to engage with people and difficulty in maintaining close relationships (Ehlers & Clark, 2000). In line with these studies, Alesina & Ferrara (2002) document that a personal trauma such as a natural disaster or divorce reduces social capital in the United States. Specifically for the Colombian case, Moya (2018) has documented that traumatic conflict-related experiences can alter individuals' behaviors and deplete their ability to make economic decisions because of severe anxiety disorders. Therefore, one could expect that social capital is also negatively affected.

Another explanation relies on the parochialism theories that point to generosity towards insiders and selfishness towards outsiders who represent a threat (Choi & Bowles, 2007). The prediction from this approach is that inter-group competition, including war, will promote individuals' pro-social behavior toward in-group members, compared to out-group members. This theory underscores the importance of the in-group's boundary relative to the out-group as a crucial feature for determining the consequences of conflict on social capital (Cassar *et al.* , 2013). Therefore, the experience of a civil war that pits one group against another might strengthen the within-group prosociality, while corroding the between groups social capital (Bauer *et al.* , 2016).

Along these lines, homogeneity within ethnic groups allows individuals to find a common ground and a rational basis for coalition (Esteban & Ray, 2011). Ethnic boundaries based on physical differences are easier to police than boundaries based on non-visible differences; this makes such boundaries a low-cost sign of intentions, since they can be used as a marker to recognize potential infiltrators, and as an effective way of enforcing group membership (Caselli & Coleman, 2013; Chandra, 2007). However, for the Colombian context such distinction is not possible as there is no polarization among religious, regional or ethnic divisions. Most of its people do not belong to an ethnicity – 90% of the population is not identified with a specific ethnic group (General Census of Population 2005). Regarding armed groups, they were formed by a military and an urban militia section. The former used camouflaged uniforms, bracelets with the group insignia, and were heavily armed. The latter did not carry visible guns and dressed like the civil population (CMH Putumayo, 2010). As such, it is not possible to distinguish if another individual is affiliated with an armed group or not, and which armed group. There are numerous examples of the "not readily identifiable" aspect of the conflict. Many victims reported later on that their perpetrators belonged to their community, showing that a large part of the conflict was within-communities, and that

(Starn *et al.* , 1999; Fumerton, 2001). In El Salvador, people tended to join and support rebel movements in response to government violence against them or their family (Wood, 2003).

groups lived together (de Memoria Historica, 2013).[14] Testimonies relating how difficult it was to distinguish criminals from civilians were common:

> *"since they were in civilian clothes, many times we did not know who was a guerrilla member"*[15]

As a consequence of all the above, I hypothesize that the inability to distinguish friend from foe in a violent environment may have hurt trust in communities. The Colombian conflict provides a framework where insurgents were among the civil population, leading to people avoiding involvement with the community to minimize the risk of being targeted. This argument could be potentially applied to other contexts where it is hard to distinguish the antagonist group.

4 Data

4.1 Social Capital

The measures of social capital come from the LAPOP survey (Latin American Public Opinion Project), conducted by the Americas Barometer[16], a nationally representative survey for individuals over 18 in rural and urban areas. It contains information on political attitudes for approximately 12,000 individuals in 55 out of 1,122 municipalities in Colombia, for the 2004-2011 period, as a repeated cross-section (there are on average 1,400 individuals per survey round). These municipalities were selected to be representative of the country based on socio-economic characteristics and population size.[17]

Social capital includes features of social organization, such as social networks, norms, and trust that facilitate coordination and cooperation for mutual benefit (Putnam *et al.* , 1993). In this paper, I measure social capital as trust in other members of the community, and participation in community organizations. These measures are an indicator of collaboration within communities and the collective ability to respond to adverse situations (Durlauf & Fafchamps, 2005; Colleta & Cullen, 2000), and they have been used in other studies on conflict. For instance, Rohner *et al.* (2013a) use the same survey design employed in this paper, but applied to the Ugandan case. Bellows & Miguel (2009) ask for interpersonal trust, attendance to community meetings in the last year, membership in local groups, participation in elections, and some questions regarding political knowledge such as the name

[14] The New York Times wrote a piece in which tells the history of a town where neighbors were perpetrators.

[15] Testimony found in CMH, Putumayo (2010).

[16] Americas Barometer selects the samples in following way: each sample is a nationally representative cross-section of all citizens of voting age obtained by (a) strictly applying random selection methods at every stage and by (b) applying sampling with probability proportionate to population size. The sample is stratified by key social characteristics in the population such as sub-national area (eg. region/department) and residential locality (urban or rural) (LAPOP, 2004-2011). Available for download here.

[17] The data include the five main cities of Colombia as well as small towns. The average population for the small towns is 140,000 inhabitants, whereas for the big cities is 4,600,000 inhabitants.

of the local councilor. The main variables are constructed using the following questions from the LatinBarometer:

- *Trust*: "How much do you trust people from your community?". I code the variable as one if the respondent answers either "I trust them a lot", or "I trust them somewhat". Otherwise, the value assigned is zero as in Rohner *et al.* (2013a). (Question $it1$).

- *Participation in community organizations*[18]: "Did you attend community meetings in the last year?" I code the variable as one if the respondent reports to have attended to at least one community meeting in the last year as in Bellows & Miguel (2009). (Question $cp8$).

- *Cooperation*: "Have you contributed to the solution of a problem in your community or from your neighbors?" It takes the value of one if the person has contributed to solving problems in his commuity, otherwise the value is zero (Question $cp5$).

In addition, LAPOP collects detailed information on the socioeconomic characteristics of surveyed individuals including age, sex, household income, and years of education. The survey also asks about participation in local and national elections, and media consumption (radio, TV, newspapers and Internet). Descriptive statistics for these variables are in Panel A and B of Table 1. On average, around 70% of individuals trust people from their community; 88% have participated in a community meeting during the last year; and, regarding socioeconomic characteristics, half of the population are women, one quarter lives in rural areas, they are aged around 37, and 6% are black.[19]

4.2 Armed Conflict

Data on the armed conflict was taken from *Centro de Estudios sobre Desarrollo Económico* (CEDE) at the School of Economics at Universidad de Los Andes in Colombia. CEDE collects data from the Observatory of Human Rights of the Vice-presidency and the National Planning Department. The original data are a compilation of reports of the national police.[20]

The data set codes different violent events by municipality location and groups involved. For this study, I construct a measure of violence that aggregates the number of attacks by rebel groups at the municipality-year level; this measure is normalized by population to

[18]The purpose of the community organizations is addressing general problems in the community (for instance, improving education, agriculture practices, security and the provision of public goods).

[19]The descriptive statistics are consistent with the ones presented by the National Department of Statistics in Colombia, DANE, so the sample is comparable to the average in the main characteristics with the rest of country.

[20]Martínez (2017) shows that these variables are consistent with a dataset produced by CERAC, a Colombian think-thank that collects information from national and local newspapers and complements it with reports from nongovernmental organizations working in remote areas.

create a rate of violence per 100,000 inhabitants for the period 2004-2011.[21]

Descriptive statistics are presented in Panel C of Table 1. The violence rate that includes all types of attacks from armed groups is 49. To provide an idea of the magnitude of this rate, it has to be noted that the average homicide rate for the period 2004-2011 is 38,5. This reveals that Colombia is a very violent country when compared to other countries with similar levels of development. For instance, the average homicide rate in Latin American countries is 25 per 100,000 inhabitants. For Honduras, it is 58; Brazil, 23; the United States, 5; and Sweden, 1.[22]

Regarding external validity, the sample of municipalities for which there is data on social capital exhibits a slightly lower rate of violence compared to the whole country. Figure 1 shows the rate of violence for both samples using the same source of data for the period studied. It can be seen that they follow the same pattern (they increase and decrease during the same years). Figure 2 presents the raw relationships between the different measures of social capital and violence. In general, there is a negative correlation between the measures of social capital and the rate of violence, but these cannot necessarily be interpreted as causal because of the identification challenges discussed in Section 5.

4.3 Coca Suitability Index

For each municipality, I have collected data on ecological and geographic characteristics from different sources. Altitude comes from the U.S Geological Survey Center, and soil pH from the Food and Agriculture Organization of the United Nations (FAO). Data on temperature, precipitation, humidity and solar radiation were taken from the Global Climate Data. These measures are available at 30 seconds or 0.0083 degrees spatial resolution, which is equivalent to approximately one square kilometer. They correspond to their average value between 1970 and 2000.[23] I think of these variables as predetermined fixed municipality characteristics, determined mostly by ecological events unrelated to the rise of coca cultivation and production. Additionally, I have included coca cultivation data for Peru and Bolivia, obtained from satellite images and used to create the external shocks to the Colombian coca market, which I will explain in detail in the next section. Data for coca cultivation in Colombia are available at the municipal level for the period of this study from the United Nations

[21] In particular I add 19 violent events to construct the rate of violence: confrontation between illegal armed groups and the state military forces, explosive terrorist attacks, incendiary terrorist attacks, attacks on police stations, attacks to private property, attacks against institutions, general attacks, general confrontations, incursions to villages, road blocks, air attacks on aircrafts belonging to the state military forces, ambushes on military/police cars, kidnappings of civilians, kidnappings of military forces, kidnappings of politicians, killings of civilians, killings of politicians, mass murders.

[22] UN Office on Drugs and Crime's International Homicide Statistics database.

[23] I have also used the median value of this measures and the index does not change much, municipalities that were more suitable for growing coca continue being more suitable for growing coca when using the median. More detailed information for the sources and construction of the variables is provided in Appendix B.

Office on Drugs and Crime. The UN monitors coca cultivation by using various types of satellite images covering the whole national territory (equivalent to 1,142,000 square kilometers). One of the major difficulties in data acquisition is the frequent cloud cover over the Colombian territory. Therefore, they use satellite with a frequent view and continuous recording (UNODC, 2009).[24]

5 Empirical Strategy

In this section, I first introduce the baseline estimating equation (Section 5.1) and construct the instrument for violence (Section 5.2). Finally, I present evidence to support the identifying assumptions (Section 5.3).

5.1 Baseline Estimating Equation

The goal of this paper is to investigate the effect of conflict on social capital. To do so, I consider the following benchmark econometric model in a repeated cross-section setting:

$$P(SocialCapital_{imt} = 1) = \beta_0 + \beta_1 Violence_{mt} + \beta_2 X_{imt} + \beta_3 Z_{mt} + \gamma_m + \delta_t + u_{imt} \qquad (1)$$

where $SocialCapital_{imt}$ is the outcome for an individual i, in municipality m, and year t. X_{imt} includes a set of individual sociodemographic variables (gender, schooling, race, income, and media consumption).[25] Z_{mt} are time-varying municipality controls, including: employment rate, pupils registered at school, tax collection, and forced displaced population. γ_m and δ_t are municipality and year fixed effects, implying that β_1 is estimated from changes in the rate of violence within the same municipality over time, compared to other municipalities in a given year. Thus, any confounding variable that has a common effect on social capital across all municipalities in the same year, such as political changes, or characteristics of municipalities that have a constant effect on social capital over time are controlled for. Standard errors are clustered at the municipality level.

5.2 Instrument for Violence

When estimating the causal effects of conflict on social capital, three primary identification challenges emerge. First, it is possible that the levels of social capital in a municipality determine the intensity of violent attacks. This raises concerns regarding reverse causality, as violence would be the consequence rather than the cause of changes in social capital. The second challenge is related to omitted variables bias, where causation cannot be disentangled from correlation if unobservable variables determine both the exposure to conflict and social

[24] The satellite has a 16-day repeat cycle, which enhances the chance for cloud free images (UNODC, 2009).

[25] Olken (2009) documents that increased signal reception in Indonesia, leads to more time watching television and listening to the radio, which is associated with less participation in social organizations and with lower self-reported trust.

capital. For instance, armed groups might be present in areas considered strategic for political reasons or valuable resources (Acemoglu *et al.* , 2013). In either case, OLS estimates will likely be biased. A war strategy could be to attack areas without community organizations to avoid civilian resistance movements (Kalyvas, 2006), which will overestimate the effect of conflict on social capital. Similarly, one can underestimate the impact of conflict if communities that are better organized are more likely to be attacked. Finally, there could be measurement error in violence, which could lead to attenuation bias.

To reduce such endogeneity concerns, I use an Instrumental Variable (IV) strategy. In particular, I take advantage of the exogenous variation in Colombian violence attributed to the eradication of coca plants in Peru and Bolivia, as the raw input required for cocaine production -the coca leaf- is only produced in these three South American countries. The eradication of coca crops is the main strategy to reduce the supply of cocaine in these countries.(UNODC, 2009)[26] Given that these external shocks affect violence disproportionately in municipalities with a potential for coca cultivation, I construct a suitability index. The index indicates which municipalities are suitable for coca cultivation based on particular ecological and geographic conditions. Therefore, $Violence_{mt}$ is instrumented for with external shocks to the Colombian cocaine market $(ln\, External\, Shock_t)^{27}$, interacted with a suitability index for producing coca $Coca\, Index_m$. Thus, I can capture exogenous variation in violence across municipalities and time as follows:

$$Violence_{mt} = a + \alpha(ln\, External\, Shock_t) * Coca\, Index_m + \lambda_2 X_{imt} + \lambda_3 Z_{mt} + \gamma_m + \delta_t + v_{mt} \quad (2)$$

In particular, these two sources of variation allow me to compare high and low coca suitable municipalities for years with high and low demand for Colombian coca. Therefore, violence is identified from within municipality variation by comparing municipalities with different suitabilities in years with different external shocks to the coca market. Given this, the first stage can be interpreted as the differential impact on violence of external shocks for municipalities that are more suitable for coca production. Put differently, the suitability index determines the intensity of the treatment (violence), as the external shocks affect the whole country. Therefore, coefficient α should be interpreted as the differential effect of an increase of 1% in external eradication efforts. Now I turn to describe the construction of the instrument.

[26] The main ingredient for the production of cocaine is cocaine alkaloid, a chemical compound that can be extracted from the leaves of coca plants. Cocaine production is a relatively simple process that can take place in small local workshops. The process of producing cocaine consists of three main steps: after being harvested and dried, coca leaves are converted into coca paste, then into cocaine base, and then into the final product, cocaine hydrochloride. The manufacturing process requires a few chemicals (precursors) such as sulfuric acid, potassium permanganate, hydrochloric acid, acetone, and ethylther, plus water, filters, and microwave ovens.

[27] The effect of the external shock is expected to be contemporaneous on violence, since coca bushes can be grown and harvested year-round, on average, four times a year (minimum three and maximum eight, depending on the variety) (Mejia & Posada, 2008). I use the logarithm of the external shocks because the effect can be interpreted as the increase in 1% in eradication efforts on the violence in Colombia.

5.2.1 External Shocks to Coca Markets

External shocks to the Colombian coca market come from calculating the extent of coca plantations that have been eradicated from the total amount of coca planted in Peru and Bolivia in a given year t as follows:

$$External\,shock_t = \frac{km^2\,of\,coca\,eradicated\,in\,Peru\,and\,Bolivia_t}{km^2\,of\,coca\,cultivated\,in\,Peru\,and\,Bolivia_t}$$

Figure 3 shows the rate of violence between 2004 and 2011, and the eradication efforts in Peru and Bolivia. Consistent with the narrative above, high levels of violence are statistically positively correlated with more eradication efforts in neighboring countries, especially for municipalities that are more suitable for coca production.

5.2.2 Constructing the Coca Index

The cross-sectional variation underlying the instrument in equation (2) is based on the idea that coca plantations are sensitive to weather and environmental conditions, which make coca cultivation exclusive to certain areas of Andean countries. For instance, if the temperature is too low, the plant does not grow, and if the temperature is too high, the leaves become very dry and lose their strength. The ecological literature has documented the relevant conditions for growing coca. Thus, with respect to temperature, this paper relies on the findings of Acock *et al.* (1996), who show that the optimum average daily temperature for leaf growth ranges between 20°C and 30°C, which is the temperature range I consider in this study. Plowman (1979) finds that coca develops in humid tropical climates in altitudes between 300 to 2000 meters above the sea level. Johnson *et al.* (1997) report that the relative humidity ranges between 55 and 85%, and also provide evidence that coca plants grow better in soil with a pH level lower than 6. In addition, coca plants need a certain amount of daylight measured by the PPFD (Photosynthetic Photon Flux Density). These conditions are summarized in Table A1. Based on these findings, the following equation is used to construct a novel measure for coca suitability:

$$coca\,index_m = \frac{\sum_{g=1}^{G_m} \mathbb{1}[300 \leqslant alt_{gm} \leqslant 2000 \wedge 20 \leqslant temp_{gm} \leqslant 30 \wedge 500 \leqslant precip_{gm} \leqslant 4000 \wedge 6 \leqslant PH \wedge 55 \leqslant Rh_{gm} \leqslant 85 \wedge PPFD \leqslant 400]}{G_m}$$

where $\mathbb{1}[\cdot]$ is an indicator function for the optimal conditions to produce coca. It takes the value of 1 when the grid g in municipality m satisfies the requirements established by the ecological literature, otherwise the value is zero. These values are added and divided by the total number of grids G in a particular municipality.[28] This creates an index between 0 and 1 for each municipality. Intuitively, it estimates the share of a municipality that is suitable for coca cultivation. A municipality with an index of 1 is completely suitable, and a

[28] The grid size is 0.0083 degrees in spatial resolution or approximately $1km^2$, the finest possible cell disaggregation, allowing me to calculate a more precise estimate of the share of a municipality that is ecologically suitable for coca production.

municipality with an index of 0 is not. Figure 4 illustrates an example of how municipalities are divided into grids to calculate whether each fulfills the ecological conditions for coca cultivation, in which case the grid is shaded.

Figure 5 displays the distribution of the index for the whole country on the left and for the sample of municipalities used in this paper on the right. Overall, the shape of the distribution of the index is comparable among samples. The main difference is in the tails, as the proportion of low suitable municipalities is slightly higher in the analyzed sample than it is for the whole country.[29] Figure 6 maps the coca index and shows that Colombia is a good case for studying violence driven by illegal crops since coca suitability is not isolated to a particular region. In fact, 50% of municipalities in Colombia are classified as potential coca producers.

Finally, I demonstrate that this index is indeed predictive of the location of coca crops during the period of interest. Table A2 provides support that findings in the ecological literature, namely, specific weather and soil conditions determine the availability of coca. The first three columns show different measures of the coca index for the whole country: *i)* the index described in this section that goes from 0 to 1, *ii)* as a dummy variable above the median and, *iii)* in terms of standard deviations above the mean. In all of the cases, the index predicts the amount of coca planted in a municipality positively. The last three columns present a similar story but for the analyzed sample. The correlations are again positive but less statistically significant compared to the whole country, perhaps because the sample consists of 55 municipalities. In particular, to facilitate the interpretation, a municipality that is one standard deviation above the coca index mean (column 6) cultivates two hectares more of coca plantations. By 2011, there were 64,000 hectares of coca planted in Colombia (UNO, 2012).

5.3 Assessing the Instrument

This subsection discusses the identifying assumptions of the instrumental variable design, which uses the interaction between external shocks to the coca market at the year-level with the suitability for growing coca at the municipality-level.

For this to be a good instrument, it needs to be relevant and valid. I begin with an intuitive and anecdotal justification of these assumptions. More formal tests are presented in Section 5.3.1 and 5.3.2. Mejia & Restrepo (2015) provide evidence that coca plantations lead to violence for two reasons: first, property rights need to be protected from other armed groups. Second, violence is required to control vast territories and communities living

[29] The main reason it is because vast territories of the Amazon are not included in my sample. However, there are few remotes villages in these territories.

in coca areas.[30] Thus, any shock to the Colombian coca market will generate changes in violence as cocaine demand is highly inelastic (Saffer & Chaloupka, 1999). A decrease in supply leads to a substantial increase in price, which in turn creates greater incentives for producing cocaine for municipalities where coca can be planted.[31] The relevance of this instrument is supported by the findings of Angrist & Kugler (2008), who show that aerial interdiction campaigns in Peru and Bolivia in 1994 led to an increase in the demand for coca cultivation in Colombia. Traditional coca growing regions experienced an increase in coca cultivation and subsequently became more violent by increasing resources available to insurgent groups.[32] However, I will use the exogenous suitability for producing coca, instead of coca planted, since most of the cultivation takes place in areas that lack state infrastructure leading to endogeneity concerns.[33] Moreover, I focus on a suitability index that does not change over time, because climatic shocks could potentially affect violence by mechanisms that also affect social capital.[34] For example, weather shocks could directly influence the feasibility of community gatherings. Such explanations can be ruled out given that local weather can be directly controlled for. For this reason, the index is based on predetermined or historical weather conditions.

Furthermore, the benefit of this approach is that it relies on ecological expertise. These conditions are in turn determined by environmental interactions of temperature, humidity, solar radiation, soil nutrients and vapor pressure that are captured in the coca index. Hence, given a set of time and municipality fixed effects, this measure combined with the exogenous shocks to the coca market is arguably an exogenous determinant of violence. Additionally, social capital measures are not likely to be affected by exogenous shocks to coca markets in neighboring countries, as they are determined by foreign policies.

[30] One hectare (10,000 m^2) of coca bushes produces, on average, between 1000 and 1200 kilograms of fresh coca leaf and 1 kilogram of cocaine (Mejia & Posada, 2008). Whereas, for instance, for coffee crops between 600 and 1000 kg can be produced in the same area (FedeCafe), and therefore the need for vast territories.

[31] See Mejia & Restrepo (2015) for a detailed theoretical model. They show that the fully inelastic demand for cocaine guarantees that Colombian coca cultivation increases when Peru and Bolivia seize more cocaine, because traffickers substitute away from these sources. In the sample of municipalities used in this study, the increase in 1% in Peru and Bolivia eradication activities increases the amount of coca planted by 3.79% in Colombia.

[32] Angrist & Kugler (2008) exploit a sharp change in the structure of the Andean drug industry: before 1994, most of the cocaine exported from Colombia was refined from coca leaf grown in Bolivia and Peru. Beginning in 1994, however, in response to increasingly effective air interdiction by American and local militaries, the so-called air bridge that ferried coca paste from growers to Colombian refiners was disrupted. In response, coca cultivation and paste production shifted to Colombia's countryside, leading to high levels of violence.

[33] For instance, raw measures of coca could be related directly to social capital measures

[34] In experimental economics, Castillo & Carter (2011) find that people that experience extensive destruction from Hurricane Mitch shared a significantly larger portion of the pie with partners in a dictator game.

5.3.1 Instrument Relevance: First Stage

Table 2 explores whether external shocks to Colombian coca crops cause violence in high suitability areas. It reports the estimates for the excluded instrument in the first stage regressions. Each panel shows a different outcome since the sample changes slightly for some variables. The fact that the results are robust to the inclusion of controls indicates the high quality of the instrument for three reasons: first, it suggests that the first stage is not explained by other variables different to those in the instrument, which would create concerns about the relevance of the instrument. Second, the inclusion of controls could pick up a small amount of the endogenous variation in conflict (as R^2 goes to 1), making the exclusion restriction invalid. Third, the instrument is not correlated with the controls, which would invalidate the identification strategy.

In all cases, the coefficients have the expected sign and are significant. The coefficient in the second column, after including controls, indicates that when the eradication efforts in the neighboring countries go up by 10%, there is an increase of 10 violent events per 100,000 inhabitants for a municipality with the highest coca index.[35] Robust (Montiel-Plueger) F-statistics accounting for clustered residuals at municipality and year level are above the conventional threshold for weak instruments.[36]

5.3.2 Instrument Validity

Although the exclusion restriction is not directly testable, I discuss its plausibility. This condition is violated if there are unobservable time varying factors correlated with cross-border eradication efforts, for municipalities more suitable for coca production. In other words, if coca producer municipalities are less inclined to trust or participate due to unobserved cultural factors or history of violence, these factors might have a direct effect on social capital. However, as long as their influence did not change with the external shocks in coca markets (other than due to the increase in violence), the instrument would be uncorrelated with the omitted variables conditional on municipality fixed effects. In contrast, problems would arise if the error term includes time varying factors that are correlated with the ecological variables. An example might be that whenever the demand for coca increases in high suitability areas, people start working together in coca fields raising the levels of trust. Therefore, for robustness purposes, municipal controls such as unemployment, amount of taxes collected and school attendance are included. These variables capture the fact that changes in incentives to grow coca might affect the local economy by displacing legal employment or school

[35] The results do not vary among variables. The fact that the coefficient for fear is more than twice that of the rest of the variables is explained because the variable fear is not available for the year 2010 and 2011, which are the least violent years during the period studied.

[36] The standard Stock-Yogo critical values for weak instruments are only valid under *i.i.d* assumptions on the residuals Kleibergen & Paap (2006). Montiel-Olea & Pflueger (2013) define weak instruments when the worst-case bias of two-stage least squares exceed 10% of the worst case bias of OLS. For a critical value of 5% the null of weak instruments ranges from 9 to 11.52. So it is always close to the Stock-Yogo rule of thumb cutoff of 10

dropout in favor of illegal activities, which might affect social capital directly. Section 6 shows that the results are not sensitive to the inclusion of these controls.

In addition, I perform a placebo test to provide evidence for the exclusion restriction. Figure 7 plots the coca index (horizontal index), and the average of trust filtered by a set of municipality and time effects (vertical axis). The left panel shows municipalities characterized by a positive number of violent episodes, while the right panel displays municipalities with no violent episodes. The relationship is negative and highly significant across municipalities experiencing violence, but it is insignificant across those that do not experience violence. Though not a formal test for the exclusion restriction, this falsification analysis suggests that suitability for coca production has an effect on trust only through the channel of violence.

Moreover, suitability for coca production may be correlated with pre-conflict levels of social capital or with other characteristics affecting social capital, even after including municipality fixed effects. However, I cannot use a measure of social capital before the conflict started, as it began over 50 years ago, and the LAPOP survey started to be collected in 2004. Accordingly, I assess the relationship between the suitability index and different measures of historical violence; i.e, before drug violence started in the country. Therefore, if it is true that the suitability index affects violence only through incentives to fight over coca, we should not see any correlation between the coca index and violence in the first half of the 20^{th} century, giving an indication that the index is not related to predetermined institutional characteristics. Table A3 is consistent with this idea and shows that the coca suitability index is not correlated with previous measures of conflict. In particular, columns 1 to 3 report the probability of having a historical conflict given the suitability for growing coca.[37] The coefficients are not statistically significant for any of the previous conflicts in Colombia.

5.3.3 Monotonicity

In this setting, the monotonicity assumption requires that municipalities that have a low suitability for growing coca and are violent, would also be violent if they had a high suitability for growing coca, and vice versa for non-violent municipalities. This assumption ensures that the 2SLS identifies the local average treatment effect (LATE), i.e. the average causal effect among the subgroup of municipalities who could have had a different rate of violence if the shock to the coca market would have been different because of their conditions for growing coca.[38]

[37]Historical conflicts are defined by following Férnandez (2012), who uses National Archives to identify municipalities with different conflicts. During the period 1901-1917 and 1918-1913, there were disputes for land property rights. Whereas the period 1948-1953 was characterized by political conflicts between Liberals and Conservatives.

[38]Following Imbens & Angrist (1994), this assumption is also known as the "no defiers" assumption, and it assumes that the instrument affects the treatment in the same direction for the entire sample.

One testable implication of this assumption is that the first stage estimates should be non-negative for any subsample. Table A4 shows that for all quartiles of the coca index the first stage estimates are positive and statistically different from zero, consistent with the monotonicity assumption.

5.3.4 Other concerns

Finally, there may be a concern for reverse causality in the first stage, for instance because eradication policies in Peru and Bolivia might respond to violence in Colombia, as a way to avoid similar patterns of violence as its neighboring country. I address these concerns in different ways. First, I check whether in times of a generalized increase in violence in Colombia, there is still variation in violence at the municipal level since the identification comes from variation in violence at municipal level. This is illustrated in Figure 8, where the dots represent municipalities with different levels of violence across time. The idea behind this Figure is that even if cross-border shocks capture some variation caused by aggregate changes in the Colombian violence rates, this does not necessarily imply a problem for my identification strategy. As long as these aggregate changes are exogenous to a given municipality, I can use them in the spirit of a Bartick or shift-share instrument.

6 Results

In this section, I first present the main results, and provide evidence that these findings are robust to concerns related to the instrument, spillovers, and migration. Next, I study heterogeneous effects and secondary outcomes to provide hints of potential mechanisms. Finally, I show that the effect of conflict on social capital also matters for institutions such as participation in elections.

6.1 Main Results

In Table 3, I study the effects of conflict on the different measures of social capital. The dependent variable is trust in other members of the community in Panel A, participation in community organizations in Panel B, and cooperation in Panel C. OLS estimations of equation (1) are reported in column 1,[39] while column 5 presents the 2SLS results. 2SLS coefficients are larger than OLS. One possible explanation for this pattern is that OLS is downward biased because armed groups tend to attack communities that have a strong local organization. As a result the decrease in social capital was not substantial, because those communities had higher social capital levels to start with. Another possibility is measurement error in violence leading to attenuation bias.

[39]The OLS coefficients are similar in magnitude to the corresponding marginal effects of a Probit model. Using OLS allows me to cluster standard error at the municipality level and they also provide an easier interpretation. In addition, OLS allows for the inclusion of fixed effects.

Table 3 also presents the reduced form estimates in column 2 without controls, which estimate the direct effect of the instrument on social capital. These coefficients show that the increase in eradication efforts in neighboring countries decreases social capital for people living in municipalities that are more suitable for coca production. Columns 3 and 4 add controls; estimates are roughly unchanged, indicating that the effect is not driven by individual and municipality characteristics correlated with violence.

Subsequent columns of Table 3 present the 2SLS results with and without adding individual and municipality controls. Remarkably, the coefficient of interest is stable across specifications even after additional covariates are included, suggesting a small amount of selection on observables.[40] To discuss the magnitude of the results, I will focus on $Trust$ since it is the variable that has been broadly used in the literature of social capital. Trust has a sample mean of 0.68 and a standard deviation of 0.56. As shown in Table 3 Panel A, the point estimate in column 7 implies that a one standard deviation increase in the rate of violence produces a 38% standard deviation reduction in trust (corresponding to 17 percentage points). This magnitude corresponds to the difference in trust levels between Germany and Colombia, according to the World Value Survey (2009). Panel B shows that conflict decreases participation in community organizations by 22% of its standard deviation (10 percentage points). Panel C has the same structure as the previous one, but shows the contribution to problem solving decreases, which decreases by 23% of a standard deviation (11 percentage points).

However, even if the effects at the mean are negative, I investigate the effects of violence along the violence distribution. For this purpose, I estimate the effect at different percentiles of violence for the entire period of study. Intriguingly, the effect is positive for low levels of violence (column 1 and 2). To give a sense of the levels of violence in municipalities in Colombia, I use other countries for comparison in Table 4. For instance, municipalities in the 8^{th} percentile of violence have the same violence rate as the US (5.33 homicides per 100,000). However, as violence becomes intense, and we move in the distribution of violence similar to countries like Mexico and Brazil, the effect becomes negative and significant (column 3 and 4). This is the case for most of the Colombian municipalities, for which the average rate of violence is 49, and situates the country in the 90^{th} percentile of the most violent countries in the world. As a result, it is not surprising that the overall estimated effects are negative. Consistent with this, Figure 9 presents negative marginal effects of conflict.

This pattern is in line with the theoretical framework presented by Jennings & Sanchez-Pages (2017). Their model predicts that conflict leads to higher investments in social capital, because of its protective effect in confronting external threats and increasing the probability of survival. However, as the threat becomes severe, social networks are disrupted and social

[40] Section 6.4 presents a discussion on the omitted variables by using the Altonji (2005)'s and Oster (2016)'s approaches.

capital is likely to fall. Overall, this provides empirical evidence that helps to reconcile the apparently contradictory results in the literature.

6.2 Mechanisms

This subsection discusses the possible mechanisms that drive the main results documented in this paper.

a. Fear to be involved with the community

In the case of Colombia and other conflicts[41], where the various warring factions are not readily identifiable, people can be especially wary and guarded about whom they can trust. As a result, conflict could have created fear of involvement with the community. To test this idea, Table 5, investigates the impact of conflict on different variables that measure fear of involvement with the community. These variables include: fear of participating in community organizations, fear of running as candidates in local polls, and fear of voting in elections.[42] The structure of the table mirrors that of Table 3: column 1 reports results from the baseline specification for OLS, while columns 2 to 7 show the reduced form and 2SLS estimations. 2SLS estimates are positive and statistically significant, suggesting the fear of dealing with neighbors that might belong to an armed group reduces the levels of social capital.

b. Inability to distinguish "the enemy"

Another possibility that explains the disruption in social capital is related to the nature of the conflict. For the Colombian scenario, as for other non-related ethnic conflicts, people are unable to apply basic cues to identify friend from foe within their communities. Accordingly, the effect should be most stringent in municipalities where the population was more heterogeneous. Table 6, breaks down the effect of conflict on social capital along two different dimensions to test this mechanism. In columns 1-2, I rely on the conventional 90^{th} to 10^{th} ratio of income distribution.[43] I find that the association between conflict and trust is stronger in more unequal municipalities (above the median of the 90-10 inequality ratio). Consistent with this story, Column 3-4 follow the definition of ethnic fractionalization used commonly in the literature[44], and it shows that the effect of conflict on social capital is higher in more racially-mixed communities.

[41] For instance, the Spanish Civil War, the Tajik War, and the Turkish and Kurdish confrontations, are conflicts in which the various warring factions are not readily identifiable.

[42] The precise questions used were: "Do you feel fear to participate in solving a problem in your community?, Do you feel fear to vote in elections of your community?, Do you feel fear to participate in a peaceful protest?, Do you feel fear to run for a political seat in your community?". The variables take the value of one if the respondent answers yes. Otherwise, the value is zero (`Questions` $der1 - der4$).

[43] This measure takes the 90^{th} percentile of the income distribution and divides by the 10^{th} percentile at municipality-year level. The mean is 3.

[44] This measures indicates the probability that two random people in a municipality belong to different groups. The index is given by $Frac = 1 - \sum_i s_i^2$. If there is only one group, then the summation term is 1, so Frac is zero Easterly & Levine (1997).

Taken together, these findings point towards conflict decreasing social capital in more heterogeneous municipalities. This interpretation is consistent with parochial theories that highlight the importance of group boundaries. However, one needs to make the caveat that these distinctions are only informative of the different groups in the municipalities, as it is difficult to correctly get information on their relationship with armed groups based on income and race, as conflict in Colombia was not ethnically driven. Nevertheless, these pieces of evidence help understand the mixed results in the literature. First, the strength of the effect seems to depend on the intensity of conflict. Second, how homogeneous a society is can also play a role in determining the effects of conflict on social capital.

6.3 Heterogeneity Analysis

Having established a significant impact of conflict on social capital, it is natural to ask whether some population groups are affected more than others. I present a heterogeneity analysis with respect to population characteristics (gender, race, having a close victim of conflict, and living in rural areas). Table 7 re-estimates separately the baseline specification by allowing an interaction between specific subsamples. The effects are still negative. However, they are not statistically significantly different, indicating that there not certain groups in the population for which conflict differentially affected their social capital.

6.4 Robustness

In order to assess the robustness of the findings, I perform a series of checks for possible confounding effects:

a. Omitted variables

The main results presented in Table 3 show that the coefficient of interest is stable across specifications even after additional covariates are included, suggesting a small amount of selection on observables. However, it is still possible that a small amount of selection on unobservables could explain the whole effect. I explore this possibility by following the Altonji (2005) omitted variable approach. Roughly speaking, the smaller the difference between the coefficients with and without controls, the less the estimate is affected by selection on observables, and so the larger the selection on unobservables needs to be, in order to explain away the entire effect of the variables of interest. This approach uses the degree of selection on observables as a guide to the degree of selection on unobservables.[45] The value of the ratio indicates that selection on unobservables would need to be 10.5 times stronger than the selection on observables, which seems highly unlikely.[46] I also perform the Oster (2016)

[45]The Altonji ratio is calculated as $\beta_f/(\beta_r - \beta_f)$, where β_r corresponds to coefficient without controls, and β_f is obtained when the full set of observable characteristics are controlled for. In my case, $\beta_f = -0.0031$ and $\beta_r = -0.0034$.

[46]More formally, the shift in the distribution of the unobservables would have to be 10.5 times as large as

test to evaluate robustness to omitted variable bias. The basic idea is that the stability of the coefficient is not enough of a sign that the omitted bias is limited. The reason is that coefficients can appear stable after the addition of controls, but this is simply because less of the outcome variable is explained by the observed variables. The δ obtained is equal to 5.7 which indicates that the unobservables would need to be more than 5 times as important as the observables to produce an effect of zero.

b. Spillover effects

So far I have discussed the impact of a municipality m being attacked by an armed group on its own population. In this subsection, I study to what degree local social capital is affected by conflict in neighboring municipalities. I perform two exercises: First, instead of using the own violence to estimate the effect on social capital, I use the average violence of neighboring municipalities.[47] Table A7 (column 1) shows that fighting in neighboring municipalities does not change the results of interest. If anything, the effect is slightly higher, but not statistically different from the coefficient in the main specification.[48] Second, I allow for spatial correlation across municipalities when estimating the standard errors. Table A7 (column 2) re-estimates the main specification allowing for spatial correlation. The standard errors are larger, but the results are still statistically significant.

c. Migration

The literature in Economics suggests that forced displacement can affect individuals' behavior, and Colombia had one of the highest number of internally displaced people worldwide by the time of the study (Moya & Carter, 2014). This could bias the results if people who trust less in others are also more likely to migrate, which in turn underestimates the effects of conflict. On the other hand, if the displaced population is discriminated against upon its arrival in municipalities, this would overestimate the effects of conflict. However, with the available data, it is not possible to identify whether an individual is a direct victim of displacement. The best approximation is controlling for the net flow of forced displaced population. However, I also restrict the sample to people who report having lived in the same municipality for the last five years. Clearly, it is an imperfect measure of forced displacement, since we do not know the motives for migration. Table A8 splits the sample between those who have migrated and those who have not. When looking at individuals that did not migrate, the effect remains the same as compared to the initial specification, whereas for those who migrated the estimates are not statistically significant. This evidence indicates that the main effects are not driven by forced displacement.

the shift in the observables to explain away the entire conflict effect.

[47]The neighbors are identified as those municipalities that share a common border.

[48]One explanation for which this effect is higher could be that the average of neighboring violence is higher than for the municipality. Therefore, people might feel more threatened.

d. Instrument

As the instrument is given by the interaction between plausibly exogenous variation in the suitability for growing coca and shocks to the Colombian coca production, the first stage of the IV can be thought of as a form of a difference-in-difference estimator. Thus, if the standard parallel trends assumption in DID is violated, the exclusion restriction underpinning the IV approach fails. Barrett & Paul (2017) point out the possibility that the results are driven by spurious correlations between the instrument, the dependent variable, and the outcome variable.

Hence, I perform an additional placebo test in the form of randomization inference. This test rests on the principle that introducing randomness into the endogenous explanatory variable of interest (a municipality's rate of violence), while holding constant the instrument, should eliminate, or at least substantially attenuate, the estimated causal relationship if indeed exogenous inter-annual shocks to the endogenous explanatory variable (violence) drive the main outcome (social capital). Therefore, I randomly assign (without replacement) the rate of violence among the 55 municipalities for every year in the sample. This "new dataset" preserves the two sources of endogeneity that Barrett & Paul (2017) worry about -the time trend and selection into the treatment- but sweeps out the source of variation by randomizing the violence among municipalities. For instance, violence in town A, cannot plausibly have caused violence in town B. This way, social capital can remain spuriously related to external shocks because neither the social capital nor the instrument are altered, but the causal mechanism has been rendered non operational by randomization.

If it is true that the causal relationship between violence and social capital is negative and the identification is unaffected by selection bias and spurious time trends, the distribution of coefficients would shift to the right relative to the original estimation, and if the share of municipalities in which violence causes changes in social capital is small relative to a large enough sample, would center around zero, because the randomization of violence would attenuate the estimated relationship between violence and social capital. From the randomization I obtain a mean of 0.007, and a median of 0.011. Both do not lie on the confidence intervals of the observed effect (lower bound=-0.007, upper bound=0.0001, observed effect=-0.0031) (See Figure A1).

Finally, it is reassuring that the LIML (Limiting Information Maximum Likelihood) estimators are almost identical to the 2SLS (See Table A5), indicating that there is no bias due to weak instruments. The LIML is an estimator that is less efficient, but also less biased by weak instruments (Angrist & Pischke, 2008). Furthermore, following Rohner *et al.* (2013a) I report the results of a specification where all variables are collapsed to the municipality level, allowing for the computation of the standard Cragg-Donald Wald F test. The F statistic is 16.38, which again suggests that weak instruments are not a problem (see

Table A6).

6.5 Extension: Effect on Voter Turnout

In an attempt to assess the economic impact of the collapse of social capital, I retrieve the estimates of the elasticity of institutional outcomes to trust. The hypothesis is that if conflict destroys trust, it might also affect other outcomes such as voter turnout. Figure 10 shows that there is negative correlation between the turnout in the peace referendum of 2016 and the rate of violence during the period of study. When this relationship is analyzed in the IV context at the municipal level, I find an elasticity of -5.26. This suggests that violence has a persistent impact on political institutions, as by 2016, most of the violent attacks had ceased.[49] Violence affects not only informal institutions but also formal institutions such as elections, which are proven to foster economic development.

7 Conclusions

In this paper, I estimate the causal effects of violence on social capital in an ethnically homogeneous context. This question has not been addressed before in the literature. To do that, I exploit the suitability of municipalities to produce coca, interacted with cross-border shocks to the Colombian coca market. A picture emerging from these findings is clear: violence perpetrated by armed groups destroys trust, participation in community organization, and contribution to community problem solving. These results seem to be driven by fear of getting involved with neighbors who might be belong to armed groups.

The results raise three important points. First, they speak to the literature on development that explores the consequences of violence by pointing to a negative effect of violence on social capital, and highlighting the fact that regardless of the particularity of the conflict, the observability of the enemy and its intensity play a role in shaping pro-social preferences. This may potentially help us understand the apparently contradictory results presented in the literature. For instance, in Sierra Leone, where the conflict has ethnic roots, one ethnicity was pitted against another generating positive developments in collective action. Whereas in Uganda, violence decreased trust towards members from other ethnicities. I speculate that the nature of the Colombian conflict, characterized by the inability to apply basic cues to identify friends or foe, has led to a disruption in social capital. However, for low-levels of violence social capital increased possibly as a way to face external threats. Second, violence does not only affect social preferences such as trust and participation in community organizations. Rather, this disruption is translated into the erosion of democratic institutions

[49]This result could be an indication that the effect on participation in community organizations was not simply a mechanical effect by which people did not attended to community meetings because the war. The reasoning is that in 2016 there was a peace deal that reduced substantially the intensity of violence, and if the mechanical argument would have driven the results then there should not be a significant effect on the referendum participation.

that have been proven relevant for economic development. Finally, the implications from a policy perspective rest on taking into account the state of social networks in post-conflict settings. This raises the challenge of re-stabilizing not only physical and human capital, but also social capital.

References

2012. World Drug Report. *United Nations Office on Drugs and Crime.*

Acemoglu, Daron, Robinson, James A., & Santos, Rafael J. 2013. The Monopoly of Violence: Evidence from Colombia. *Journal of the European Economic Association*, **11**(S1), 5–44.

Acemoglu, Daron, García-Jimeno, Camilo, & Robinson, James A. 2015. State Capacity and Economic Development: A Network Approach. *American Economic Review*, **105**(8), 2364–2409.

Acock, Mary, Lydon, John, Johnson, Emanuel, & Collins, Ronald. 1996. Effects of Temperature and Light Levels on Leaf Yield and Cocaine Content in Two Erythroxylum Species. *Annals of Botany*, **78**(1), 49 – 53.

Aghion, Philippe, Algan, Yann, Cahuc, Pierre, & Shleifer, Andrei. 2010. Regulation and Distrust. *The Quarterly Journal of Economics*, **125**(3), 1015–1049.

Alesina, Alberto, & Ferrara, Eliana La. 2002. Who Trusts Others? *Journal of Public Economics*, **85**, 207–34.

Algan, Yann, & Cahuc, Pierre. 2010. Inherited Trust and Growth. *American Economic Review*, **100**(5), 2060–92.

Angrist, J.D., & Pischke, J.S. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Angrist, Joshua D., & Kugler, Adriana D. 2008. Rural Windfall or a New Resource Curse? Coca, Income, and Civil Conflict in Colombia. *The Review of Economics and Statistics*, **90**(2), 191–215.

Bagley, Bruce. 2012. *Drug Trafficking and Organized Crime in the Americas.*

Bandiera, Oriana, & Rasul, Imran. 2006. Social Networks and Technology Adoption in Northern Mozambique. *The Economic Journal*, **116**(514), 869–902.

Barrett, Christopher, & Paul, Christian. 2017. Revisiting the effect of food aid on conflict: a methodological caution. *Policy Research Working Paper; no. WPS 8171. Washington, D.C. : World Bank Group.*

Bauer, Michal, Blattman, Christopher, Chytilova, Julie, Henrich, Joseph, Miguel, Edward, & Mitts, Tamar. 2016. Can War Foster Cooperation? *Journal of Economic Perspectives*, **30**(3), 249–74.

Bellows, John, & Miguel, Edward. 2009. War and local collective action in Sierra Leone. *Journal of Public Economics*, **93**(11-12), 1144–1157.

BenYishay, Ariel, & Mobarak, A. Mushfiq. 2014 (May). *Social Learning and Communication.* Working Paper 20139. National Bureau of Economic Research.

Besley, Timothy, & Reynal-Querol, Marta. 2012 (Feb.). *The Legacy of Historical Conflict Evidence from Africa.* STICERD - Economic Organisation and Public Policy Discussion Papers Series 036. Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Cadena, Ximena, & Zuluaga, Victor. 2012. *Vulnerabilidad a choques y desastres naturales.Encuesta Longitudinal Colombiana.* Boletin de Divulgacion ELCA No.2 1026. Universidad de los Andes.

Caselli, Francesco, & Coleman, Wilbur John. 2013. On The Theory of Ethnic Conflict. *Journal of the European Economic Association*, **11**, 161–192.

Cassar, Alessandra, Grosjean, Pauline, & Whitt, Sam. 2013. Legacies of Violence: Trust and Market Development. *Journal of Economic Growth*, **18**(3), 285–318.

Castillo, Marco, & Carter, Michael. 2011 (June). *Behavioral Responses to Natural Disasters.* Working Papers 1026. George Mason University, Interdisciplinary Center for Economic Science.

Chandra, K. 2007. *Why Ethnic Parties Succeed: Patronage and Ethnic Head Counts in India.* Cambridge Studies in Comparati. Cambridge University Press.

Choi, Jung-Kyoo, & Bowles, Samuel. 2007. The Coevolution of Parochial Altruism and War. *Science*, **318**(5850), 636–640.

Coleman, J.S. 1994. *Foundations of Social Theory.* Belknap Series. Belknap Press of Harvard University Press.

Colleta, N.J., & Cullen, M. 2000. The Nexus between Violent Conflict, Social Capital and Social Cohesion: Case Studies from Cambodia and Rwanda. *World Bank.*

Collier, Paul, Hoeffler, Anke, & Rohner, Dominic. 2009. Beyond Greed and Grievance: Feasibility and Civil War. *Oxford Economic Papers*, **61**(1), 1.

Conley, Timothy G. 2016. *Spatial Econometrics.* London: Palgrave Macmillan UK. Pages 1–9.

Conley, Timothy G., & Udry, Christopher R. 2010. Learning about a New Technology: Pineapple in Ghana. *American Economic Review*, **100**(1), 35–69.

de Memoria Historica, Centro. 2013. *Basta Ya. Informe General.* Departamento Nacional de Paneacion.

Dube, Oeindrila, & Vargas, Juan. 2013. Commodity Price Shocks and Civil Conflict: Evidence from Colombia. *Review of Economic Studies*, **80**(4), 1384–1421.

Durlauf, Steven N., & Fafchamps, Marcel. 2005. Social Capital. *Chap. 26, pages 1639–1699 of:* Aghion, Philippe, & Durlauf, Steven (eds), *Handbook of Economic Growth.* Handbook of Economic Growth, vol. 1. Elsevier.

Easterly, William, & Levine, Ross. 1997. Africa's Growth Tragedy: Policies and Ethnic Divisions. *The Quarterly Journal of Economics*, **112**(4), 1203–1250.

Ehlers, Anke, & Clark, David M. 2000. A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, **38**(4), 319 – 345.

Esteban, Joan, & Ray, Debraj. 2011. A Model of Ethnic Conflict. *Journal of the European Economic Association*, **9**(3), 496–521.

Fafchamps, Marcel, & Lund, Susan. 2003. Risk-Sharing Networks in Rural Philippines. *Journal of Development Economics*, **71**(2), 261 – 287.

Foster, Andrew D., & Rosenzweig, Mark R. 2001. Imperfect Commitment, Altruism, And The Family: Evidence From Transfer Behavior In Low-Income Rural Areas. *The Review of Economics and Statistics*, **83**(3), 389–407.

Fumerton, Mario. 2001. Rondas Campesinas in the Peruvian Civil War: Peasant Self-defence Organisations in Ayacucho. *Bulletin of Latin American Research*, **20**(4), 470–497.

Férnandez, Manuel. 2012. Violencia y derechos de propiedad: El caso de la violencia en Colombia. *Ensayos sobre Política Económica*, **30**(69).

Glennerster, Rachel, Miguel, Edward, & Rothenberg, Alexander D. 2013. Collective Action in Diverse Sierra Leone Communities. *The Economic Journal*, **123**(568), 285–316.

Guiso, Luigi, Sapienza, Paola, & Zingales, Luigi. 2004. The Role of Social Capital in Financial Development. *American Economic Review*, **94**(3), 526–556.

Guiso, Luigi, Sapienza, Paola, & Zingales, Luigi. 2006. Does Culture Affect Economic Outcomes? *Journal of Economic Perspectives*, **20**(2), 23–48.

Hancock, Landon, & Mitchell, Christopher. 2007. *Zones of peace*.

Homans, George C. 1958. Social Behavior as Exchange. *American Journal of Sociology*, **63**(6), 597–606.

Imbens, Guido W., & Angrist, Joshua D. 1994. Identification and Estimation of Local Average Treatment Effects. *Econometrica*, **62**(2), 467–475.

Jennings, Colin, & Sanchez-Pages, Santiago. 2017. Social capital, conflict and welfare. *Journal of Development Economics*, **124**, 157 – 167.

Johnson, Emanuel L., Campbell, T. Austin, & Foy, Charles D. 1997. Effect of Soil pH on Mineral Element Concentrations of Two Erythroxylum Species. *Journal of Plant Nutrition*, **20**(11), 1503–1515.

Kalyvas, S.N. 2006. *The Logic of Violence in Civil War*. Cambridge Studies in Comparative Politics. Cambridge University Press.

Kaplan, Oliver. 2013. Protecting civilians in civil war: The institution of the ATCC in Colombia. *Journal of Peace Research*, **50**(3), 351–367.

Kaplan, O.R., Laitin, D.D., Karl, T.L., Schultz, K., Weinstein, J.M., & of Political Science, Stanford University. Department. 2010. *Civilian Autonomy in Civil War*.

Kleibergen, Frank, & Paap, Richard. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, **133**(1), 97–126.

Martínez, Luis R. 2017. Transnational insurgents: Evidence from Colombia's FARC at the border with Chávez's Venezuela. *Journal of Development Economics*, **126**, 138 – 153.

Mejia, Daniel, & Posada, Carlos Esteban. 2008. Cocaine Production and Trafficking: What Do We Know? *Policy Research Working Paper. Development Research Group. World Bank*.

Mejia, Daniel, & Restrepo, Pascual. 2015. Bushes and Bullets: Illegal Cocaine Markets and Violence in Colombia. *Working Paper MIT*.

Molina, M.A., & Castaño, C. 2001. *Mi confesión: Carlos Castaño revela sus secretos.* Editorial Oveja Negra.

Montiel-Olea, José-Luis, & Pflueger, Carolin. 2013. A Robust Test for Weak Instruments. *Journal of Business & Economic Statistics*, **31**(3), 358–369.

Morales, Juan S. 2018. The impact of internal displacement on destination communities: Evidence from the Colombian conflict. *Journal of Development Economics*, **131**, 132 – 150.

Moya, Andres, & Carter, Michael. 2014 (Sept.). *Violence and the Formation of Hopelessness and Pessimistic Prospects of Upward Mobility in Colombia.* NBER Working Papers 20463. National Bureau of Economic Research, Inc.

Moya, Andrés. 2018. Violence, psychological trauma, and risk attitudes: Evidence from victims of violence in Colombia. *Journal of Development Economics*, **131**, 15 – 27.

Nannicini, Tommaso, Stella, Andrea, Tabellini, Guido, & Troiano, Ugo. 2013. Social Capital and Political Accountability. *American Economic Journal: Economic Policy*, **5**(2), 222–50.

Nunn, Nathan, & Wantchekon, Leonard. 2011. The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review*, **101**(7), 3221–52.

Olken, Benjamin A. 2009. Do Television and Radio Destroy Social Capital? Evidence from Indonesian Villages. *American Economic Journal: Applied Economics*, **1**(4), 1–33.

Otis, John. 2014. *The FARC and Colombias Illegal Drug Trade.*

Plowman, Timothy. 1979. Botanical Perspectives on Coca. *Journal of Psychedelic Drugs*, **11**(1-2), 103–117. PMID: 522163.

Putnam, R.D. 2020. *Bowling Alone: Revised and Updated: The Collapse and Revival of American Community.* Simon & Schuster.

Putnam, Robert D., Leonardi, Robert, & Nanetti, Raffaella. 1993. *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton: Princeton University Press.

Rohner, Dominic, Thoenig, Mathias, & Zilibotti, Fabrizio. 2013a. Seeds of distrust: conflict in Uganda. *Journal of Economic Growth*, **18**(3), 217–252.

Rohner, Dominic, Thoenig, Mathias, & Zilibotti, Fabrizio. 2013b. War Signals: A Theory of Trade, Trust, and Conflict. *The Review of Economic Studies*, **80**(3), 1114.

Saffer, Henry, & Chaloupka, Frank. 1999. THE DEMAND for ILLICIT DRUGS. *Economic Inquiry*, **37**(3), 401–411.

Starn, O., Mignolo, W.D., Silverblatt, I., & Saldívar-Hull, S. 1999. *Nightwatch: The Politics of Protest in the Andes.* Latin America Otherwise. Duke University Press.

Suarez, Alfredo Rangel. 2000. Parasites and Predators: Guerrillas and the Insurrection Economy of Colombia. *Journal of International Affairs*, **53**(2), 577–601.

UNODC. 2009. *World Drug Report.* Tech. rept. United Nations Office on Drugs and Crime.

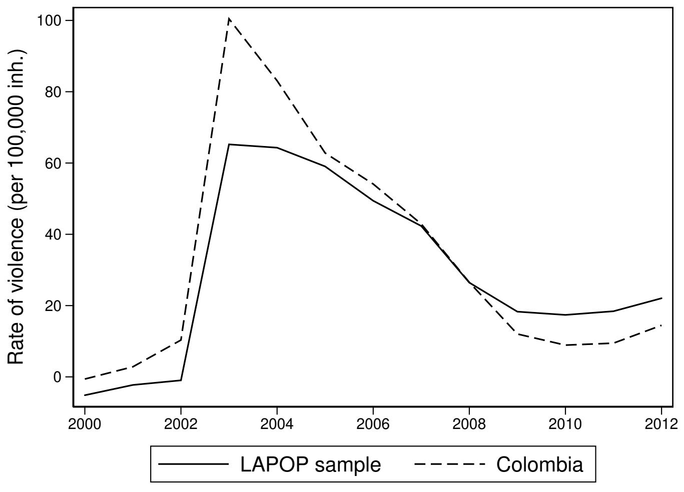
Vargas, Juan. 2009. Military Empowerment and Civilian Targeting in Civil War. *Documentos de Trabajo, Facultad de Economía, Universidad del Rosario*, 5–44.

Voors, Maarten J., Nillesen, Eleonora E. M., Verwimp, Philip, Bulte, Erwin H., Lensink, Robert, & Van Soest, Daan P. 2012. Violent Conflict and Behavior: A Field Experiment in Burundi. *American Economic Review*, **102**(2), 941–64.

Wood, E.J. 2003. *Insurgent Collective Action and Civil War in El Salvador*. Cambridge Studies in Comparative Politics. Cambridge University Press.

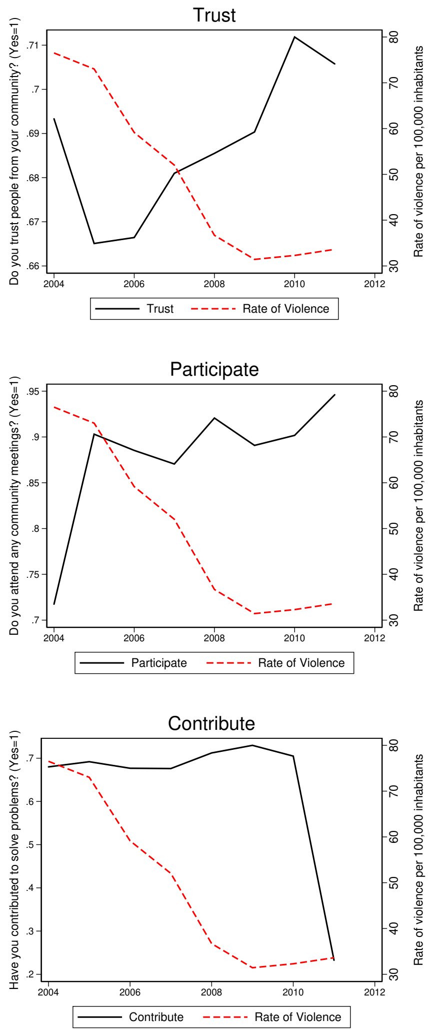
Figures

Figure 1: Comparison of the rate of violence between the sample of municipalities used in this study and the entire country



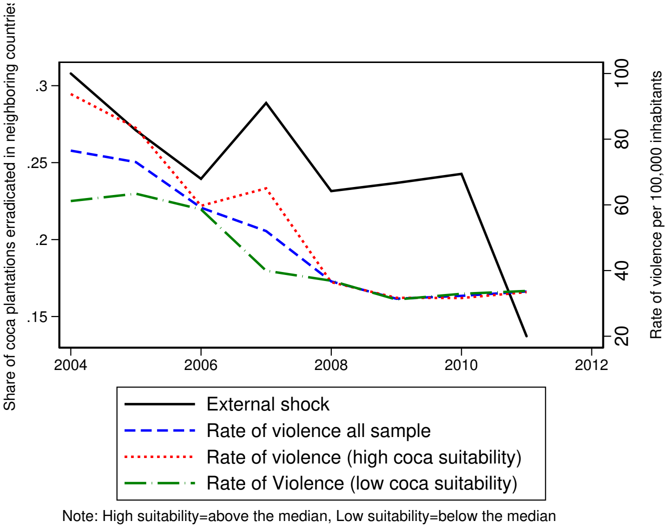
Note: The figure shows yearly averages of the rate of violence for the municipalities used in this paper (solid line), and for the entire country (dashed line). The sample contains 55 out of 1,121 municipalities in Colombia. They were selected by the LAPOP survey (Latin American Public Opinion Project) based on being nationally representative for both rural and urban areas in terms of socio-economic characteristics and population size. The rate of violence corresponds to the sum of 19 violent indicators divided by 2005 population from the Census (which is the latest year available for the period studied). Data for violence come from CEDE.

Figure 2: Rate of violence vs. average of social capital measures by year



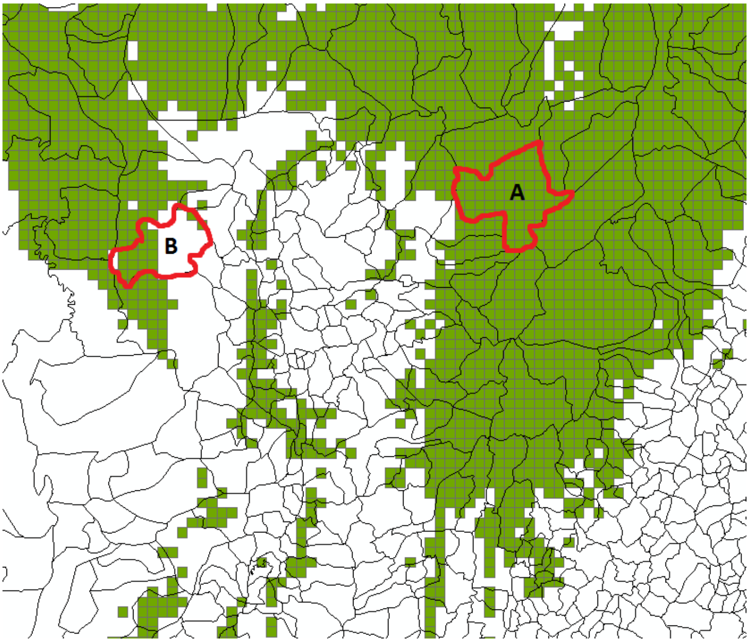
Note: The graphs show yearly averages of the rate of violence and the four measures of social capital employed in this paper. The social capital measures are in solid lines and they correspond to the yearly average of individual questions from the LAPOP survey. The rate of violence (dashed line) is the sum of 19 violent indicators divided by population.

Figure 3: Eradication in Peru and Bolivia is positively related to violence in Colombia



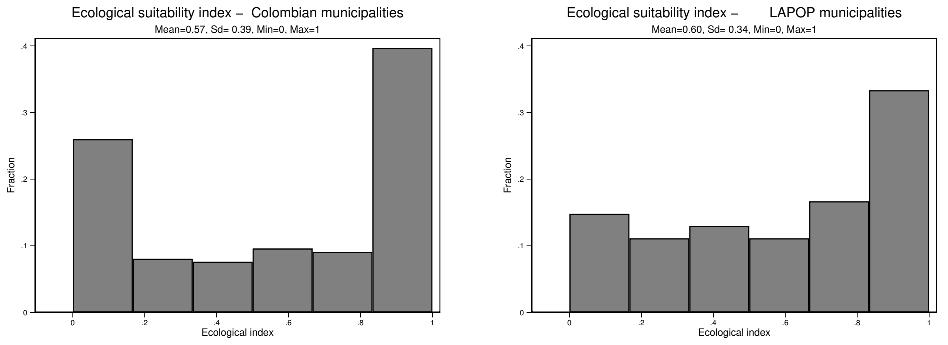
Note: The graph presents the yearly average rate of violence for municipalities with different suitability for coca production. The solid line depicts the amount of hectares that eradicated in Peru and Bolivia across time.

Figure 4: Example of how municipalities are divided into grids to calculate the share of land that is suitable for coca production



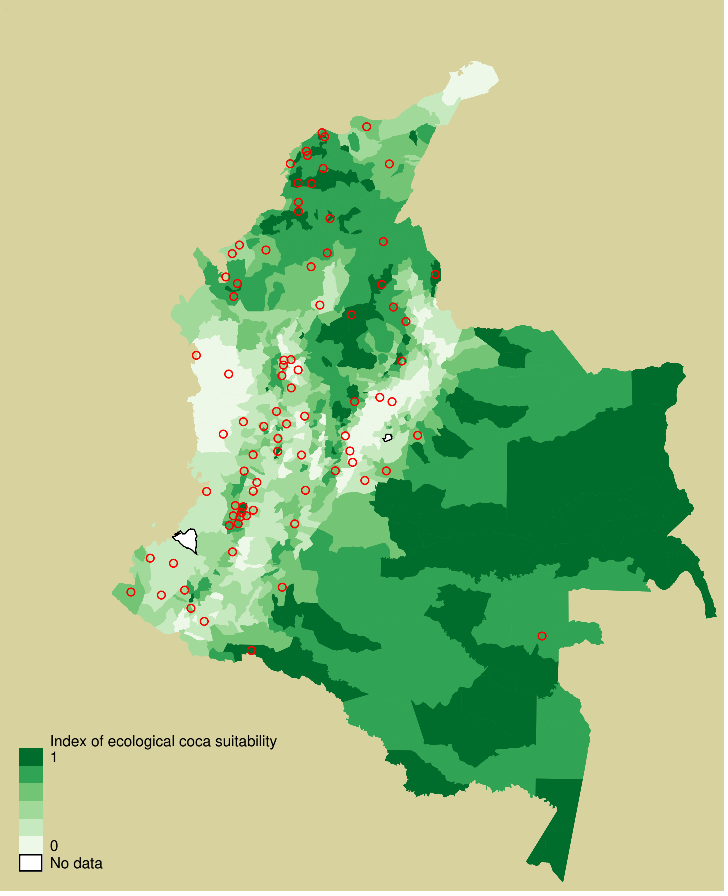
Note: The figure gives an example of how municipalities are divided into grids to calculate the share of land that is suitable fo coca production. For instance, Municipality "A" is completely suitable for coca production, while only 0.5 of the municipality B is suitable because only half of the grids in the municipality fulfill the ecological requirements for coca cultivation.

Figure 5: Distribution of the coca index



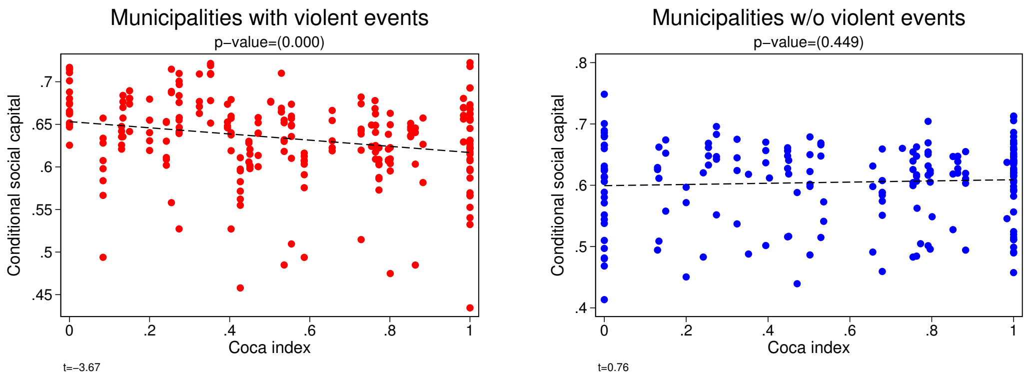
Note: The histograms show the distribution of the coca suitability index. The left graph is for all the municipalities in Colombia. The graph in the right corresponds for the municipalities that are included in the sample used in this paper.

Figure 6: Map of coca suitability for Colombia. Darker municipalities are more suitable for coca production



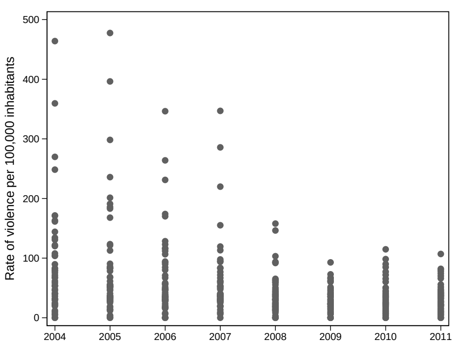
Note: The map presents the geographic distribution of the coca suitability index. The index goes from 1 to 0. Red dots are the 55 municipalities in the studied sample. The figure suggests that the coca index is widely distributed throughout most of the country.

Figure 7: Placebo test for exclusion restriction of the instrument



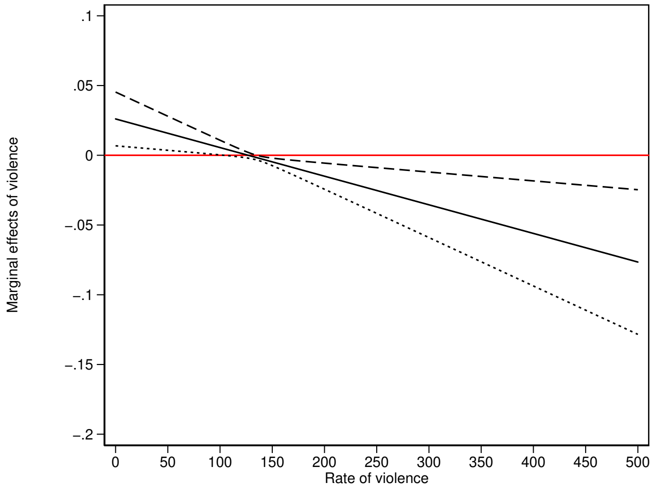
Note: The figures show the average trust filtered by a set of municipality and time effects. The right side presents the municipalities that have positive rate of violence. The left side displays municipalities with out violent events. There is only a negative relationship between the instrument-coca suitability index- and the conditional main measure of social capital (trust), suggesting that the only way in which the instrument affects social capital is through the levels of violence.

Figure 8: Variation of violence within municipalities across time



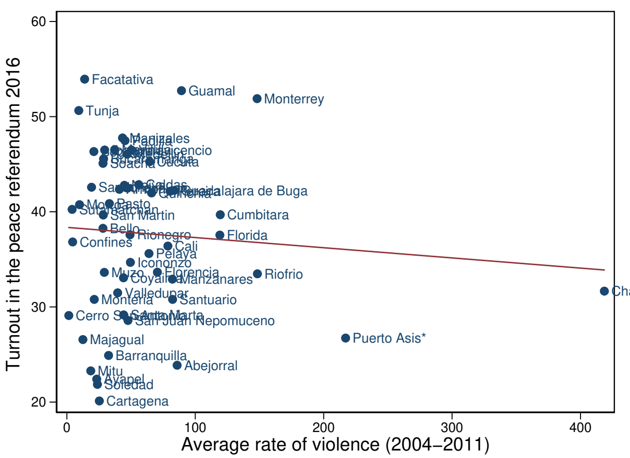
Note: The graph shows the yearly rate of violence for each municipality, which are represented by dots. The purpose of the graph is to show that there is municipality variation in the rate of violence across years as the identification strategy comes from variation in violence at municipal level.

Figure 9: Non-linear effects of violence on trust



Note: The figure plots the marginal effects of conflict on trust. Confidence intervals at 95%.

Figure 10: Turnout in Referendum vs. Rate of Violence



Note: The graph plots the average rate of violence for the period 2004-2011 for all the municipalities included in the sample, against the turnout in the peace referendum in 2016.

Tables

Table 1: Descriptive Statistics

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
Panel A: Social Capital					
Trust (=1)	11,651	0.687	0.464	0	1
Participation in community organizations (=1)	11,855	0.600	0.490	0	1
Cooperation to community problem solving (=1)	11,835	0.637	0.481	0	1
Fear to participate in community organizations (=1)	8,578	0.365	0.482	0	1
Fear to run for local elections (=1)	6,411	0.531	0.499	0	1
Fear to vote in local elections (=1)	8,002	0.154	0.361	0	1
Mean of all variables (=1)	6,411	0.560	0.225	0	1
Panel B: Individual controls					
Women (=1)	11,871	0.502	0.500	0	1
Years of education	11,871	8.877	4.350	0	17
Age	11,871	36.96	14.77	18	99
Black (=1)	11,871	0.0580	0.231	0	1
Income category	11,871	3.940	1.802	0	10
Media consumption (tv, newspaper, radio=1)	11,871	0.910	0.286	0	1
Close victim of conflict (=1)	11,871	0.320	0.467	0	1
Panel C: Municipality variables					
Rate of violence (per 100,000 pop)	11,871	49.26	56.61	0	477.6
Rate of violence general (per 100,000 pop)	11,871	7.438	22.76	0	264
Rate of violence selective (per 100,000 pop)	11,871	41.82	39.56	0	295.9
Index for coca suitability	11,871	0.508	0.373	0	1.000
Area of coca cultivated	11,871	36.12	273.0	0	4,531
# pupils	11,871	278,908	470,909	425	1.365e+06
Tax collection index	11,871	67.08	10.12	0	94.19
# forced displaced pop	11,871	0.0126	0.112	0	1

Note: This table shows the summary statistics at individual and municipality level. Panel A presents the measures of social capital from the LAPOP survey. Panel B displays the socio-economic characteristics of the sample which are also obtained from the LAPOP survey. Panel C comes from CEDE data. All the variables are presented as the average across years for the period 2004-2011.

Table 2: First stage results

Dependent variable in the first stage: Rate of violence	(1) *First Stage* Rate of violence	(2) *First Stage + Controls* Rate of violence
Panel A: Trust		
Coca index*Ln(External Shock)	96.43***	100.69***
	(27.70)	(28.97)
F test Cragg-Donald	218.67	227.67
F test Kleibergen-Paap	12.11	12.08
F test Montiel-Plueger	12.14	12.10
R2	0.65	0.48
N	11,682	11,682
Panel B: Participation		
Coca index*Ln(External Shock)	95.76***	100.11***
	(27.49)	(28.71)
F test Cragg-Donald	222.12	231.67
F test Kleibergen-Paap	12.12	12.15
F test Montiel-Plueger	12.15	12.18
R2	0.65	0.54
N	11,889	11,889
Panel C: Contribute		
Coca index*Ln(External Shock)	95.17***	99.50***
	(27.17)	(28.37)
F test Cragg-Donald	219.99	229.39
F test Kleibergen-Paap	12.26	12.29
F test Montiel-Plueger	12.29	12.32
R2	0.65	0.54
N	11,869	11,869
Panel E: Fear		
Coca index*Ln(External Shock)	226.35***	229.14***
	(68.20)	(79.68)
F test Cragg-Donald	178.80	169.79
F test Kleibergen-Paap	11.01	8.26
F test Montiel-Plueger	11.04	8.29
R2	0.72	0.66
N	6,252	6,252
Panel F: Mean of all variables		
Coca index*Ln(External Shock)	95.82***	100.18***
	(27.47)	(28.69)
F test Cragg-Donald	222.73	232.32
F test Kleibergen-Paap	12.16	12.18
F test Montiel-Plueger	12.18	12.21
N	6,252	6,252

Note: The table shows the results from the first stage of an instrumental variable regression. Standard error in parenthesis (robust clustered at the municipality level). All regressions include municipality and year fixed effects. The dependent variable is the rate of violence. Each panel shows a different outcome since the sample changes sightly for some variables. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 3: Main results of the effect of conflict on social capital

Dependent variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	OLS	———Reduced form———			————-2SLS————-		
Panel A: Trust (mean=0.68)							
Coca index*Shock	-0.0002						
(0.0001)							
Rate of violence		-0.3322***	-0.3170***	-0.3172***	-0.0034***	-0.0033***	-0.0031***
		(0.0894)	(0.0886)	(0.0906)	(0.0010)	(0.0009)	(0.0009)
Observations	11,682	11,682	11,682	11,682	11,682	11,682	11,682
R-squared	0.0247	0.0257	0.0454	0.0467			
Panel B: Participation in community organizations (mean=0.60)							
Rate of violence	-0.0004***	-0.1829**	-0.1981**	-0.1896**	-0.0019**	-0.0021**	-0.0019**
	(0.0001)	(0.0926)	(0.0915)	(0.0937)	(0.0010)	(0.0010)	(0.0009)
Observations	11,889	11,889	11,889	11,889	11,889	11,889	11,889
R-squared	0.0290	0.0287	0.0548	0.0552			
Panel C: Contribution to community problem solving (mean=0.63)							
Rate of violence	-0.0002	-0.2044**	-0.2062**	-0.2038**	-0.0021**	-0.0022**	-0.0020**
	(0.0001)	(0.0875)	(0.0860)	(0.0881)	(0.0009)	(0.0009)	(0.0009)
Observations	11,869	11,869	11,869	11,869	11,869	11,869	11,869
R-squared	0.0126	0.0129	0.0472	0.0477			
Indiv. controls	No	No	Yes	Yes	No	Yes	Yes
Muni. controls	No	No	No	Yes	No	No	Yes

Note: Column (1) presents the OLS results, Column (2) the reduced form, Column (3) the reduced form including individual controls, Column (4) adds municipal controls, Column (5) the 2SLS estimation, and Column (6) and (7) the 2SLS estimation with individual and municipality controls. Standard error in parenthesis (robust clustered at the municipality level). All regressions include municipality and year fixed effects. Individual control variables include: gender, schooling, age, race, income, media consumption on radio, TV and newspapers, dummy for rural area. Municipality controls: number of students attending to school, fiscal performance, GDP. The dependent variable is the rate of violence, which corresponds to the sum of 19 violent indicators divided by population. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 4: Effect of conflict on social capital at different points of the distribution - Comparison with other countries

Dependent variable:	(1)	(2)	(3)	(4)
Trust	Mean World	US	Mexico	Brazil
	2SLS	2SLS	2SLS	2SLS
Rate of violence	0.4078**	0.4093**	-0.0028***	-0.0058***
	(0.1937)	(0.2012)	(0.0008)	(0.0018)
Constant	1.0462***	1.0555***	0.7523***	0.9066***
	(0.1206)	(0.1275)	(0.0446)	(0.0927)
Observations	874	902	10,003	7,627
Rate of violence	4.3	5.3	13	23

Note: The table shows the effect of conflict on different parts of the violence distribution, which corresponds to the average level of violence in other countries. The coefficients are estimated by using 2SLS. Individuals and municipality controls are included. Standard error in parenthesis (robust clustered at the municipality level). All regressions include municipality and year fixed effects. Individual control variables include: gender, schooling, age, race, income, media consumption on radio, TV and newspapers, dummy for rural area. Municipality controls: number of students attending to school, fiscal performance, GDP. The dependent variable is the rate of violence, which corresponds to the sum of 19 violent indicators divided by population. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 5: Mechanism I: fear to get involved with the community

Dependent variable	(1) OLS	(2)	(3) Reduced form	(4)	(5)	(6) 2SLS	(7)
Panel A: Fear to participate in community org. (mean=0.36)							
Rate of violence	0.0001	0.5646***	0.5599***	0.4405**	0.0026***	0.0025***	0.0020**
	(0.0002)	(0.2068)	(0.2044)	(0.2117)	(0.0009)	(0.0009)	(0.0010)
Observations	8,578	8,578	8,578	8,578	8,578	8,578	8,578
R-squared	0.0157	0.0164	0.0430	0.0441			
Panel B: Fear to run in local elections (mean=0.53)							
Rate of violence	0.0003	0.8328***	0.8356***	0.7872***	0.0037***	0.0037***	0.0034***
	(0.0002)	(0.2440)	(0.2375)	(0.2465)	(0.0011)	(0.0011)	(0.0011)
Observations	6,411	6,411	6,411	6,411	6,411	6,411	6,411
R-squared	0.0313	0.0327	0.0867	0.0873			
Panel C : Fear to vote in local elections (mean=0.15)							
Rate of violence	0.0001	0.0036	0.0217	-0.0445	0.0000	0.0001	-0.0002
	(0.0001)	(0.1598)	(0.1583)	(0.1639)	(0.0007)	(0.0007)	(0.0007)
Observations	8,002	8,002	8,002	8,002	8,002	8,002	8,002
R-squared	0.0201	0.0201	0.0424	0.0434			
Indiv. controls	No	No	Yes	Yes	No	Yes	Yes
Muni. controls	No	No	No	Yes	No	No	Yes

Note: Column (1) presents the OLS results, Column (2) the reduced form, Column (3) the reduced form including individual controls, Column (4) adds municipal controls, Column (5) the 2SLS estimation, and Column (6) and (7) the 2SLS estimation with individual and municipality controls. Standard error in parenthesis (robust clustered at the municipality level). All regressions include municipality and year fixed effects. Individual control variables include: gender, schooling, age, race, income, media consumption on radio, TV and newspapers, dummy for rural area. Municipality controls: number of students attending to school, fiscal performance, GDP. The dependent variable is the rate of violence, which corresponds to the sum of 19 violent indicators divided by population. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 6: Mechanism II: inability to distinguish enemy

Dependent variable: *Trust*	(1) Below 90/10 ratio 2SLS	(2) Above 90/10 ratio 2SLS	(3) Below elf mean 2SLS	(4) Above elf mean 2SLS
Rate of violence	-0.0025***	-0.0147*	-0.0025**	-0.0107**
	(0.0009)	(0.0086)	(0.0010)	(0.0043)
Observations	7,966	3,716	5,848	5,834

Note: Column (1) presents the IV results for observations below the 90/10 ratio, Column (2) for above the 90/10 ratio. Column (3) results for observations belows the ethnic-linguistic fractionalization index (elf). Column (4) for above the elf. Standard error in parenthesis (robust clustered at the municipality level). All regressions include municipality and year fixed effects. Individual control variables include: gender, schooling, age, race, income, media consumption on radio, TV and newspapers, dummy for rural area. Municipality controls: number of students attending to school, fiscal performance, GDP. The dependent variable is the rate of violence, which corresponds to the sum of 19 violent indicators divided by population. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table 7: Heterogeneous effect of conflict on trust (each row represents the interaction between violence and characteristics of the population estimated from separate regressions)

Dependent variable: *Trust*	Coefficient from interaction
Rate of violence*Close victim (=1)	-0.0019
	(0.0021)
Rate of violence*Women (=1)	-0.0003
	(0.0017)
Rate of violence*Black (=1)	-0.0039
	(0.0088)
Rate of violence*Rural (=1)	-0.0010
	(0.0020)
Observations	11,651

Note: The table shows the heterogeneous effect of conflict on different groups of the population. I present the interaction coefficient between the rate of violence and a dummy variable for whether the person has a close relative who was victim of conflict, whether the person is woman, black, or whether the persons lives in rural areas. Standard error in parenthesis (robust clustered at the municipality level). All regressions include municipality and year fixed effects. Individual control variables include: gender, schooling, age, race, income, media consumption on radio, TV and newspapers, dummy for rural area. Municipality controls: number of students attending to school, fiscal performance, GDP. The dependent variable is the rate of violence, which corresponds to the sum of 19 violent indicators divided by population. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

A Figures and Tables

Table A1: Ecological conditions for growing coca

Variables	Optimum range	Units	Source
Altitude	300-2000	masl	Plowman (1979)
Temperature	20-30	C	Acock *et al.* (1996)
Precipitation	500-4000	mm year	Plowman (1979)
PH	<=6	pH	Johnson *et al.* (1997)
Relative humidity	55-85	%	Johnson *et al.* (1997)
Photosynthetic Photon Flux Density	<=400	$\mu mol\, m^{-}2s^{-}1$	Acock *et al.* (1996)

Note: This table summarizes the optimal conditions for for growing coca according to the ecological literature. These characteristics are used to create a coca suitability index.

Table A2: The coca index predicts positively the coca cultivation for all the municipalities in the country and for the municipalities in the LAPOP sample

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	——Area of coca planted——					
Coca index	1.4490***			3.2297**		
	(0.2123)			(1.3201)		
Coca index (=1 if above the median)		0.8864***			1.0131	
		(0.1679)			(0.9555)	
Coca index (1 sd. above the mean)			0.8260***			1.9997**
			(0.1210)			(0.8174)
Sample	Colombia	Colombia	Colombia	LAPOP	LAPOP	LAPOP
Observations	1,116	1,116	1,116	55	55	55
R-squared	0.0401	0.0243	0.0401	0.1032	0.0212	0.1032

Note: Robust standard errors in parenthesis. The area of coca planted is measured in hectares, which is equivalent to 10,000 km^2. Column 1, 2 and 3 shows the relationship between the coca index and the area of coca planted for the entire Colombia. Column 4, 5 and 6 presents the same information but for the sample used in this paper. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Coca index vs. historical measures of conflict

Dependent variable	(1) Land conflicts 1901-1917 (=1)	(2) Land conflicts 1918-1931 (=1)	(3) Political violence 1948-1953 (=1)
Coca index	-0.0163	0.222	0.0190
	(0.161)	(0.183)	(0.1256)
Constant	0.213*	0.188	0.1003
	(0.107)	(0.122)	(0.0839)
Observations	55	55	55

Note: The table includes as dependent variable dummies for whether there were historic conflicts in a municipality during different periods of time. All the regressions include municipality and year fixed effects. All columns show that there is not a correlation between the coca index and the different measures of historical violence. *** p<0.01, ** p<0.05, * p<0.1.

Table A4: Test for monotonicity assumption - First stage

Dependent variable: *Rate of violence*	(1) 1^{st}*quartile*	(2) 2^{nd}*quartile*	(3) 3^{rd}*quartile*	(4) 4^{th}*quartile*
Coca index*Ln(External Shock)	3.6220***	6.9847***	4.1395*	136.1161***
	(0.8294)	(0.4407)	(2.1221)	(20.0467)
Observations	2,988	2,974	2,969	2,974
R-squared	0.7849	0.7456	0.4191	0.7111

Note: The table shows the results from the first stage of an instrumental variable regression for the different quartiles in the distribution of violence. Standard error in parenthesis (robust clustered at the municipality level). The dependent variables is the rate of violence. All regressions include municipality and year fixed effects. The results come from a 2SLS estimation. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table A5: LIML estimators

	(1)	(2)	(3)	(4)
Dependent variable	Trust	Participation	Contribution	Mean of all
Rate of violence	-0.0031***	-0.0019	-0.0020**	-0.0024***
	(0.0011)	(0.0012)	(0.0010)	(0.0008)
Observations	11,682	11,889	11,869	11,905
R-squared	0.0002	0.0584	0.1295	0.0194

Note: This table replicates the results from Table 3, under LIML estimators (The LIML is an estimator that is less efficient, but also less biased to weak instruments (Angrist & Pischke, 2008)). Standard error in parenthesis (robust clustered at the municipality level). All regressions include municipality and year fixed effects. Individual control variables include: gender, schooling, age, race, income, media consumption on radio, TV and newspapers, dummy for rural area. Municipality controls: number of students attending to school, fiscal performance, GDP. The dependent variable is the rate of violence, which corresponds to the sum of 19 violent indicators divided by population. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table A6: IV estimation with all variables collapsed to the municipality level

	(1)	(2)
Variables	Rate of violence	Trust
Coca index* Ln(External Shock)	57.66***	
	(14.42)	
Rate of violence		-0.0015**
		(0.0007)
F Cragg-Donald	16.38	
N	440	440

Note: Column (1) presents the first stage, and column (2) the 2SLS estimates for the sample collapsed at municipality level. So that the standard F Cragg-Donald can be computed as the errors are not clustered. Robust standard errors in parenthesis. p<0.01, ** p<0.05, * p<0.1.

Table A7: Conflict and Trust: Allowing for spatial correlation

	(1)	(2)
Estimation	*Using neighboring violence*	*Spatial correlation*
	Trust	Trust
	2SLS	2SLS
Rate of violence	-0.0035**	-0.0031*
	(0.0017)	(0.0020)
Observations	11,682	11,682

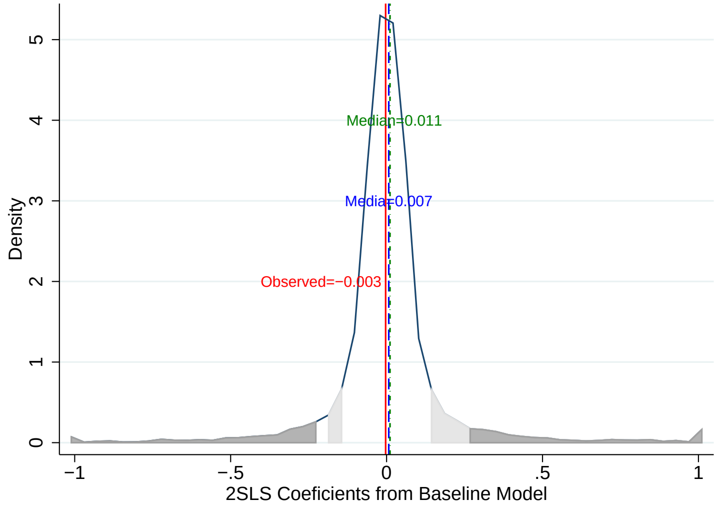
Note: Column (1) presents the estimation of the basic specification, but replacing the violence for the average of violence in the neighboring municipalities. Column (2) replicates the basic estimation but allowing the standard error to be spatially correlated as in (Conley, 2016). All regressions include municipality and year fixed effects. Individual control variables include: gender, schooling, age, race, income, media consumption on radio, TV and newspapers, dummy for rural area. Municipality controls: number of students attending to school, fiscal performance, GDP, night-light intensity. The dependent variable is the rate of violence, which corresponds to the sum of 19 violent indicators divided by population. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Table A8: Main results of the effect of conflict on social capital for migrants and non-migrants

Dependent variable: Trust	(1)	(2)
Sample	*Migration=1*	*Migration=0*
Rate of violence	0.0002	-0.0032***
	(0.0031)	(0.0010)
Observations	829	10,853

Note: Column (1) presents the estimation of the basic specification (equation 1) for individuals who have migrated in the last 5 years. Column (2) replicates the same estimation but for individuals who have not migrated in the last 5 years. Standard error in parenthesis (robust clustered at the municipality level). All regressions include municipality and year fixed effects. Individual control variables include: gender, schooling, age, race, income, media consumption on radio, TV and newspapers, dummy for rural area. Municipality controls: number of students attending to school, fiscal performance, GDP, night-light intensity. The dependent variable is the rate of violence, which corresponds to the sum of 19 violent indicators divided by population. $*p < 0.1$; $**p < 0.05$; $***p < 0.01$.

Figure A1: Distribution of 2SLS Coefficient Estimates Using Randomized Conflict Allocations



Note: The density plot depicts the distribution of 2SLS coefficient estimates using the set of baseline controls with 10,000 draws of randomized allocation of violence among municipalities in Colombia. The dark shaded area indicates the bottom top 5%. The light shaded area shows the top and bottom 10%. The kernel density function and percentiles are estimated on the full set of 10,000 iteration.

B Data

1. Detailed individual controls

    - Age: Continuous variable are you? (question *q*2)
    - Rural: Dummy variable that varies on individual level. 1=Urban, 2=Rural (question *ur*)
    - Education: Continuous variable. Years of education (question *ed*)
    - Gender: 0=Male, 1=Female (question *q*1)
    - Black: Takes the value of 1 if individual report himself to be black. (question *cetid*)
    - Radio: “Do you listen the news in the radio? ” Takes the value of 1 if individual report to listen the radio ” Every day” or ” at least once a week” , 0 if the answer is ” seldom” or ” never” (question *A*1)
    - TV: “Do you watch the news in the TV?” Takes the value of 1 if individual report to watch the TV ” Every day” or "at least once a week", 0 if the answer is "seldom” or ” never” (question *A*2).
    - Newspapers: “Do you read the newspapers?” Takes the value of 1 if individual report to read the newspapers ” Every day” or "at least once a week", 0 if the answer is "seldom” or "never" (question *A*3).
    - Internet: “Do you read the news on the Internet?” Takes the value of 1 if individual report to read news on the Internet ” Every day” or ” at least once a week” , 0 if the answer is ” seldom” or ” never” (question *A*4).
    - Work: “Do you work?” (question *exc*1).

    Imputation of missing data The method for imputing the missing data consists on calculating the mean value per year and municipality (Only 0.36% of the observations were missing values).

2. Data used for constructing coca suitability index
    Come form different sources and it is available at 30 seconds or 0.0083 degrees spatial resolution, which is approximate ($\sim 1\,km^2$)[50].

    Temperature

    - Mean temperature per year for period 1970-2000
    - Source: *CliMond* (Version 2)
    - Freely available at: WorldClim - Global Climate Data (Free climate data for ecological modeling and GIS)
    - Variable used: *Bio*001

    Precipitation

    - Mean precipitation per year (*mm*) for period 1970-2000

[50]For replication run program $Index_replication.py$ in *Python* and *ArcGis*.

- Source: $CliMond$ (Version 2)
- Freely available at: WorldClim - Global Climate Data (Free climate data for ecological modeling and GIS)
- Variable used: $Bio012$

Relative Humidity

- This variable is estimated by using conventional formula for relative humidity (Unwin 1980)[51]:

$$Relative\,humidity = \frac{Vapor\,pressure * 100\%}{Saturated\,vapor\,pressure}$$

- information for $vapor\,pressure$ comes from $CliMond$ (Version 2). Freely available at: WorldClim. Whereas information for $Saturated\,vapor\,pressure$ is constructed by using temperature data and following formula (Mitchell et al. (2004))[52]:

$$Saturated\,vapor\,pressure = 6.107 * exp\left(\frac{17.38 * Temperature}{239 + Temperature}\right)$$

Altitude

- Meters above sea
- Source: U.S Geological Survey's Center for Earth Resources Observation and Science (EROS), with contribution of National Aeronautics and Space Administration (NASA), United Nations Environment Programme/Global Resource Information Database (UNEP/GRID), U.S. Agency for International Development (USAID), Instituto Nacional de Estadistica Geografica e Informatica (INEGI) of Mexico, Geographical Survey Institute (GSI) of Japan, Manaaki Whenua Landcare Research of New Zealand, and Scientific Committee on Antarctic Research (SCAR).
- Freely available at: GTOPO30

Ph Soil

- Measure for acidity and alkalinity of soil
- Source: Harmonized World Soil Database $v\,1.2$
- Freely available at: FAO Soil, created in 2004
- Variable used: $T_P H_H 20$

[51]More information here
[52]More information here.

Chapter III

Comparing Labor Income Risk Across Countries

William F. Maloney*

Melissa Rubio†

Abstract

This paper generates comparable measures of labor market risk across the development process from repeated cross sections of labor market surveys, and identifies patterns across demographic and employment categories. It identifies a striking and very significant negative correlation between risk and the level of development. Developing country workers do seem to face more labor income risk. The downward gradient reflects greater labor market risk among the self-employed and lower female labor market participation.

*Development Economics Research Group and Equitable Growth, Finance and Institutions, The World Bank.

†University of Gothenburg, Sweden

1 Introduction

The development process is commonly characterized as a progressive rise in incomes. However, equally important has been a quest to reduce the risk that workers and households face, partly through the development of institutions such as employment or health insurance. Mainstream theory incorporates risk aversion into its standard utility functions and an empirical literature suggests that differences in risk are of first order importance.[1] On these grounds alone, a measure of income risk merits a place among standard welfare proxies such as GDP per capita income, or the GINI coefficient for income distribution.

However, there are several other reasons why income risk should be of greater concern. First, education is an asset as any other and hence a rise in risk in the return to asset (labor market risk) will depress investment in human capital. Krebs (2003) shows that reducing US income risk to zero would lead to an accumulation in human capital that contributed 0.5% to additional growth. Hence, should poor or informal workers face more risk than formal workers, then we may find a development trap where risky jobs impede the accumulation of capital that would facilitate getting a less risky job.

Second, risk complicates the interpretation of some common measures of welfare. First, as Deaton & Paxson (1994); Bourguignon *et al.* (2004); Storesletten *et al.* (2004) and others note, permanent labor market shocks cause the variance of incomes to rise with age. Hence, even if two countries have an identical initial distribution of incomes as workers enter the labor force, differing levels of labor market risk will lead to differences in the measured income distribution. Further, the demographic profile will also affect measured distribution: a relatively young country will show a lower GINI than an older country even though the underlying income process is identical.

In addition, Krebs *et al.* (2013) show a direct link between income distribution and income mobility through risk. They demonstrate that though an increase in measured mobility (for instance, the Hart index $m_t = 1 - corr(lny_{i0}, lny_{it})$) is generally considered welfare improving, in fact a large component is precisely risk, that is, *bad* mobility. Hence, many conventional measures of mobility are therefore difficult to map to welfare. Indeed, in Mexico, they find the welfare effect of risk-related mobility to be of the same magnitude as that of "good" mobility.

Finally, there is an association of self-employment, the bulk of which is informal in developing countries, with more "vulnerable" types of jobs, which has led to the conclusion that these are inferior positions. Substantial evidence from subjective surveys in Brazil, Mexico and Ghana suggests that this is not necessarily true (See Maloney (1999, 2004); Falco *et al.*

[1] For example, Krebs (2003) simulates that were US income risk to fall to zero, this would be the equivalent of a gain 0.5% growth per year. Krebs *et al.* (2010) show that the increased risk due to trade adjustments with liberalization had substantial welfare losses.

(2015)). However, to date, it has not been possible to separate out transitory movements in income reflecting the bunching of orders, or seasonal shifts from permanent shocks that are of welfare import. Relatedly, it is sometimes asserted that women generally face more risk in their activities.

Yet, to date, we have no common measure that would allow us to compare income risk across countries or in fact establish stylized facts about labor market risk across the development process, demographic categories or job types. In part, this is because panel data sets that would allow variations in individual incomes across long periods of time are few, especially in developing countries, making it difficult to follow individual incomes over time.[2] However, as Deaton & Paxson (1994) and Bourguignon *et al.* (2004), among others note, under the assumption that individual earning dynamics obey some basic properties and follow a simple stochastic process, the main parameters of this process can be recovered from repeated cross-sectional data. The methodology is based on pseudo-panel techniques as pioneered by Deaton & Paxson (1994).[3] The evolution of the variance of incomes within a cohort over time revealed in repeated cross sections offers important information on the nature of those processes. For instance, if shocks to income follow a random walk and are hence permanent, then cohort variance follows a linear trend with cohort age whose slope is the variance of those permanent income shocks and hence, of permanent income risk. In contrast, if income shocks are purely temporary (I win the lottery but tomorrow I'm again dependent on my fixed salary), then the variance of risk does not increase with cohort age and the variance of the permanent component of risk is zero.

This paper exploits the simple property of the unit root process that the slope of the age-variance linear relationship derivable from repeated cross sections is a relevant measure of income risk. The unit root income process assumption has been widely employed (MaCurdy (1982); Abowd & Card (1989); Meghir & Pistaferri (2004); Carroll (1996); Carroll & Samwick (1997a); Gourinchas & Parker (2002); Gottschalk & Moffitt (1994)) and the survey of the empirical findings by Meghir & Pistaferri (2011) in the Handbook of Labor Economics concludes that a (near) random walk component in earnings is not rejected by the existing evidence (see also Hubbard *et al.* (1994); Storesletten *et al.* (2004); Krebs *et al.* (2013); Baker & Solon (2003); Hryshko (2012)), although debate remains.[4] A central debate is the importance of allowing for heterogeneity in growth rates in the estimation, which Guvenen (2009) find reduces the autoregressive component below 1. However, Baker & Solon (2003) find heterogeneity and a random walk component in using a Canadian panel

[2]Loose comparisons of Japan, US, UK and Taiwan in this spirit were offered by Ohtake & Saito (1998) who propose a method to construct synthetic panel data from cross sections to measure poverty mobility.

[3]In our case, the pseudo-panel is formed by following cohorts of randomly selected individuals born in a year interval over time in successive cross-sectional surveys. As in Deaton & Paxson (1994) the idea is as follows: if it may be assumed that all individuals within a cohort face a stochastic earning process that has common characteristics, these characteristics may be recovered at the aggregate level, without observing actual earnings paths. Furthermore, they show that the basic earning dynamics parameter –i.e. the persistence of earnings shocks from one period to the next–recovered from repeated cross-sectional data, or a pseudo-panel are not significantly different from those estimated from a true panel.

[4]See for example,Guvenen (2009); Guvenen & Smith (2014); Gustavsson & Österholm (2014).

while Hryshko (2012) finds little evidence for heterogeneity in the PSID in the United States, and cannot reject the random walk.[5] As seen below, making this assumption allows us to back out some striking stylized facts across countries, demographic categories and job types.

Our analysis uses countries in the Western Hemisphere, for which we can draw data from the The International Income Distribution Data Set (I2D2 - World Bank). The data have a high degree of common survey design and processing. These 20 countries offer substantial variance across income levels, and common demographic and job type variables. The main finding is that overall labor market risk decreases with the level of development. We also document heterogeneous findings along different dimensions: women and the self-employed experience more variance in shocks to their permanent income.

We mainly focus on the risk pertaining to permanent shocks, both because families can smooth transitory shocks (the welfare consequences of which are relatively minor, compared to permanent income shocks[6]). In addition, transitory shocks are inseparable from measurement error in income (Moffitt and Gottschalk (1993), Carroll and Samwick (1995)), making it difficult to make comparisons across differently collected data sets.

We also find that a standard distribution statistic -the Theil index- increases when taking into account labor market risk. In particular, we simulate what would happen to inequality in the US if we allocate labor market risk from Latin American countries. We also show that lowering the labor market risk in Latin American countries would improve inequality measures.

The remainder of the paper is organized as follows. Section 2 presents the theoretical model that motives the empirical analysis. Section 3 describes the data. Section 4 presents the main results and heterogeneous analysis across different socio-demographic groups. Section 5 discusses the difference between the estimates of risk for the US and Latin America. Section 6 incorporates our measure of risk on a standard distribution statistic to show how risk can also affect inequality. Section 7 concludes.

2 Measuring Income Risk

2.1 Theory: Income Process and Risk

At any period in time, individuals in the economy are hit by shocks, both good and bad. These shocks, to the degree that they are unpredictable, constitute risk. Some of these shocks will be short lived. Perhaps a self-employed person has a low (high) business period

[5] The PSID is the Panel Study of Income Dynamics in the US. This study began in 1986 with a nationally representative sample of over 18,000 individuals. It is the longest running longitudinal household survey in the world.

[6] Blundell *et al.* (2008) argue that considerable consumption smoothing takes place in response to transitory shocks but much less for permanent shock. See also Aiyagari (1994), Heaton & Lucas (1996), Levine & Zame (2002).

for a month. Income will fall (rise), but may well soon achieve its previous levels. Similarly, winning the lottery will increase income for one period, after which it will return to previous levels. A worker, knowing these to be short lived is likely to self-insure at the individual, family, or perhaps village level. The welfare impacts are likely to be small for the majority of the population not living a the subsistence level.

Some shocks, however, are persistent. Losing a job may lead to a permanent loss in income if the next job is not as good as the previous. Similarly, an adverse health shock may permanently lower a worker's productivity. In the aggregate, the overall dynamism of the economy and churning in the labor market may generate opportunities for gain, but also make some worker's skills obsolete. Each of these shocks, by virtue of being permanent, are hard to insure against and hence the welfare impacts are large.

Consider a large number of workers indexed by i. For notational ease, we focus on one cohort of workers who enter the labor market for the first time in period $t = 0$ so that $t = 0, 1, \ldots$ stands for both calendar time and age (experience) of the worker. Let y_{it} stand for the labor income of worker i in period t. Following a longstanding tradition in microeconometrics, we postulate that the log of y_{it} is a random variable that is the sum of two components, a persistent component, ω_{it}, and a transitory component, η_{it}[7].

In addition, we set the mean of lny_{it} to μ. In short, we have:

$$log\, y_{it} = \omega_{it} + \eta_{it} + \mu \ . \tag{1}$$

The persistent component, ω_{it}, follows an AR(1) process:[8]

$$\omega_{i,t+1} = \rho\omega_{it} + \epsilon_{i,t+1} \ , \tag{2}$$

where ρ is a parameter measuring the persistence of shocks. The term ϵ denotes a stochastic innovation to labor income, which we assume to be i.i.d. over time and across individuals. We further assume that the transitory component of labor income, η_{it}, is i.i.d. over time and across individuals. Moreover, η_{it} and $\epsilon_{i,t+n}$ are uncorrelated for all t and n. All random variables are normally distributed so that labor income is log-normally distributed. More specifically, we assume that $\epsilon_{it} \sim N(0, \sigma_\epsilon^2)$, $\eta_{it} \sim N(0, \sigma_\eta^2)$, and $\omega_{i0} \sim N(0, \sigma_{\omega_0}^2)$ as in Meghir & Pistaferri (2004).

Put differently, the variance of earnings of a cohort of workers increases as the cohort ages. Two identical workers beginning with identical incomes receive permanent shocks in one period which generates variance in their incomes. But the next period they receive

[7]See Gottschalk & Moffitt (1994) and Carroll & Samwick (1997b) for similar specifications and Baker and Solon (2003) for a detailed discussion.

[8]In Meghir & Pistaferri (2004) the log earnings are generated by the sum of a random walk process, where innovations are i.i.d. Gottschalk & Moffitt (1994) also allows the permanent component to follow a random walk.

another shock on top of the previous shock. This, on average, leads to even greater variance in incomes. $\rho < 1$ implies that some fraction of this shock dissipates. However, as ρ in the literature is frequently estimated as close to 1 , as workers age these shocks compound to generate a larger and larger variance in incomes within a given cohort. Deaton & Paxson (1994) explore the distribution of consumption for several countries broadly through this lens.

Equations (1) and (2) together imply that:

$$ln\, y_{it} = \rho^t \omega_{i0} + \sum_{n=0}^{t-1} \rho^{t-n-1} \epsilon_{i,n+1} + \eta_{it} + \mu \ . \tag{3}$$

Thus, labor income in period t is determined by initial condition, ω_0, and stochastic changes, the latter being represented by the transitory shocks, η, and permanent shocks, ϵ. From (3) and our assumptions about ϵ, η, and ω_0 it follows that expected labor income is $E[lny_{it}] = \mu$ and labor income uncertainty before ω_{i0} is known is given by

$$var[lny_{it}] \ = \ \begin{cases} \rho^{2t}\sigma^2_{\omega_0} + \sigma^2_\eta + \frac{1-\rho^{2t}}{1-\rho^2}\sigma^2_\epsilon \quad if\ \rho \neq 1 \\ \sigma^2_{\omega_0} + \sigma^2_\eta + t\sigma^2_\epsilon \quad if\ \rho = 1 \end{cases} . \tag{4}$$

Krebs *et al.* (2013) show that, the case where $\rho < 1$, it is possible to identify all the parameters above when both panel and repeated cross sectional data are available. This situation is unusual in the developing world, and for a variety of reasons cross-sectional data are more common than panel data. Panel data can be costly and pose a logistical challenges. However, if we make the simplifying assumption that the permanent component follows a random walk (ρ=1), then estimation is substantially simplified, and estimates could be generated for any country with repeated cross sections.

In this case, the second line of (4) becomes a linear relationship between variance of income and Age (t). Figure 1 shows this graphically. The coefficient on age yields σ^2_ϵ, and this will be our measure of the permanent income risk. The intercept $\sigma^2_{\omega_0} + \sigma^2_\eta$ yields a combination of the initial distribution of income when workers enter the work force at, say, age 20, and transitory shocks or measurement error.[9] From the point of view of distribution, $\sigma^2_{\omega_0}$ is arguably a "base" measure of inequality of a workers entering the work force. It is, of course, affected by measurement error, as all other distribution statistics, but it is measured before labor market shocks, and abstracts from demographic differences across countries. Across the rest of life, the distribution of within-cohort income is importantly determined by good and bad shocks. An important part of the measured distribution of income is therefore risk.

This also points to an important compositional issue in standard measures of distribution. Imagine two countries where the distribution for any cohort is identical. For instance 20 year olds have the identical variance in both cases, and 40 year olds have identical vari-

[9]See also Bourguignon *et al.* (2004) for a comparison of pseudo panel vs. actual panel estimates.

ances, although greater than that for 20 years old. But suppose that the second country has relatively more 40 year olds. It will appear to have a "worse" distribution of income even though the underlying income dynamics are exactly the same. Back of the envelope calculations suggest, for example, that were relatively young Bolivia to have Europe's age structure, the Gini might be 7 points higher. In principle, it is possible to use the variances calculated above to simulate how this demographic evolution will affect measured distribution.

2.2 Estimating Risk

In practice, since we are interested in the unpredicted part of income movements (e.g., not arising from, for instance, the simple accumulation of seniority or experience) the relevant income measure is, in fact, the mincer residual: log labor income, $ln\ y_{it}$, is specified as in (1) and it is assumed that the deterministic mean component, μ, depends on $x_{it} = (x'_{it}, z_{it})$, where z_{it} denotes the age of worker i in year t and x'_{it} is vector of observable individual characteristics beyond age (education, education^2, gender). We also make the assumption $\mu_t(x'_{it}, z_{it}) = \lambda_t + \lambda(x') \cdot x'_{it} + \sum_z \lambda(z)\delta(z_{it})$, where λ_t is a constant that varies by calendar time period (thus absorbing the effects of macroeconomic factors such as aggregate productivity growth and aggregate economic fluctuations on income), $\lambda(x')$ is a vector of coefficients for the vector of worker characteristics x', and $\delta(z_{it})$ are age-dummies, giving the predictable component of income. Thus, log labor income can be written as for each of the countries in our sample:

$$lny_{it} = \lambda_t + \lambda(x') \cdot x_{it} + \sum_z \lambda(z)\delta(z_{it}) + v_{it} \qquad (1')$$

$$v_{it} = \omega_{it} + \eta_{it}$$

Equation (1') resembles a typical Mincer specification for labor income for which the residual, v_{it}, is the sum of two unobserved stochastic components, ω_{it} and η_{it}. As in Carroll & Samwick (1997b), the residual from equation (1'), v_{it}, can then be used to estimate, in a second step, the parameters of interest.

For notational simplicity, assume that all individuals i "are born" in period $t = 0$, so that t and z simultaneously stand for age of the individual and calendar time. Equations (1) and (2) which describe our labor income process imply that the change in residual income variance with age is given by:

$$var[v_{iz}] = var[(\omega_{iz} + \eta_{iz})] = \sigma^2_\eta + \sigma^2_{\omega 0} + z * \sigma^2{}_\epsilon \qquad (4')$$

Equation (4') links the changes in cross sectional residual income variances for any age cohort z with our parameters of interest and can be estimated by OLS as:

$$var[(\omega_{iz} + \eta_{iz})] = \sigma^2_0 + \sigma^2_\epsilon * Age \qquad (5)$$

where $\sigma^2_0 = \sigma^2_{\omega_0} + \sigma^2_\eta$ and the coefficient on Age gives the cross-sectional variance.

Figure 1 plots this relationship between variance and age and shows that the first two terms, initial distribution and transitory shocks/measurement error, correspond to the constant and the permanent component of risk is backed out from the slope of the age relationship. To measure the impact on either σ_0^2 or the permanent share of risk σ_ϵ^2 in a given demographic category, or sector of employment, τ, we can interact a dummy for the category of interest with age and the coefficient will measure the difference of category τ's variance with that of the "base" category'.

$$var[v_{iz}] = \sigma_0^2 + \delta\sigma_{0\tau}^2 + \sigma_\epsilon^2 * Age + \delta\sigma_\tau^2 * Age \quad (6)$$

where clearly σ_0^2 and $\sigma_\epsilon^2 * Age$ correspond to the omitted "base" category. The coefficients on the dummies interacted with age give the difference with the base.

3 Data

The labor market outcomes data come from the International Income Distribution Survey (I2D2). The I2D2 is a harmonized collection of nationally representative household and labor surveys from 1990- 2013 constructed and maintained by the research group of the World Bank.[10] They have standardized the variable definitions across economies and time periods. The I2D2 set includes repeated cross-section data from Latin American countries. These data contain information on on monthly wages, hours worked, schooling, age, gender, industry, and experience at the individual level. These survey variables are comparable across countries. Family aid workers and apprentices were eliminated because their wages do not reflect market productivity. The unemployed and people who work in voluntary services were excluded as well. Table A1 presents the country summary statistics for the main labor outcomes used in this paper.

We limit our samples to the working life of prime age workers 25-55.[11] The main reason to chose this age range is that we do not have reliable information on how many hours young people work, and workers are more likely to retire after 55 years in informal economies and live from family's support. An important concern related with the data in all countries is the presence of outliers and following much of the literature, we delete the extreme values by trimming the top and bottom 0.5 percent of the earnings distribution within age-education-year cells (Gottschalk & Moffitt (1994); Blundell *et al.* (2013)).

[10] Only 25% of the observations were collected before 2000, and 75% of the observations come from the period 2000-2013.

[11] Though other literature, for instance, Storesletten *et al.* (2004) use a broader age group, our preliminary results suggest that workforce entry takes longer and retirement starts earlier in many developing countries.

4 Results

4.1 Levels of Risk

As discussed in section 2.2, we begin by regressing income on standard Mincerian variables to remove predictable determinants of income such as education, and evolution over time, such as the accumulation of experience. We then regress the residuals from the Mincer regression on a constant, a full set of age dummies, and year dummies, and plot the coefficients on the age dummies as in equation 4'.[12] As in Deaton & Paxson (1994), the slope of the relation between age and variance is what drives our estimate of σ_ϵ^2.

Figures 2 plot the variance of earnings against age for Latin America and the US and finds that in every case i) The variance increases in age, and ii) The plot is generally linear, consistent with a near-unit root income process. The figures are in line with the evidence presented by Storesletten *et al.* (2004) for the US, who used the PSID. They show an intercept of 0.31, whereas in our case, Figure 2 shows a slighter higher intercept of 0.33 for the US.[13] Though there is substantial deviation from trend, there is a significant downward relationship between the level of permanent risk faced by workers, and the log of GDP in development.

Figure 3 plots the estimated variances for each of the 20 countries against income. What emerges is a very significant downward relationship between overall labor market risk and consistent with the observed upward slopes, all estimates are above zero and, now there is a downward gradient between income and risk. The US here is around .002 while the mean for the poorer countries like Bolivia, Honduras and Bolivia is close to 5 times that. Poorer countries show much higher measures of income risk.

In what follows, we attempt to identify what may lie beneath this gradient.

4.2 Disaggregating Measured Income Risk

4.2.1 Estimation Approach

Clearly these differences can arise either because of higher variance within categories, or differing labor market composition of sectors with intrinsically different variances. Figure 4, for example, shows systematic differences in share of self-employment, share of women in the work forces, and educational attainment across the development process. To identify the risk associated with individual sectors or categories, we interact a corresponding indicator variable (dummy) with the age variable. We necessarily include the dummy free standing as well which can be interpreted as differences in either measurement error or the transitory component. Again, we focus exclusively on the permanent component, both because arguably families can smooth transitory shocks and hence they have a lesser welfare effect,

[12] We used household sampling weights in the main estimations.

[13] The results are similar whether the wages are measured weekly or yearly, consistent with Card (2004) (not shown).

and because the measurement error component makes it difficult to compare across countries.

Tables 1 and 2 present the approach going forward for a representative LDC country, Mexico, and advanced country, the US using more complete data sets to identify what types of workers have shown a higher association with risk. Both tables present very similar stories. The first columns capture the simple estimate of σ_{ϵ}^2, the coefficient on age, and confirms our graphical findings: the US at .003 and Mexico at .0093 or triple. It is important to highlight that observed differences, particularly between Mexico and the US are not likely to be a question of differences in data management or collection, since most measurement error, scale effects, or other factors that are likely to be invariant to age are captured by the constant as discussed in section 2.2.

The second columns add a gender dummy making the omitted category (the coefficient on age) effectively the risk of men. Again, the interactive term, women*age, captures the difference in measured risk while the free standing woman term captures differences in transitory shocks and measurement error. The coefficient on the former is not significant in Mexico and is negative in the US suggesting that women either experience lower risk, or the same level as men. The age term, which now captures risk pertaining to men, changes very little, suggesting that our aggregate differences in risk are not being driven by differences in gender composition between the two countries.

The third column analogously interacts age with self-employed status. In both countries, the risk to salaried workers, the corresponding coefficient on age, falls. This makes sense in the US since the interactive coefficient is positive suggesting that the self-employed face perhaps double the risk as salaried employees. The Mexico data offers somewhat of a puzzle, however, since while transitory variations appear higher as captured by the very large and significant coefficient on the self-employment dummy, the permanent component of risk appears actually lower. We will return to this puzzle below.

Column 4 analogously controls for education, making the omitted category (age) effectively workers with secondary education. Again, we find inconsistent results between the US and Mexico. In the latter, those with no secondary education show higher risk and those with tertiary lower, while the reverse is the case in the US. This might be explained because a composition effect coming from the fact there are more individuals with secondary education than with tertiary education in Mexico than in the US, relative to the population, as secondary education is the omitted category for comparison.

Finally, the last column combines all covariates so we can isolate, for instance, the influence of self employment vs. education which we know to be highly correlated. Here, the age variable is capturing male salaried workers with a secondary education. We see that in both countries, the measured risk for these "average" workers has fallen substantially, in Mexico by a factor of 4, in the US by a factor of 3. The previous results largely hold although women now appear for face more risk then men in Mexico, but the previous finding in the US of less female risk no disappears. But show no difference in the US. There is now no difference between those with secondary and no secondary education in ether country.

4.3 Estimates for the Complete Sample

These exercises are meant to clarify the approach but clearly point to a need for a broader sample of countries to establish any stylized facts around demographic or employment categories. The next set of figures plot the corresponding compound coefficients (Male, salaried, primary education + interactive term) for the full regressions on each individual country with whisker plots, against income. The solid line represents the average estimate for the sample and the dashed line the zero line against which the whisker plots can be referenced.

Figure 5 plots the age coefficient representing salaried males with primary education. The majority of estimates are statistically different from zero. A linear relationship between the risk faced by salaried males and income levels emerge. The observed pattern thus appears consistent with the gradient presented in figure 3.

Figure 6 suggests consistent premium for self-employment. For most countries, dots allow us to reject that the interactive term is zero. Among those coefficients significantly different from zero, Bolivia, Honduras, and Nicaragua suggest self-employment is riskier; whereas Chile, Argentina and the US are less risky. Thus, the self employed arguably share the same downward relationship with development found among salaried males. The higher risk estimated for self-employment in poorer countries helps to understand the linear downward slope found in figure 3.

Figure 7 suggests slightly lower risk for women on average although there appears to be an upward trend with income. Peru, Chile, Uruguay and the US show no significant gender gap in risk. While Ecuador, Colombia, Brazil, Venezuela, Mexico and Argentina show modestly higher female risk, around .001. Lower income Bolivia, Nicaragua, Honduras, and Guatemala suggest lower risk for women. Together, a positive relationship emerges with income: Women face relatively less risk relative to men in poorer countries than rich and substantially so. Figure 4 suggest that this also contributes to the gradient since in those countries women participate perhaps 10pp less than in the US. However, again the gradient is driven by effectively the four poorest countries.

Figure 8 and 9 present the risk profile for secondary and post-secondary workers. The relationship to income levels is not linear, workers with secondary education seem to face less risk in poor countries. The risk profile of post-secondary workers in 9 shows a negative but insignificant difference with secondary workers and no significant gradient.

In sum, the gradient in figure 3 may be partially driven by a risk gradient of prime age salaried males with secondary education and perhaps a lower participation of women. The impact of increased education across the development process is not clear as both secondary and post secondary workers show lower risk than secondary workers and contract as offsetting margins in the development process.

We might expect compositional effects to explain part of the pattern in figure 3. For instance, weighting the difference in risk by the difference in post secondary education from roughly 10 percent of the work force in Honduras to 70 percent in the United states can account for 0.000795 and the difference in self employment from 50 percent to 10 percent another 0.0004888 for a total of 0.0012838. This can only explain under 10 percent of the total gap in figure 3.

On the other hand, within the salaried, male, primary educated category, we can explain .003-.004 points and within self-employment, .005. Hence the mystery shifts to what macro or institutional drivers might be of this within group variance in risk. Next attempts to understand whether the differences between Latin American countries and the US are explained by compositional effects (e.g. poorer countries having more self-employed workers), or just because certain characteristics of workers in poorer countries (e.g. self-employed are inherently facing more risk in poorer countries.)

5 What Explains the Difference Between Risk in the US and Latin America?

Previously, we tried to estimate which socio-demographic groups of the population were explaining the high levels of risk found in poorer countries. In this subsection, we attempt to identify which covariates may drive differences in risk between the United States and countries in Latin America. In addition, this exercise allows us to determine which part of the risk that we estimated is explained by compositional effects.

One way to decompose the risk gap between the US and Latin American countries is by using an Oaxaca-Blinder decomposition. This allows to split the gap in two components. The first part, due to differing covariates (differences in coefficients). The second component is a composition effect, which is also called the explained effect (by differences in covariates). However, we can have two countries with similar levels of risk, but they can have distinct underlying distributions. Therefore, we use the Machado & Mata (2005) decomposition, and the Recentered Influence Function (RIF) (Firpo *et al.* , 2009). Both allow for a decomposition of the gap between the US and Latin America for the whole distribution.

Table 3, Panel A, presents the raw gap from a simple quantile comparision. Panel B shows Machado-Mata and Melly estimates (MM). Panel C, the RIF estimates. We present the estimates for the 10^{th} , 50^{th} and 90^{th} percentiles. The results are broadly consistent across measures. Panel A shows that gap in all cases is largest at the top of the distribution. Panel B and C show the amount of this gap that can be explained by socio-economic and demographic characteristics. Overall the MM method suggests statistically significant explanatory power of 40% at the 50^{th} percentile. Whereas the remaining 60% is explained by unobserved characteristics, or factors that make Latin American countries more risky. In this paper, however, we can not establish which are these unobserved factors. The RIF presents a similar decomposition but it allow us to identify which variables are most responsible. For instance, post-secondary education workers seem to be explaining lower levels of risk in the US.

6 Risk and the Measurement of Inequality

The fact that standard distribution statistics abstract from mobility across the distribution is well established. However, what is also true is that labor market risk as discussed here is not only a bad type of mobility- risk enters negatively in most standard utility functions, but also confuses distribution statistics on two counts.

First, two societies with identical distributions of wages of job market entrants, but very different variance of shocks, the riskier society will appear more unequal. It is, but often our idea of inequality corresponds closer to $\sigma^2_{\omega_0}$, the beginning distribution of endowments, rather than the riskiness of the subsequent income trajectory. For example, assume two countries with identical $\sigma^2_{\omega_0}$, but one has a less risk averse population, is more entrepreneurial and hence is more often self-employed. Therefore, σ^2_{ϵ} is higher by choice. Income distribution will appear to be worse in the second case, but the implications for welfare are unclear. This calculus, of course, varies to the degree that the choice of the riskier activity is not voluntary, although a substantial literature argues that the self-employed in developing countries are, in fact, voluntarily self-employed.

Second, since wage distribution increases with age, necessarily countries with older populations will mechanically appear more unequal. Hence, the United States is more unequal than Japan, an aging country, than statistics indicate. Further, aging populations will show a natural worsening of income distribution, again, with unclear welfare implications.

How important are these effects? As one exercise, we employ the decomposability property of the Theil statistic. Under the assumption of log normality employed above, Cowell (2011) shows that the Theil-L (or General Entropy 0 index) can be decomposed into:

$$I = \sum_{c=1}^{C} p_c \frac{\sigma^2_{\epsilon,c}}{2} + \sum_{c=1}^{C} p_c ln(\frac{\bar{Y}}{\bar{Y}_c}) (7)$$

Where $\bar{Y}$ is average income of the population in a country, $\bar{Y}_c$ is average income in each group c, C is the number of groups and p_c is the share of the total population in group c. If we treat C as simply the age bins across time for which we have estimated labor market risk, then we can can simply aggregate across age groups to get the component of measured distribution which is due to shocks to income.

In Figure 10, we estimate the Theil index by using the different levels of risk across the Latin American countries, and the US parameters. The purpose is to simulate how the risk would be, if we instead of having low levels of risk as in the US, the US has now the levels of risk of Latin American countries. We see that the higher the levels of risk the more unequal the US would be.

Figure 11 presents a similar exercise, but instead it calculates the Theil index for all Latin American countries, and compares how their Theil index would be if they had the same parameters, but the US levels risk. In all cases, we see that if Latin American countries would have lower levels of risk, even having the same income distribution, and population composition, inequality would decrease.

7 Conclusion

This paper has sought to generate comparable estimates of labor market risk across the development process and to establish some stylized facts across demographic and job categories. Using consistent data for the Americas, we identify a negative relationship of overall labor market risk with development. We do find additional risk to being self employed on average. We also find slightly higher risk for women. The downward relationship with development seems a combination of an across the board downward relationship, higher risk among the self-employed for the poorest countries which also have a large informal sector, and which also show lower risk for women, but lower female participation; and the interaction with the transitory component of self employment with declining labor market share across the life cycle. We also find that labor market risk can influence traditional inequality measures.

References

Abowd, John M, & Card, David. 1989. On the Covariance Structure of Earnings and Hours Changes. *Econometrica*, **57**(2), 411–45.

Aiyagari, S. Rao. 1994. Uninsured Idiosyncratic Risk and Aggregate Saving. *The Quarterly Journal of Economics*, **109**(3), 659–684.

Baker, Michael, & Solon, Gary. 2003. Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records. *Journal of Labor Economics*, **21**(2), 267–288.

Blundell, Richard, Pistaferri, Luigi, & Preston, Ian. 2008. Consumption Inequality and Partial Insurance. *American Economic Review*, **98**(5), 1887–1921.

Blundell, Richard, Low, Hamish, & Preston, Ian. 2013. Decomposing changes in income risk using consumption data. *Quantitative Economics*, **4**(1), 1–37.

Bourguignon, Francois, Goh, Chor-ching, & Kim, Dae Il. 2004 (Aug.). *Estimating individual vulnerability to poverty with pseudo-panel data*. Policy Research Working Paper Series 3375. The World Bank.

Carroll, Christopher D. 1996 (Oct.). *Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis*. NBER Working Papers 5788. National Bureau of Economic Research, Inc.

Carroll, Christopher D., & Samwick, Andrew A. 1997a. The nature of precautionary wealth. *Journal of Monetary Economics*, **40**(1), 41–71.

Carroll, Christopher D., & Samwick, Andrew A. 1997b. The nature of precautionary wealth. *Journal of Monetary Economics*, **40**(1), 41–71.

Cowell, Frank. 2011. *Measuring inequality*. Oxford University Press.

Deaton, Angus, & Paxson, Christina. 1994. Intertemporal Choice and Inequality. *Journal of Political Economy*, **102**(3), 437–67.

Falco, Paolo, Maloney, William F, Rijkers, Bob, & Sarrias, Mauricio. 2015. Heterogeneity in subjective wellbeing: An application to occupational allocation in Africa. *Journal of Economic Behavior & Organization*, **111**, 137–153.

Firpo, Sergio, Fortin, Nicole M., & Lemieux, Thomas. 2009. Unconditional Quantile Regressions. *Econometrica*, **77**(3), 953–973.

Gottschalk, Peter, & Moffitt, Robert. 1994. The Growth of Earnings Instability in the U.S. Labor Market. *Brookings Papers on Economic Activity*, **25**(2), 217–272.

Gourinchas, Pierre-Olivier, & Parker, Jonathan A. 2002. Consumption Over the Life Cycle. *Econometrica*, **70**(1), 47–89.

Gustavsson, Magnus, & Österholm, Pär. 2014. Does the labor-income process contain a unit root? Evidence from individual-specific time series. *Journal of Economic Dynamics and Control*, **47**, 152 – 167.

Guvenen, Fatih. 2009. An empirical investigation of labor income processes. *Review of Economic Dynamics*, **12**(1), 58 – 79.

Guvenen, Fatih, & Smith, Anthony A. 2014. Inferring Labor Income Risk and Partial Insurance From Economic Choices. *Econometrica*, **82**(6), 2085–2129.

Heaton, John, & Lucas, Deborah J. 1996. Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing. *Journal of Political Economy*, **104**(3), 443–487.

Hryshko, Dmytro. 2012. Labor income profiles are not heterogeneous: Evidence from income growth rates. *Quantitative Economics*, **3**(2), 177–209.

Hubbard, R Glenn, Skinner, Jonathan, & Zeldes, Stephen P. 1994. Expanding the Life-Cycle Model: Precautionary Saving and Public Policy. *American Economic Review*, **84**(2), 174–79.

Krebs, Tom. 2003. Human Capital Risk and Economic Growth. *The Quarterly Journal of Economics*, **118**(2), 709–744.

Krebs, Tom, Krishna, Pravin, & Maloney, William. 2010. Trade Policy, Income Risk, and Welfare. *The Review of Economics and Statistics*, **92**(3), 467–481.

Krebs, Tom, Krishna, Pravin, & Maloney, William. 2013 (Jan.). *Income Mobility and Welfare*. IMF Working Papers 13/24. International Monetary Fund.

Levine, David K., & Zame, William R. 2002. Does Market Incompleteness Matter? *Econometrica*, **70**(5), 1805–1839.

Machado, José A. F., & Mata, José. 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. *Journal of Applied Econometrics*, **20**(4), 445–465.

MaCurdy, Thomas E. 1982. The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics*, **18**(1), 83–114.

Maloney, William F. 1999. Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico. *The World Bank Economic Review*, **13**(2), 275–302.

Maloney, William F. 2004. Informality revisited. *World development*, **32**(7), 1159–1178.

Meghir, Costas, & Pistaferri, Luigi. 2004. Income Variance Dynamics and Heterogeneity. *Econometrica*, **72**(1), 1–32.

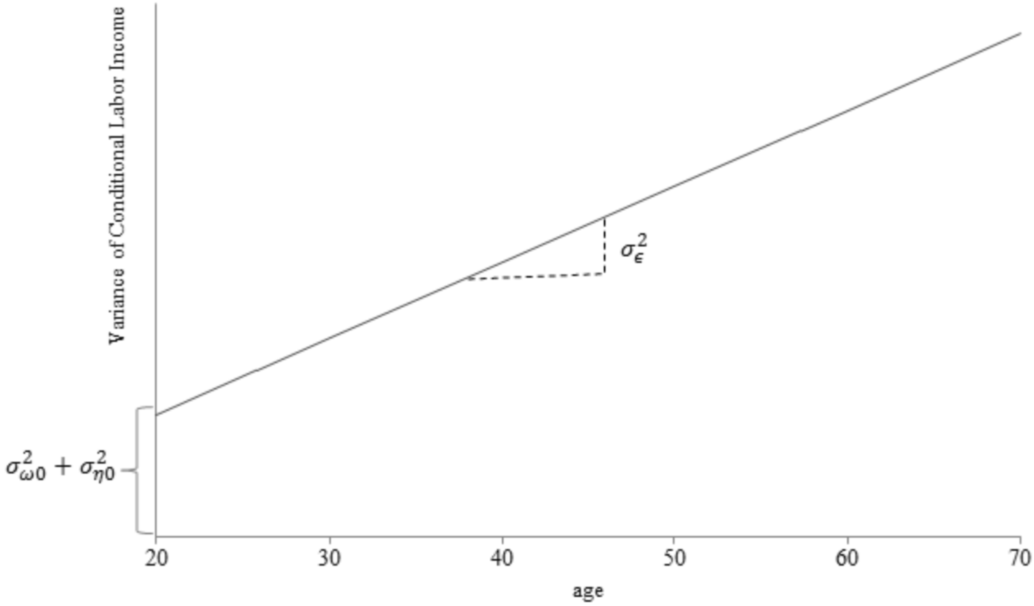
Meghir, Costas, & Pistaferri, Luigi. 2011. *Earnings, Consumption and Life Cycle Choices*. Handbook of Labor Economics, vol. 4. Elsevier. Chap. 9, pages 773–854.

Ohtake, Fumio, & Saito, Makoto. 1998. Population Aging and Consumption Inequality in Japan. *Review of Income and Wealth*, **44**(3), 361–81.

Storesletten, Kjetil, Telmer, Chris I., & Yaron, Amir. 2004. Cyclical Dynamics in Idiosyncratic Labor Market Risk. *Journal of Political Economy*, **112**(3), 695–717.

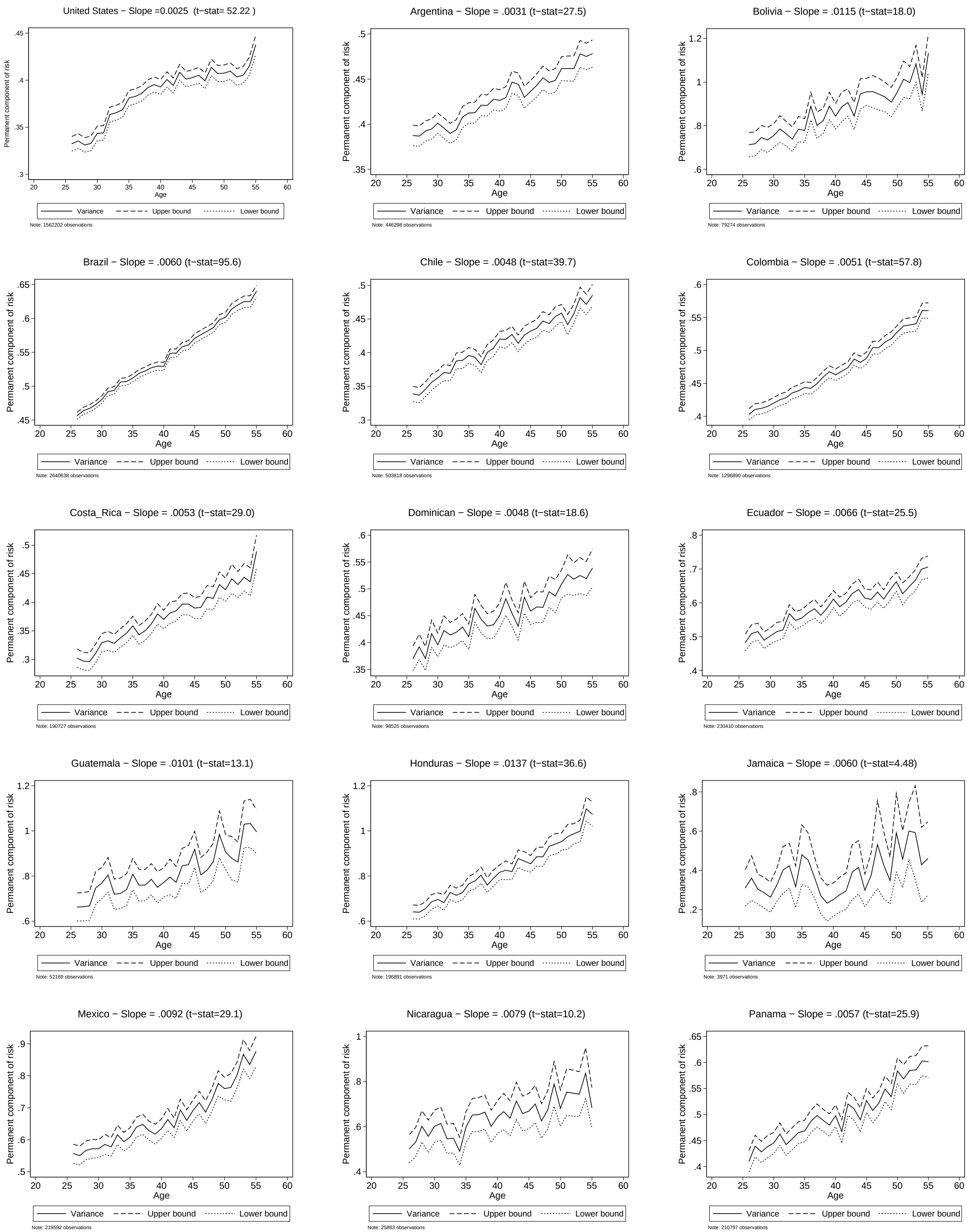
Figures

Figure 1: Income Variance vs. Age

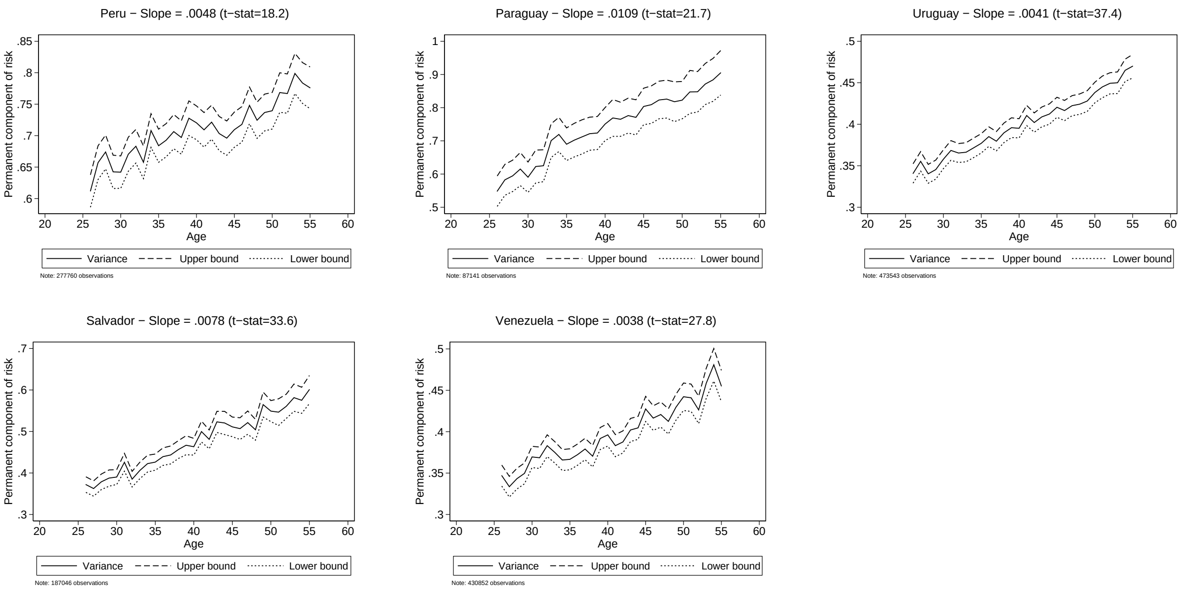


Note: This figure plots the relationship between variance and age as presented in equation 5. It shows the initial distribution and the permanent income. The intercept $\sigma_{\omega_0}^2 + \sigma_{\eta}^2$ yields a combination of the initial distribution of income, and the slope σ_{ϵ}^2 measures the permanent income risk.

Figure 2: Income variance vs. Risk - Latin America and the US

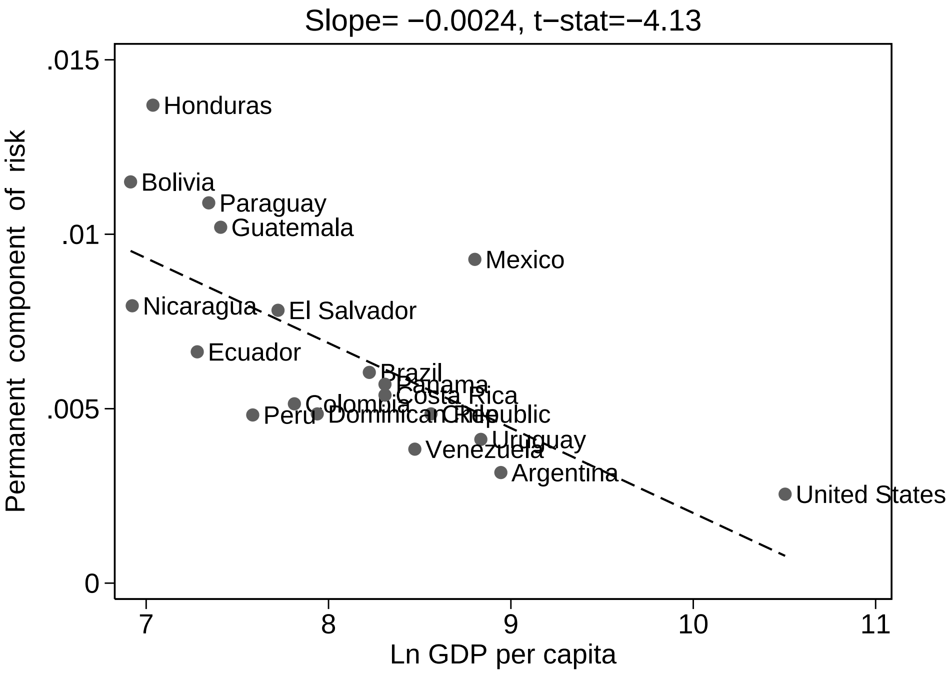


Note: Variance is the coefficient on age from a regression of the Mincer residual squared on age and year dummies. Estimates for different countries.



Note: Variance is the coefficient on age from a regression of the Mincer residual squared on age and year dummies. Estimates are for different countries from equation 5. Data source: I2D2. For sample size see Table A1.

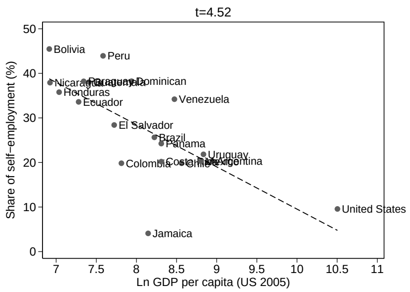
Figure 3: Income risk vs. Ln of GDP per capita. Latin America



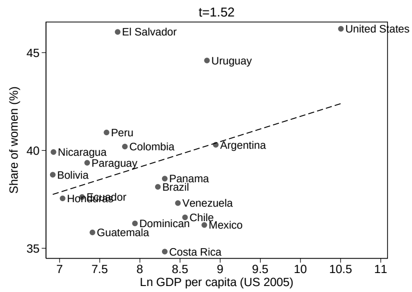
Note: This figure plots the estimated variances of permanent income for each for each country against Ln GDP per capita. Data source: I2D2. Sample size: 20 countries (slopes from Figure 2)

Figure 4: Share of different groups vs. Ln of GDP per capita

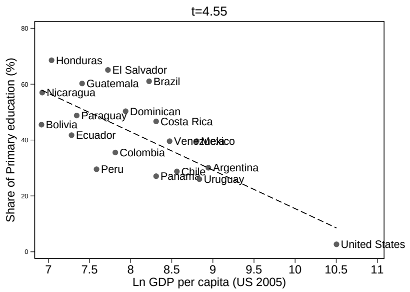
(a) Share of self-employment



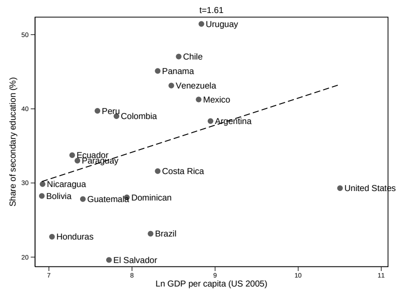
(b) Share of women



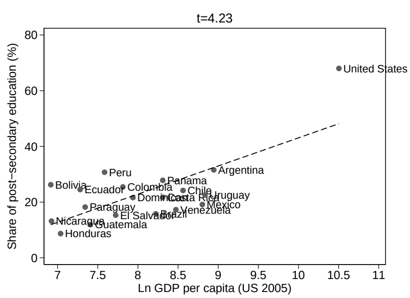
(c) Share of Primary education



(d) Share of secondary education

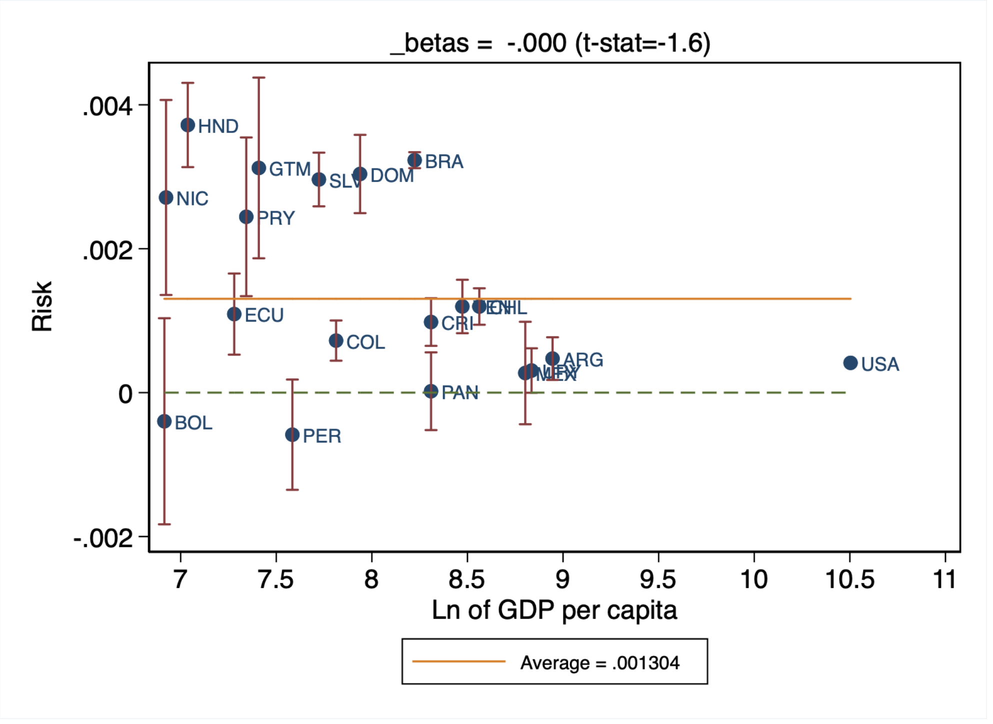


(e) Share of post-secondary education



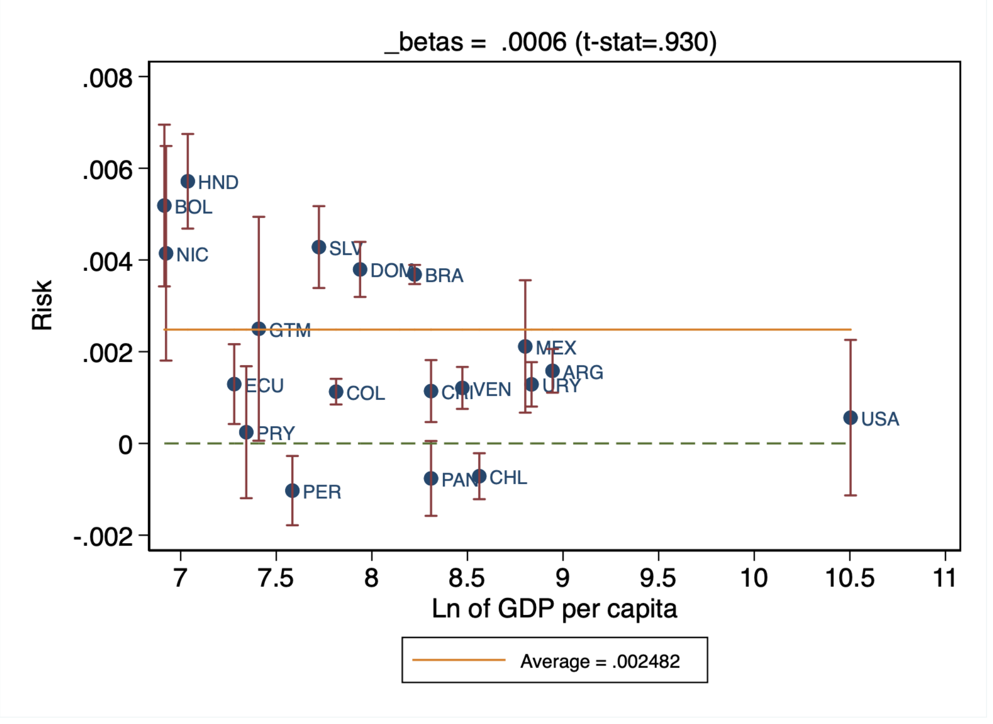
Note: This figure plots the share of different groups in the population against the Ln of the GDP per capita. Data source: I2D2. Sample size: 20 countries.

Figure 5: Risk, Salaried males



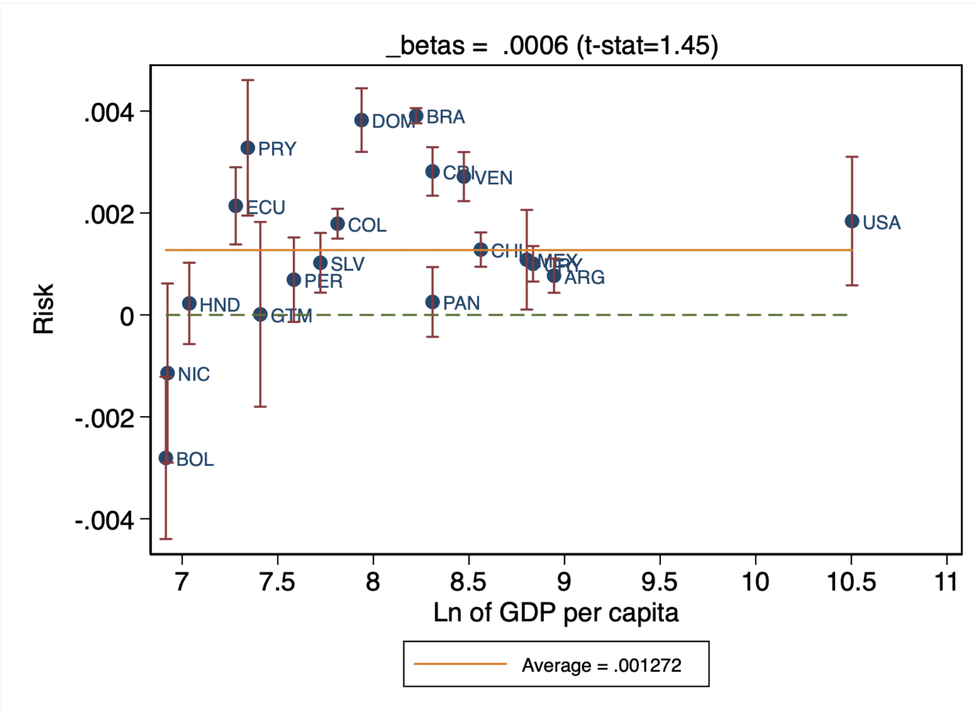
Note: This figure plots the risk of salaried males from an estimation of risk against a dummy for gender for each country: $e^2 =_0 + \alpha_1 age + \alpha_2 women + v$. Data source: I2D2. See Table A1 for descriptive statistics.

Figure 6: Risk, Self-employment



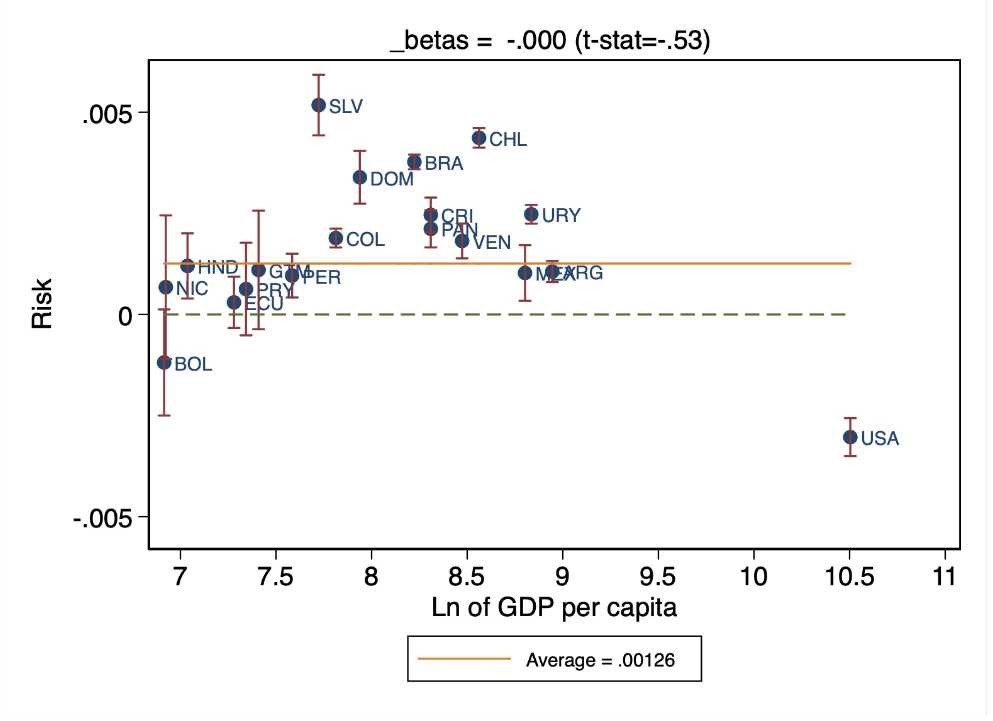
Note: This figure plots the risk of self-employed workers from an estimation of risk against a dummy for self-employment, interacted with age, for each of the countries in our sample: $e^2 =_0 + \alpha_1 age + \alpha_2 self\ employed + \alpha_3 self\ employed * age + v$. Data source: I2D2. See Table A1 for descriptive statistics.

Figure 7: Risk, Women



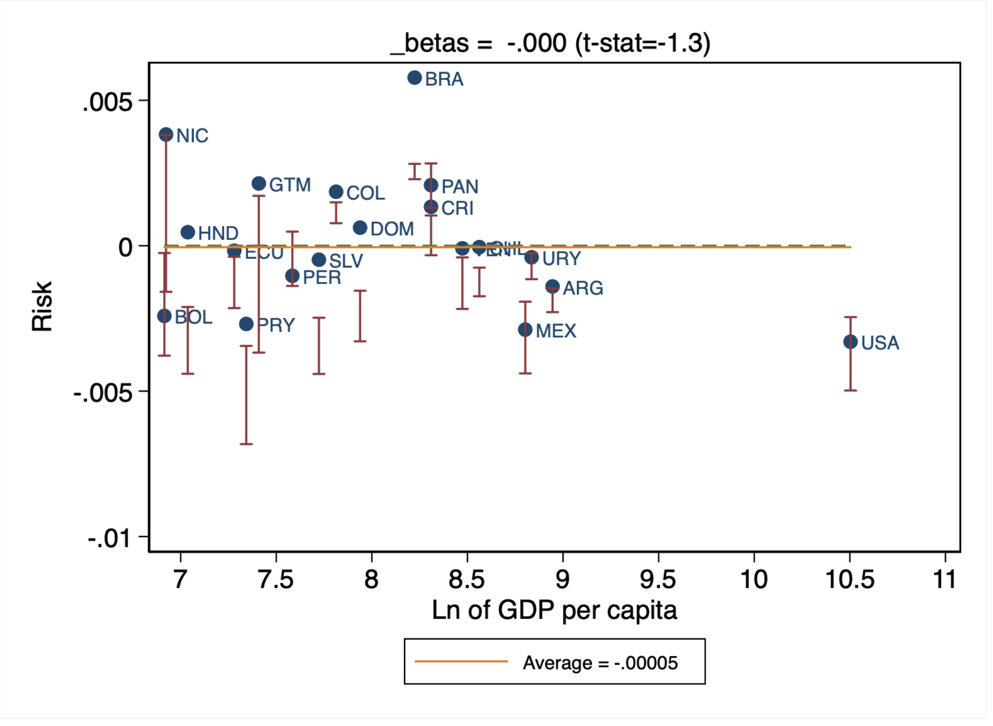
Note: This figure plots the risk of female workers from an estimation of risk against a dummy for gender, interacted with age, for each of the countries in our sample: $e^2 =_0 + \alpha_1 age + \alpha_2 women + \alpha_3 women * age + v$ Data source: I2D2. See Table A1 for descriptive statistics.

Figure 8: Risk, Secondary education



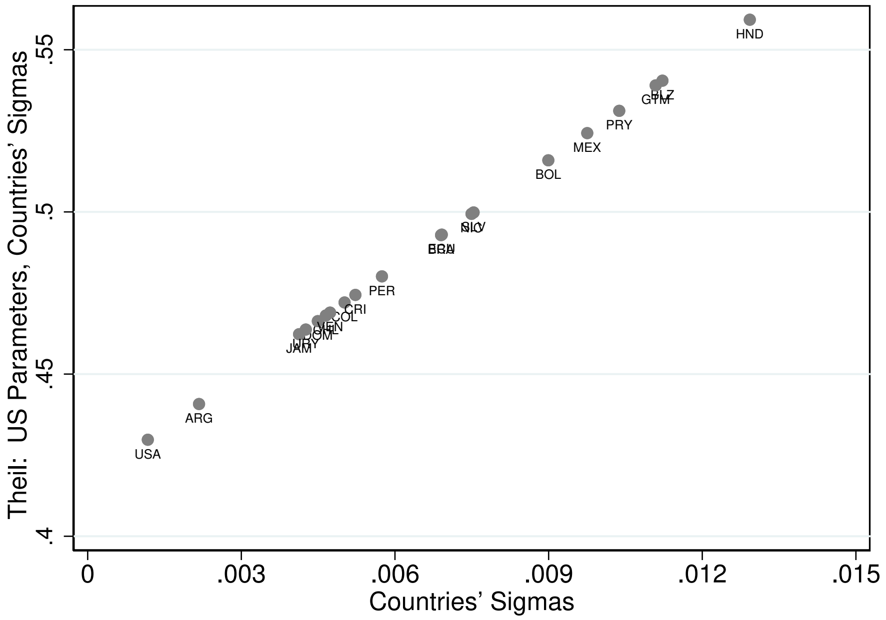
Note: This figure plots the risk of secondary education workers from an estimation of risk against a dummy for gender, interacted with age, for each of the countries in our sample: $e^2 =_0 +\alpha_1 age + \alpha_2 secondary\ education + \alpha_3 secondary\ education * age + v$. Data source: I2D2. See Table A1 for descriptive statistics.

Figure 9: Risk, Post-secondary education



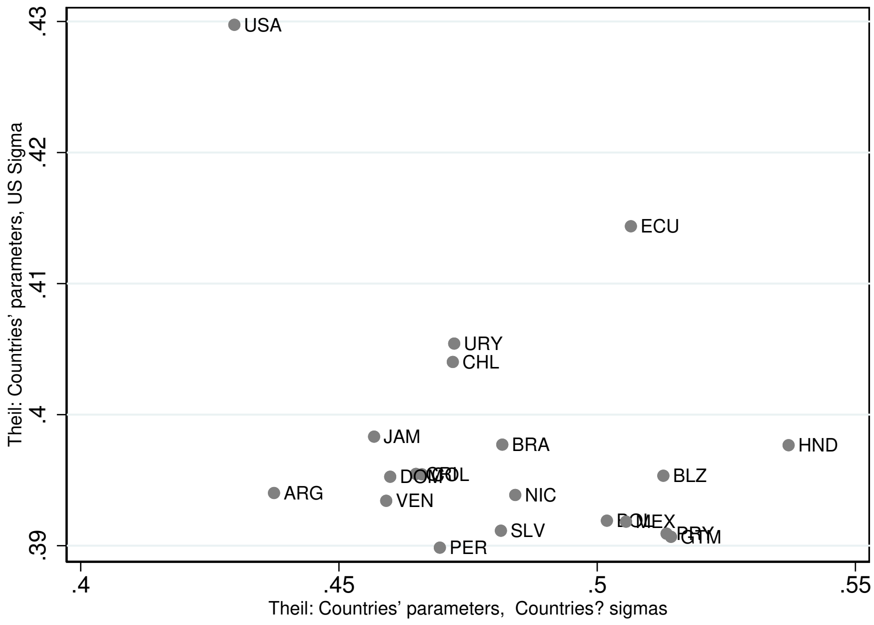
Note: This figure plots the risk of post-secondary education workers from an estimation of risk against a dummy for gender, interacted with age, for each of the countries in our sample: $e^2 =_0 + \alpha_1 age + \alpha_2 post - secondary\ education + \alpha_3 post - secondary\ education * age + v$. Data source: I2D2. See Table A1 for descriptive statistics.

Figure 10: Theil index vs. Sigma



Note: This figure plots a Theil index calculated with US parameters (initial distribution of income, population composition, and income distribution), but it varies the risk component taking the values of Latin American countries. Data source: I2D2. See Table A1 for descriptive statistics.

Figure 11: Theil index



Note: This figure plots a Theil index for each of the countries in the sample against a Theil calculated with country parameters, but with the US component of risk. It shows that once countries get the risk component of the US, inequality measured by the Theil decreases. Data source: I2D2. See Table A1 for descriptive statistics.

Tables

Table 1: Decomposing the risk - Mexico

	(1)	(2)	(3)	(4)	(5)
Age	0.00929***	0.00909***	0.00228***	0.00682***	0.00230***
	(0.000337)	(0.000385)	(0.000365)	(0.000540)	(0.000519)
Self			0.880***		0.832***
			(0.0281)		(0.0416)
Self*age			0.0018		0.00243***
			(0.0001)		(0.0014)
Women		0.145***			0.119***
		(0.0282)			(0.0270)
Women*age		0.00104			0.00136*
		(0.000749)			(0.000718)
No educ				0.131***	0.0472
				(0.0283)	(0.0295)
Post-secondary				0.0856**	0.107***
				(0.0348)	(0.0310)
No educ*age				0.00149**	-0.000520
				(0.000727)	(0.000775)
Post-secondary*age				-0.00381***	-0.00259***
				(0.000921)	(0.000833)
Constant	0.293***	0.234***	0.339***	0.319***	0.217***
	(0.0127)	(0.0144)	(0.0139)	(0.0200)	(0.0188)
Observations	219,593	219,593	219,593	219,593	219,593
R-squared	0.004	0.009	0.084	0.011	0.095

Note: This table shows the decomposition of risk among demographic categories for Mexico. Data source: I2D2. See Table A1 for descriptive statistics.

Table 2: Decomposing the risk - United States

	(1)	(2)	(3)	(4)	(5)
Age	0.0025***	0.003871***	0.002369***	0.002306***	0.001522***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Self			0.153715***		0.160912***
			(0.014)		(0.014)
Self*age			0.002586***		0.002328***
			(0.000)		(0.000)
Women		0.020757***			-0.005055
		(0.005)			(0.005)
Women*age		0.00192			0.001***
		(0.002)			(0.000)
No educ				-0.012426	-0.014962
				(0.023)	(0.023)
Post-secondary				-0.000134	0.005034
				(0.002)	(0.000)
No educ*age				-0.001114**	-0.000847
				(0.001)	(0.001)
Post-secondary*age				-0.00343***	-0.0037***
				(0.002)	(0.000)
Constant	0.245117***	0.235338***	0.262093***	0.247583***	0.258098***
	(0.003)	(0.004)	(0.002)	(0.004)	(0.004)
Observations	4,606,775	4,606,775	4,606,775	4,606,775	4,606,775
R-squared	0.002	0.002	0.016	0.005	0.018

Note: This table shows the decomposition of risk among demographic categories for the United States. Data source: I2D2. See Table A1 for descriptive statistics.

Table 3: Risk Decomposition

	10th Percentile	50th Percentile	90th Percentile
	A: Raw risk		
	0.001	0.046	0.414
	B. Decomposition method: Machado-Mata-Melly		
	0.001***	0.040***	0.396***
	(0.000)	(0.001)	(0.005)
Total explained by characteristics	0.002***	0.016***	0.357***
	(0.000)	(0.000)	(0.002)
Total explained by coefficients	-0.001***	0.024***	0.039***
	(0.000)	(0.000)	(0.004)
	C: Decomposition method: RIF regression		
Difference	0.0011***	0.0414***	0.3711***
	(0.0002)	(0.0003)	(0.0033)
Total explained by characteristics	0.0025***	0.0172***	0.3331***
	(0.0001)	(0.0002)	(0.0024)
Total explained by coefficients	-0.0014***	0.0242***	0.0381***
	(0.0002)	(0.0004)	(0.0039)
C1: Explained by characteristics			
Permanent component	0.0023	0.0154	0.2938
Age	-0.0003***	-0.0018***	-0.0012
	(0.0000)	(0.0001)	(0.0007)
Women*age	0.0001	-0.0006***	-0.0079***
	(0.0000)	(0.0001)	(0.0011)
Self-employed*Age	0.0012***	0.0054***	0.1184***
	(0.0002)	(0.0005)	(0.0058)
Secondary Educ*Age	-0.0000	-0.0003***	-0.0085***
	(0.0000)	(0.0000)	(0.0005)
Post-Secondary Educ*Age	0.0013***	0.0127***	0.1930***
	(0.0004)	(0.0008)	(0.0078)
Transitory component	0.0002	0.0017	0.0292
Women	-0.0003***	-0.0004***	-0.0016**
	(0.0000)	(0.0001)	(0.0008)
Self-employed	0.0032***	0.0221***	0.1922***
	(0.0002)	(0.0005)	(0.0059)
Secondary Educ	-0.0001	0.0003***	0.0072***
	(0.0000)	(0.0001)	(0.0010)
Post-Secondary Educ	-0.0026***	-0.0203***	-0.1586***
	(0.0003)	(0.0008)	(0.0075)
C2: Unexplained by characteristics			
Permanent component	0.0138	0.061	0.4173
Age	0.0141***	0.0951***	1.0658***
	(0.0045)	(0.0093)	(0.1131)
Women*age	-0.0014**	-0.0092***	-0.2222***
	(0.0006)	(0.0012)	(0.0125)
Self-employed*Age	-0.0014***	-0.0066***	-0.0453***
	(0.0001)	(0.0001)	(0.0012)
Secondary Educ*Age	-0.0000	-0.0071**	-0.1705***
	(0.0014)	(0.0029)	(0.0357)
Post-Secondary Educ*Age	0.0025	-0.0061	-0.2105***
	(0.0031)	(0.0063)	(0.0766)
Transitory component	0.0193	0.1182	0.9889
Women	-0.0004	0.0008	0.1400***
	(0.0006)	(0.0012)	(0.0130)
Self-employed	0.0013***	0.0062***	0.0400***
	(0.0001)	(0.0001)	(0.0011)
Secondary Educ	0.0059***	0.0339***	0.3292***
	(0.0015)	(0.0031)	(0.0388)
Post-Secondary Educ	0.0125***	0.0773***	0.4797***
	(0.0032)	(0.0067)	(0.0833)

Note: This table shows the decomposition of risk among demographic categories for the United States

Table A1: Descriptive Statistics: I2D2

	Argentina (1974-2012)					*Belize (1993-1999)*				
Variables	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	446601	38.62	8.675	25	55	9211	36.75	8.116	25	55
Schooling	442992	10.98	3.912	0	20	9140	7.901	3.519	0	22
Working Hours	443647	42.06	18	0	147	9211	44.02	14.88	1	99
Wage	446601	213.2	2812	0.174	181481	9211	5.364	6.062	0.13	100
Women (=1)	446601	0.408	0.491	0	1	9211	0.245	0.43	0	1
Rural (=1)	446601	0	0	0	0	9211	0.559	0.497	0	1
Paid worker (=1)	446601	0.76	0.427	0	1	9211	0.692	0.462	0	1
Employer (=1)	446601	0.0422	0.201	0	1	9211	0.0801	0.271	0	1
Self-employed (=1)	446601	0.198	0.398	0	1	9211	0.227	0.419	0	1
	Bolivia (1992-2012)					*Brazil (1981-2012)*				
Variables	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	79300	38.35	8.532	25	55	2.64E+006	37.56	8.415	25	55
Schooling	78824	9.263	5.256	0	23	2.63E+006	7.051	4.787	0	19
Working Hours	79300	46.98	20.4	1	147	2.64E+006	43.36	13.43	1	99
Wage	79300	8.345	9.964	0.0399	104.9	2.64E+006	447.8	2216	0.144	50000
Women (=1)	75677	0.385	0.487	0	1	2.64E+006	0.381	0.486	0	1
Rural (=1)	79300	0.237	0.425	0	1	2.64E+006	0.143	0.35	0	1
Paid worker (=1)	79300	0.479	0.5	0	1	2.64E+006	0.694	0.461	0	1
Employer (=1)	79300	0.0677	0.251	0	1	2.64E+006	0.0475	0.213	0	1
Self-employed (=1)	79300	0.453	0.498	0	1	2.64E+006	0.259	0.438	0	1
	Chile (1987-2011)					*Costa Rica (1989-2009)*				
	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	504042	38.68	8.553	25	55	190761	37.75	8.288	25	55
Schooling	495761	9.878	4.069	0	22	190267	8	4.249	0	19
Working Hours	489058	45.13	17.27	0	135	190761	46.67	16.64	1	98
Wage	504042	1325	1577	15.62	21824	190761	812.9	1009	6	14583
Women (=1)	504042	0.331	0.471	0	1	190761	0.324	0.468	0	1
Rural (=1)	504042	0.29	0.454	0	1	190761	0.58	0.494	0	1
Paid worker (=1)	504042	0.759	0.428	0	1	190761	0.715	0.451	0	1
Employer (=1)	504042	0.0243	0.154	0	1	190761	0.0726	0.259	0	1
Self-employed (=1)	504042	0.217	0.412	0	1	190761	0.212	0.409	0	1

	Colombia (1996-2012)					*Dominican Republic (1996-2012)*				
	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	1.30E+006	38.53	8.573	25	55	98504	38.06	8.4	25	55
Schooling	1.30E+006	9.455	4.72	0	20	98495	8.58	4.981	0	24
Working Hours	1.17E+006	48.88	17.91	1	150	98504	42.52	13.43	1	99
Wage	1.30E+006	3923	4249	82.98	41808	98504	51.49	51.55	2	581.4
Women (=1)	1.30E+006	0.445	0.497	0	1	98504	0.357	0.479	0	1
Rural (=1)	1.30E+006	0.085	0.279	0	1	98504	0.313	0.464	0	1
Paid worker (=1)	1.30E+006	0.528	0.499	0	1	98504	0.545	0.498	0	1
Employer (=1)	1.30E+006	0.0442	0.206	0	1	98504	0.0392	0.194	0	1
Self-employed (=1)	1.30E+006	0.427	0.495	0	1	98504	0.416	0.493	0	1
	Ecuador (1994-2012)					*Guatemala (2000-2011)*				
	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	230482	39.03	8.679	25	55	52163	37.84	8.695	25	55
Schooling	228532	9.047	4.946	0	23	51959	5.53	4.873	0	20
Working Hours	213807	42.71	16.69	1	140	51138	45.3	19.63	0	140
Wage	230482	506.3	2697	0.0319	105000	52163	1879	2047	5.75	23226
Women (=1)	230482	0.371	0.483	0	1	52163	0.368	0.482	0	1
Rural (=1)	230482	0.374	0.484	0	1	52163	0.38	0.486	0	1
Paid worker (=1)	230482	0.593	0.491	0	1	52163	0.564	0.496	0	1
Employer (=1)	230482	0.0591	0.236	0	1	52163	0.0636	0.244	0	1
Self-employed (=1)	230482	0.348	0.476	0	1	52163	0.373	0.483	0	1
	Honduras (1991-2011)					*Mexico (1989-2012)*				
	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	196902	37.81	8.613	25	55	219593	38.06	8.448	25	55
Schooling	195928	6.598	4.818	0	22	219564	8.663	4.94	0	22
Working Hours	196902	43.45	18.02	1	148	149779	44.6	17.12	1	150
Wage	196902	24.43	33.92	0.192	697.8	219593	491.3	2359	0.108	72014
Women (=1)	196902	0.388	0.487	0	1	219593	0.362	0.481	0	1
Rural (=1)	196902	0.382	0.486	0	1	219593	0.242	0.428	0	1
Paid worker (=1)	196902	0.534	0.499	0	1	219593	0.728	0.445	0	1
Employer (=1)	196902	0.127	0.333	0	1	219593	0.0613	0.24	0	1
Self-employed (=1)	196902	0.34	0.474	0	1	219593	0.21	0.407	0	1

	Jamaica (1990-2002)					*Nicaragua (1993-2009)*				
	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	3969	36.84	8.476	25	55	25857	37.61	8.529	25	55
Schooling	3805	9.909	2.837	0	18	25760	6.436	4.866	0	22
Working Hours	3969	40.13	6.96	8	49	25857	47.77	18.31	1	147
Wage	3969	98.12	106.9	0.913	1433	25857	15.27	18.79	0.176	242
Women (=1)	3953	0.501	0.5	0	1	25857	0.399	0.49	0	1
Rural (=1)	3969	0.517	0.5	0	1	25857	0.335	0.472	0	1
Paid worker (=1)	3969	0.944	0.23	0	1	25857	0.562	0.496	0	1
Employer (=1)	3969	0.00932	0.0961	0	1	25857	0.0426	0.202	0	1
Self-employed (=1)	3969	0.0423	0.201	0	1	25857	0.395	0.489	0	1
	Panama (1989-2012)					*Peru (1996-2012)*				
	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	210926	38.4	8.468	25	55	277802	38.83	8.552	25	55
Schooling	210843	10.09	4.682	0	22	277231	9.245	4.667	0	20
Working Hours	210926	41.66	15.12	1	98	277802	42.72	21.3	1	99
Wage	210926	2.478	2.535	0.0799	34.38	277802	4.314	4.911	0.026	66.74
Women (=1)	210926	0.368	0.482	0	1	277802	0.397	0.489	0	1
Rural (=1)	210926	0.385	0.487	0	1	277802	0.334	0.471	0	1
Paid worker (=1)	210926	0.707	0.455	0	1	277802	0.456	0.498	0	1
Employer (=1)	210926	0.0303	0.171	0	1	277802	0.0733	0.261	0	1
Self-employed (=1)	210926	0.263	0.44	0	1	277802	0.471	0.499	0	1
	Paraguay (1990-2012)					*El Salvador (1991-2009)*				
	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	87141	38.47	8.597	25	55	186996	37.44	8.481	25	55
Schooling	87085	8.084	4.625	0	18	186857	7.116	5.171	0	22
Working Hours	87141	46.37	19.82	1	150	186996	44.04	15.48	1	99
Wage	87141	6925	8919	74.36	154733	186996	8.89	10.67	0.0297	113.1
Women (=1)	87141	0.39	0.488	0	1	186996	0.453	0.498	0	1
Rural (=1)	87141	0.402	0.49	0	1	186996	0.341	0.474	0	1
Paid worker (=1)	87141	0.477	0.499	0	1	186996	0.652	0.476	0	1
Employer (=1)	87141	0.0842	0.278	0	1	186996	0.0475	0.213	0	1
Self-employed (=1)	87141	0.422	0.494	0	1	186996	0.3	0.458	0	1

	Uruguay (1989-2012)					*Venezuela (1989-2006)*				
	N	Mean	sdt. Dev.	min	max	N	Mean	sdt. Dev.	min	max
Age	473920	39.76	8.727	25	55	430964	37.81	8.316	25	55
Schooling	473182	9.964	3.9	0	22	430018	8.567	4.379	0	19
Working Hours	473920	41.62	16.4	1	108	430964	41.45	12.65	1	99
Wage	473920	60.77	63.36	0.0646	700	430964	1564	1976	4.167	25000
Women (=1)	473920	0.445	0.497	0	1	430964	0.386	0.487	0	1
Rural (=1)	473920	0.0635	0.244	0	1	318785	0.86	0.347	0	1
Paid worker (=1)	473920	0.732	0.443	0	1	430964	0.605	0.489	0	1
Employer (=1)	473920	0.0477	0.213	0	1	430964	0.0661	0.248	0	1
Self-employed (=1)	473920	0.22	0.414	0	1	430964	0.329	0.47	0	1

	United States (1990-2010)				
	N	Mean	sdt. Dev.	min	max
Age	4606775	40.05	8.654	25	55
Schooling	4606775	13.81	2.664	0	20
Working Hours	4606775	41.54	11.17	1	99
Wage	4606775	20.47	19.54	0.237	195.3
Women (=1)	4606775	0.469	0.499	0	1
Rural (=1)	4606775	0.279	0.449	0	1
Paid worker (=1)	4606775	0.9	0.3	0	1
Employer (=1)	4606775	0	0	0	0
Self-employed (=1)	4606775	0.1	0.3	0	1

Previous doctoral theses in the Department of Economics, Gothenburg

Avhandlingar publicerade innan serien Ekonomiska Studier startades
(Theses published before the series Ekonomiska Studier was started):

Östman, Hugo (1911), Norrlands ekonomiska utveckling
Moritz, Marcus (1911), Den svenska tobaksindustrien
Sundbom, I. (1933), Prisbildning och ändamålsenlighet
Gerhard, I. (1948), Problem rörande Sveriges utrikeshandel 1936/38
Hegeland, Hugo (1951), The Quantity Theory of Money
Mattsson, Bengt (1970), Cost-Benefit analys
Rosengren, Björn (1975), Valutareglering och nationell ekonomisk politik
Hjalmarsson, Lennart (1975), Studies in a Dynamic Theory of Production and its Applications
Örtendahl, Per-Anders (1975), Substitutionsaspekter på produktionsprocessen vid massaframställning
Anderson, Arne M. (1976), Produktion, kapacitet och kostnader vid ett helautomatiskt emballageglasbruk
Ohlsson, Olle (1976), Substitution och odelbarheter i produktionsprocessen vid massaframställning
Gunnarsson, Jan (1976), Produktionssystem och tätortshierarki – om sambandet mellan rumslig och ekonomisk struktur
Köstner, Evert (1976), Optimal allokering av tid mellan utbildning och arbete
Wigren, Rune (1976), Analys av regionala effektivitetsskillnader inom industribranscher
Wästlund, Jan (1976), Skattning och analys av regionala effektivitetsskillnader inom industribranscher
Flöjstad, Gunnar (1976), Studies in Distortions, Trade and Allocation Problems
Sandelin, Bo (1977), Prisutveckling och kapitalvinster på bostadsfastigheter
Dahlberg, Lars (1977), Empirical Studies in Public Planning
Lönnroth, Johan (1977), Marxism som matematisk ekonomi
Johansson, Börje (1978), Contributions to Sequential Analysis of Oligopolistic Competition

<u>Ekonomiska Studier</u>, utgivna av Nationalekonomiska institutionen vid Göteborgs Universitet. Nr 1 och 4 var inte doktorsavhandlingar. (The contributions to the department series 'Ekonomiska Studier' where no. 1 and 4 were no doctoral theses):

2. **Ambjörn, Erik** (1959), Svenskt importberoende 1926-1956: en ekonomisk-statistisk kartläggning med kommentarer
3. **Landgren, K-G.** (1960), Den "Nya ekonomien" i Sverige: J.M. Keynes, E. Wigfors och utecklingen 1927-39
5. **Bigsten, Arne** (1979), Regional Inequality and Development: A Case Study of Kenya
6. **Andersson, Lars** (1979), Statens styrning av de kommunala budgetarnas struktur (Central Government Influence on the Structure of the Municipal Budget)
7. **Gustafsson, Björn** (1979), Inkomst- och uppväxtförhållanden (Income and Family Background)
8. **Granholm, Arne** (1981), Interregional Planning Models for the Allocation of Private and Public Investments
9. **Lundborg, Per** (1982), Trade Policy and Development: Income Distributional Effects in the Less Developed Countries of the US and EEC Policies for Agricultural

Commodities

10. **Juås, Birgitta** (1982), Värdering av risken för personskador. En jämförande studie av implicita och explicita värden. (Valuation of Personal Injuries. A comparison of Explicit and Implicit Values)
11. **Bergendahl, Per-Anders** (1982), Energi och ekonomi - tillämpningar av input-output analys (Energy and the Economy - Applications of Input-Output Analysis)
12. **Blomström, Magnus** (1983), Foreign Investment, Technical Efficiency and Structural Change - Evidence from the Mexican Manufacturing Industry
13. **Larsson, Lars-Göran** (1983), Comparative Statics on the Basis of Optimization Methods
14. **Persson, Håkan** (1983), Theory and Applications of Multisectoral Growth Models
15. **Sterner, Thomas** (1986), Energy Use in Mexican Industry.
16. **Flood, Lennart** (1986), On the Application of Time Use and Expenditure Allocation Models.
17. **Schuller, Bernd-Joachim** (1986), Ekonomi och kriminalitet - en empirisk undersökning av brottsligheten i Sverige (Economics of crime - an empirical analysis of crime in Sweden)
18. **Walfridson, Bo** (1987), Dynamic Models of Factor Demand. An Application to Swedish Industry.
19. **Stålhammar, Nils-Olov** (1987), Strukturomvandling, företagsbeteende och förväntningsbildning inom den svenska tillverkningsindustrin (Structural Change, Firm Behaviour and Expectation Formation in Swedish Manufactury)
20. **Anxo, Dominique** (1988), Sysselsättningseffekter av en allmän arbetstidsförkortning (Employment effects of a general shortage of the working time)
21. **Mbelle, Ammon** (1988), Foreign Exchange and Industrial Development: A Study of Tanzania.
22. **Ongaro, Wilfred** (1988), Adoption of New Farming Technology: A Case Study of Maize Production in Western Kenya.
23. **Zejan, Mario** (1988), Studies in the Behavior of Swedish Multinationals.
24. **Görling, Anders** (1988), Ekonomisk tillväxt och miljö. Förorenings-struktur och ekonomiska effekter av olika miljövårdsprogram. (Economic Growth and Environment. Pollution Structure and Economic Effects of Some Environmental Programs).
25. **Aguilar, Renato** (1988), Efficiency in Production: Theory and an Application on Kenyan Smallholders.
26. **Kayizzi-Mugerwa, Steve** (1988), External Shocks and Adjustment in Zambia.
27. **Bornmalm-Jardelöw, Gunilla** (1988), Högre utbildning och arbetsmarknad (Higher Education and the Labour Market)
28. **Tansini, Ruben** (1989), Technology Transfer: Dairy Industries in Sweden and Uruguay.
29. **Andersson, Irene** (1989), Familjebeskattning, konsumtion och arbetsutbud - En ekonometrisk analys av löne- och inkomstelasticiteter samt policysimuleringar för svenska hushåll (Family Taxation, Consumption and Labour Supply - An Econometric Analysis of Wage and Income Elasticities and Policy Simulations for Swedish Households)
30. **Henrekson, Magnus** (1990), An Economic Analysis of Swedish Government Expenditure
31. **Sjöö, Boo** (1990), Monetary Policy in a Continuous Time Dynamic Model for Sweden

32. **Rosén, Åsa** (1991), Contributions to the Theory of Labour Contracts.
33. **Loureiro, Joao M. de Matos** (1992), Foreign Exchange Intervention, Sterilization and Credibility in the EMS: An Empirical Study
34. **Irandoust, Manuchehr** (1993), Essays on the Behavior and Performance of the Car Industry
35. **Tasiran, Ali Cevat** (1993), Wage and Income Effects on the Timing and Spacing of Births in Sweden and the United States
36. **Milopoulos, Christos** (1993), Investment Behaviour under Uncertainty: An Econometric Analysis of Swedish Panel Data
37. **Andersson, Per-Åke** (1993), Labour Market Structure in a Controlled Economy: The Case of Zambia
38. **Storrie, Donald W.** (1993), The Anatomy of a Large Swedish Plant Closure
39. **Semboja, Haji Hatibu Haji** (1993), Energy and Development in Kenya
40. **Makonnen, Negatu** (1993), Labor Supply and the Distribution of Economic Well-Being: A Case Study of Lesotho
41. **Julin, Eva** (1993), Structural Change in Rural Kenya
42. **Durevall, Dick** (1993), Essays on Chronic Inflation: The Brazilian Experience
43. **Veiderpass, Ann** (1993), Swedish Retail Electricity Distribution: A Non-Parametric Approach to Efficiency and Productivity Change
44. **Odeck, James** (1993), Measuring Productivity Growth and Efficiency with Data Envelopment Analysis: An Application on the Norwegian Road Sector
45. **Mwenda, Abraham** (1993), Credit Rationing and Investment Behaviour under Market Imperfections: Evidence from Commercial Agriculture in Zambia
46. **Mlambo, Kupukile** (1993), Total Factor Productivity Growth: An Empirical Analysis of Zimbabwe's Manufacturing Sector Based on Factor Demand Modelling
47. **Ndung'u, Njuguna** (1993), Dynamics of the Inflationary Process in Kenya
48. **Modén, Karl-Markus** (1993), Tax Incentives of Corporate Mergers and Foreign Direct Investments
49. **Franzén, Mikael** (1994), Gasoline Demand - A Comparison of Models
50. **Heshmati, Almas** (1994), Estimating Technical Efficiency, Productivity Growth And Selectivity Bias Using Rotating Panel Data: An Application to Swedish Agriculture
51. **Salas, Osvaldo** (1994), Efficiency and Productivity Change: A Micro Data Case Study of the Colombian Cement Industry
52. **Bjurek, Hans** (1994), Essays on Efficiency and Productivity Change with Applications to Public Service Production
53. **Cabezas Vega, Luis** (1994), Factor Substitution, Capacity Utilization and Total Factor Productivity Growth in the Peruvian Manufacturing Industry
54. **Katz, Katarina** (1994), Gender Differentiation and Discrimination. A Study of Soviet Wages
55. **Asal, Maher** (1995), Real Exchange Rate Determination and the Adjustment Process: An Empirical Study in the Cases of Sweden and Egypt
56. **Kjulin, Urban** (1995), Economic Perspectives on Child Care
57. **Andersson, Göran** (1995), Volatility Forecasting and Efficiency of the Swedish Call Options Market
58. **Forteza, Alvaro** (1996), Credibility, Inflation and Incentive Distortions in the Welfare State
59. **Locking, Håkan** (1996), Essays on Swedish Wage Formation
60. **Välilä, Timo** (1996), Essays on the Credibility of Central Bank Independence

61. **Yilma, Mulugeta** (1996), Measuring Smallholder Efficiency: Ugandan Coffee and Food-Crop Production
62. **Mabugu, Ramos E.** (1996), Tax Policy Analysis in Zimbabwe Applying General Equilibrium Models
63. **Johansson, Olof** (1996), Welfare, Externalities, and Taxation; Theory and Some Road Transport Applications.
64. **Chitiga, Margaret** (1996), Computable General Equilibrium Analysis of Income Distribution Policies in Zimbabwe
65. **Leander, Per** (1996), Foreign Exchange Market Behavior Expectations and Chaos
66. **Hansen, Jörgen** (1997), Essays on Earnings and Labor Supply
67. **Cotfas, Mihai** (1997), Essays on Productivity and Efficiency in the Romanian Cement Industry
68. **Horgby, Per-Johan** (1997), Essays on Sharing, Management and Evaluation of Health Risks
69. **Nafar, Nosratollah** (1997), Efficiency and Productivity in Iranian Manufacturing Industries
70. **Zheng, Jinghai** (1997), Essays on Industrial Structure, Technical Change, Employment Adjustment, and Technical Efficiency
71. **Isaksson, Anders** (1997), Essays on Financial Liberalisation in Developing Countries: Capital mobility, price stability, and savings
72. **Gerdin, Anders** (1997), On Productivity and Growth in Kenya, 1964-94
73. **Sharifi, Alimorad** (1998), The Electricity Supply Industry in Iran: Organization, performance and future development
74. **Zamanian, Max** (1997), Methods for Mutual Fund Portfolio Evaluation: An application to the Swedish market
75. **Manda, Damiano Kulundu** (1997), Labour Supply, Returns to Education, and the Effect of Firm Size on Wages: The case of Kenya
76. **Holmén, Martin** (1998), Essays on Corporate Acquisitions and Stock Market Introductions
77. **Pan, Kelvin** (1998), Essays on Enforcement in Money and Banking
78. **Rogat, Jorge** (1998), The Value of Improved Air Quality in Santiago de Chile
79. **Peterson, Stefan** (1998), Essays on Large Shareholders and Corporate Control
80. **Belhaj, Mohammed** (1998), Energy, Transportation and Urban Environment in Africa: The Case of Rabat-Salé, Morocco
81. **Mekonnen, Alemu** (1998), Rural Energy and Afforestation: Case Studies from Ethiopia
82. **Johansson, Anders** (1998), Empirical Essays on Financial and Real Investment Behavior
83. **Köhlin, Gunnar** (1998), The Value of Social Forestry in Orissa, India
84. **Levin, Jörgen** (1998), Structural Adjustment and Poverty: The Case of Kenya
85. **Ncube, Mkhululi** (1998), Analysis of Employment Behaviour in Zimbabwe
86. **Mwansa, Ladslous** (1998), Determinants of Inflation in Zambia
87. **Agnarsson, Sveinn** (1998), Of Men and Machines: Essays in Applied Labour and Production Economics
88. **Kadenge, Phineas** (1998), Essays on Macroeconomic Adjustment in Zimbabwe: Inflation, Money Demand, and the Real Exchange Rate
89. **Nyman, Håkan** (1998), An Economic Analysis of Lone Motherhood in Sweden
90. **Carlsson, Fredrik** (1999), Essays on Externalities and Transport
91. **Johansson, Mats** (1999), Empirical Studies of Income Distribution

92. **Alemu, Tekie** (1999), Land Tenure and Soil Conservation: Evidence from Ethiopia
93. **Lundvall, Karl** (1999), Essays on Manufacturing Production in a Developing Economy: Kenya 1992-94
94. **Zhang, Jianhua** (1999), Essays on Emerging Market Finance
95. **Mlima, Aziz Ponary** (1999), Four Essays on Efficiency and Productivity in Swedish Banking
96. **Davidsen, Björn-Ivar** (2000), Bidrag til den økonomisk-metodologiske tenkningen (Contributions to the Economic Methodological Thinking)
97. **Ericson, Peter** (2000), Essays on Labor Supply
98. **Söderbom, Måns** (2000), Investment in African Manufacturing: A Microeconomic Analysis
99. **Höglund, Lena** (2000), Essays on Environmental Regulation with Applications to Sweden
100. **Olsson, Ola** (2000), Perspectives on Knowledge and Growth
101. **Meuller, Lars** (2000), Essays on Money and Credit
102. **Österberg, Torun** (2000), Economic Perspectives on Immigrants and Intergenerational Transmissions
103. **Kalinda Mkenda, Beatrice** (2001), Essays on Purchasing Power Parity, RealExchange Rate, and Optimum Currency Areas
104. **Nerhagen, Lena** (2001), Travel Demand and Value of Time - Towards an Understanding of Individuals Choice Behavior
105. **Mkenda, Adolf** (2001), Fishery Resources and Welfare in Rural Zanzibar
106. **Eggert, Håkan** (2001), Essays on Fisheries Economics
107. **Andrén, Daniela** (2001), Work, Sickness, Earnings, and Early Exits from the Labor Market. An Empirical Analysis Using Swedish Longitudinal Data
108. **Nivorozhkin, Eugene** (2001), Essays on Capital Structure
109. **Hammar, Henrik** (2001), Essays on Policy Instruments: Applications to Smoking and the Environment
110. **Nannyonjo, Justine** (2002), Financial Sector Reforms in Uganda (1990-2000): Interest Rate Spreads, Market Structure, Bank Performance and Monetary Policy
111. **Wu, Hong** (2002), Essays on Insurance Economics
112. **Linde-Rahr, Martin** (2002), Household Economics of Agriculture and Forestry in Rural Vienam
113. **Maneschiöld, Per-Ola** (2002), Essays on Exchange Rates and Central Bank Credibility
114. **Andrén, Thomas** (2002), Essays on Training, Welfare and Labor Supply
115. **Granér, Mats** (2002), Essays on Trade and Productivity: Case Studies of Manufacturing in Chile and Kenya
116. **Jaldell, Henrik** (2002), Essays on the Performance of Fire and Rescue Services
117. **Alpizar, Francisco, R.** (2002), Essays on Environmental Policy-Making in Developing Countries: Applications to Costa Rica
118. **Wahlberg, Roger** (2002), Essays on Discrimination, Welfare and Labor Supply
119. **Piculescu, Violeta** (2002), Studies on the Post-Communist Transition
120. **Pylkkänen, Elina** (2003), Studies on Household Labor Supply and Home Production
121. **Löfgren, Åsa** (2003), Environmental Taxation – Empirical and Theoretical Applications
122. **Ivaschenko, Oleksiy** (2003), Essays on Poverty, Income Inequality and Health in Transition Economies

123. **Lundström, Susanna** (2003), On Institutions, Economic Growth and the Environment
124. **Wambugu, Anthony** (2003), Essays on Earnings and Human Capital in Kenya
125. **Adler, Johan** (2003), Aspects of Macroeconomic Saving
126. **Erlandsson, Mattias** (2003), On Monetary Integration and Macroeconomic Policy
127. **Brink, Anna** (2003), On the Political Economy of Municipality Break-Ups
128. **Ljungwall, Christer** (2003), Essays on China's Economic Performance During the Reform Period
129. **Chifamba, Ronald** (2003), Analysis of Mining Investments in Zimbabwe
130. **Muchapondwa, Edwin** (2003), The Economics of Community-Based Wildlife Conservation in Zimbabwe
131. **Hammes, Klaus** (2003), Essays on Capital Structure and Trade Financing
132. **Abou-Ali, Hala** (2003), Water and Health in Egypt: An Empirical Analysis
133. **Simatele, Munacinga** (2004), Financial Sector Reforms and Monetary Policy in Zambia
134. **Tezic, Kerem** (2004), Essays on Immigrants' Economic Integration
135. INSTÄLLD
136. **Gjirja, Matilda** (2004), Efficiency and Productivity in Swedish Banking
137. **Andersson, Jessica** (2004), Welfare Environment and Tourism in Developing Countries
138. **Chen, Yinghong** (2004), Essays on Voting Power, Corporate Governance and Capital Structure
139. **Yesuf, Mahmud** (2004), Risk, Time and Land Management under Market Imperfections: Applications to Ethiopia
140. **Kateregga, Eseza** (2005), Essays on the Infestation of Lake Victoria by the Water Hyacinth
141. **Edvardsen, Dag Fjeld** (2004), Four Essays on the Measurement of Productive Efficiency
142. **Lidén, Erik** (2005), Essays on Information and Conflicts of Interest in Stock Recommendations
143. **Dieden, Sten** (2005), Income Generation in the African and Coloured Population – Three Essays on the Origins of Household Incomes in South Africa
144. **Eliasson, Marcus** (2005), Individual and Family Consequences of Involuntary Job Loss
145. **Mahmud, Minhaj** (2005), Measuring Trust and the Value of Statistical Lives: Evidence from Bangladesh
146. **Lokina, Razack Bakari** (2005), Efficiency, Risk and Regulation Compliance: Applications to Lake Victoria Fisheries in Tanzania
147. **Jussila Hammes, Johanna** (2005), Essays on the Political Economy of Land Use Change
148. **Nyangena, Wilfred** (2006), Essays on Soil Conservation, Social Capital and Technology Adoption
149. **Nivorozhkin, Anton** (2006), Essays on Unemployment Duration and Programme Evaluation
150. **Sandén, Klas** (2006), Essays on the Skill Premium
151. **Deng, Daniel** (2006), Three Essays on Electricity Spot and Financial Derivative Prices at the Nordic Power Exchange
152. **Gebreeyesus, Mulu** (2006), Essays on Firm Turnover, Growth, and Investment Behavior in Ethiopian Manufacturing

153. **Islam, Nizamul Md.** (2006), Essays on Labor Supply and Poverty: A Microeconometric Application
154. **Kjaer, Mats** (2006), Pricing of Some Path-Dependent Options on Equities and Commodities
155. **Shimeles, Abebe** (2006), Essays on Poverty, Risk and Consumption Dynamics in Ethiopia
156. **Larsson, Jan** (2006), Four Essays on Technology, Productivity and Environment
157. **Congdon Fors, Heather** (2006), Essays in Institutional and Development Economics
158. **Akpalu, Wisdom** (2006), Essays on Economics of Natural Resource Management and Experiments
159. **Daruvala, Dinky** (2006), Experimental Studies on Risk, Inequality and Relative Standing
160. **García, Jorge** (2007), Essays on Asymmetric Information and Environmental Regulation through Disclosure
161. **Bezabih, Mintewab** (2007), Essays on Land Lease Markets, Productivity, Biodiversity, and Environmental Variability
162. **Visser, Martine** (2007), Fairness, Reciprocity and Inequality: Experimental Evidence from South Africa
163. **Holm, Louise** (2007), A Non-Stationary Perspective on the European and Swedish Business Cycle
164. **Herbertsson, Alexander** (2007), Pricing Portfolio Credit Derivatives
165. **Johansson, Anders C.** (2007), Essays in Empirical Finance: Volatility, Interdependencies, and Risk in Emerging Markets
166. **Ibáñez Díaz, Marcela** (2007), Social Dilemmas: The Role of Incentives, Norms and Institutions
167. **Ekbom, Anders** (2007), Economic Analysis of Soil Capital, Land Use and Agricultural Production in Kenya
168. **Sjöberg, Pål** (2007), Essays on Performance and Growth in Swedish Banking
169. **Palma Aguirre, Grisha Alexis** (2008), Explaining Earnings and Income Inequality in Chile
170. **Akay, Alpaslan** (2008), Essays on Microeconometrics and Immigrant Assimilation
171. **Carlsson, Evert** (2008), After Work – Investing for Retirement
172. **Munshi, Farzana** (2008), Essays on Globalization and Occupational Wages
173. **Tsakas, Elias** (2008), Essays on Epistemology and Evolutionary Game Theory
174. **Erlandzon, Karl** (2008), Retirement Planning: Portfolio Choice for Long-Term Investors
175. **Lampi, Elina** (2008), Individual Preferences, Choices, and Risk Perceptions – Survey Based Evidence
176. **Mitrut, Andreea** (2008), Four Essays on Interhousehold Transfers and Institutions in Post-Communist Romania
177. **Hansson, Gustav** (2008), Essays on Social Distance, Institutions, and Economic Growth
178. **Zikhali, Precious** (2008), Land Reform, Trust and Natural Resource Management in Africa
179. **Tengstam, Sven** (2008), Essays on Smallholder Diversification, Industry Location, Debt Relief, and Disability and Utility
180. **Boman, Anders** (2009), Geographic Labour Mobility – Causes and Consequences
181. **Qin, Ping** (2009), Risk, Relative Standing and Property Rights: Rural Household Decision-Making in China

182. **Wei, Jiegen** (2009), Essays in Climate Change and Forest Management
183. **Belu, Constantin** (2009), Essays on Efficiency Measurement and Corporate Social Responsibility
184. **Ahlerup, Pelle** (2009), Essays on Conflict, Institutions, and Ethnic Diversity
185. **Quiroga, Miguel** (2009), Microeconomic Policy for Development: Essays on Trade and Environment, Poverty and Education
186. **Zerfu, Daniel** (2010), Essays on Institutions and Economic Outcomes
187. **Wollbrant, Conny** (2010), Self-Control and Altruism
188. **Villegas Palacio, Clara** (2010), Formal and Informal Regulations: Enforcement and Compliance
189. **Maican, Florin** (2010), Essays in Industry Dynamics on Imperfectly Competitive Markets
190. **Jakobsson, Niklas** (2010), Laws, Attitudes and Public Policy
191. **Manescu, Cristiana** (2010), Economic Implications of Corporate Social Responsibility and Responsible Investments
192. **He, Haoran** (2010), Environmental and Behavioral Economics – Applications to China
193. **Andersson, Fredrik W.** (2011), Essays on Social Comparison
194. **Isaksson, Ann-Sofie** (2011), Essays on Institutions, Inequality and Development
195. **Pham, Khanh Nam** (2011), Prosocial Behavior, Social Interaction and Development: Experimental Evidence from Vietnam
196. **Lindskog, Annika** (2011), Essays on Economic Behaviour: HIV/AIDS, Schooling, and Inequality
197. **Kotsadam, Andreas** (2011), Gender, Work, and Attitudes
198. **Alem, Yonas** (2011), Essays on Shocks, Welfare, and Poverty Dynamics: Microeconometric Evidence from Ethiopia
199. **Köksal-Ayhan, Miyase Yesim** (2011), Parallel Trade, Reference Pricing and Competition in the Pharmaceutical Market: Theory and Evidence
200. **Vondolia, Godwin Kofi** (2011), Essays on Natural Resource Economics
201. **Widerberg, Anna** (2011), Essays on Energy and Climate Policy – Green Certificates, Emissions Trading and Electricity Prices
202. **Siba, Eyerusalem** (2011), Essays on Industrial Development and Political Economy of Africa
203. **Orth, Matilda** (2012), Entry, Competition and Productivity in Retail
204. **Nerman, Måns** (2012), Essays on Development: Household Income, Education, and Female Participation and Representation
205. **Wicks, Rick** (2012), The Place of Conventional Economics in a World with Communities and Social Goods
206. **Sato, Yoshihiro** (2012), Dynamic Investment Models, Employment Generation and Productivity – Evidence from Swedish Data
207. **Valsecchi, Michele** (2012), Essays in Political Economy of Development
208. **Teklewold Belayneh, Hailemariam** (2012), Essays on the Economics of Sustainable Agricultural Technologies in Ethiopia
209. **Wagura Ndiritu, Simon** (2013), Essays on Gender Issues, Food Security, and Technology Adoption in East Africa
210. **Ruist, Joakim** (2013), Immigration, Work, and Welfare
211. **Nordén, Anna** (2013), Essays on Behavioral Economics and Policies for Provision of Ecosystem Services
212. **Yang, Xiaojun** (2013), Household Decision Making, Time Preferences, and

Positional Concern: Experimental Evidence from Rural China

213. **Bonilla Londoño, Jorge Alexander** (2013), Essays on the Economics of Air Quality Control
214. **Mohlin, Kristina** (2013), Essays on Environmental Taxation and Climate Policy
215. **Medhin, Haileselassie** (2013), The Poor and Their Neighbors: Essays on Behavioral and Experimental Economics
216. **Andersson, Lisa** (2013), Essays on Development and Experimental Economics: Migration, Discrimination and Positional Concerns
217. **Weng, Qian** (2014), Essays on Team Cooperation and Firm Performance
218. **Zhang, Xiao-Bing** (2015), Cooperation and Paradoxes in Climate Economics
219. **Jaime Torres, Monica Marcela** (2015) Essays on Behavioral Economics and Policy Design
220. **Bejenariu, Simona** (2015) Determinants of Health Capital at Birth: Evidence from Policy Interventions
221. **Nguyen, Van Diem** (2015) Essays on Takeovers and Executive Compensation
222. **Tolonen, Anja** (2015) Mining Booms in Africa and Local Welfare Effects: Labor Markets, Women's Empowerment and Criminality
223. **Hassen, Sied** (2015) On the Adoption and Dis-adoption of Household Energy and Farm Technologies
224. **Moursli, Mohamed-Reda** (2015) Corporate Governance and the Design of Board of Directors
225. **Borcan, Oana** (2015) Economic Determinants and Consequences of Political Institutions
226. **Ruhinduka, Remidius Denis** (2015) Essays on Field Experiments and Impact Evaluation
227. **Persson, Emil** (2016) Essays on Behavioral and Experimental Economics, Cooperation, Emotions and Health.
228. **Martinangeli, Andrea** (2017) Bitter divisions: inequality, identity and cooperation
229. **Björk, Lisa** (2017) Essays on Behavioral Economics and Fisheries: Coordination and Cooperation
230. **Chegere, Martin Julius** (2017) Post-Harvest Losses, Intimate Partner Violence and Food Security in Tanzania
231. **Villalobos-Fiatt, Laura** (2017) Essays on forest conservation policies, weather and school attendance
232. **Yi, Yuanyuan** (2017) Incentives and Forest Reform: Evidence from China
233. **Kurz, Verena** (2017) Essays on behavioral economics: Nudges, food consumption and procedural fairness
234. **Yashodha, Yashodha** (2017) Contract Choice and Trust in Informal Groundwater Markets
235. **Meles, Tensay Hadus** (2017) Power Outages, Increasing Block Tariffs and Billing Knowledge
236. **Mühlrad, Hanna** (2018) The Impact of Reproductive and Birth Technologies on Health, Fertility and Labor Outcomes
237. **Gakii Gatua, Josephine** (2018) Primary Health Care Interventions and Social Ties in Kenya
238. **Felgendreher, Simon** (2018) Essays on behavioral economics and the effects of the colonial rule on Java
239. **Demeke, Eyoual** (2019) Essays on Environmental and Behavioral Economics
240. **Kiss, Tamas** (2019) Predictability in Equity Markets: Estimation and Inference
241. **Khomenko, Maksym** (2019) Essays on the Design of Public Policies and .

Regulations

242. **Munkanjari, Samson** (2019) Climate Policy and Financial Markets
243. **Schürz, Simon** (2020) Economic and Intergenerational Decision-Making in Families
244. **Lau, Debbie** (2020) Empirical Essays on Education and Health Policy Evaluation
245. **Ho, Hoang-Anh** (2020) Essays on Culture, Institutions, and Economic Development
246. **Rubio Ramos, Melissa** (2020) The Economics of Coercive Institutions, Conflict, and Development