UNIVERSITÄT HOHENHEIM
Fakultät Wirtschafts- und Sozialwissenschaften

Three Essays in Empirical Economics

Dissertation
zur Erlangung des Grades eines
Doktors der Wirtschaftswissenschaften
(Dr. oec.)

vorgelegt der
Fakultät der Wirtschafts– und Sozialwissenschaften
der Universität Hohenheim

von
Sebastian Daniel Seiffert, M.Sc.

Stuttgart–Hohenheim
2019

Die vorliegende Arbeit wurde im März 2019 von der Fakultät Wirtschafts– und Sozialwissenschaften der Universität Hohenheim als Dissertation zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaften (Dr. oec.) angenommen.

Datum der mündlichen Doktorprüfung: 17. Oktober 2019

Dekan: Prof. Dr. Karsten Hadwich
Prüfungsvorsitz: Prof. Dr. Alfonso Sousa-Poza
Erstgutachter: Prof. Dr. Thomas Beißinger
Zweitgutachter: Prof. Dr. Klaus Prettner

Acknowledgements

In the course of the last couple of years I went through highs and lows. While the work on this thesis has been highly challenging and sometimes also frustrating, it also enabled me to acquire new skills, broaden my scientific horizon and led to many cherishable insights both professionally as well as personally. This contributed a great deal in making me the person who I am today.

First and foremost, I want to express my deep gratitude towards my supervisor and mentor Prof. Dr. Thomas Beißinger. Not only is he a great supervisor and person to work with, he also was always very encouraging and supportive throughout the whole duration of this dissertation project, both on a professional as and personal level.

I further want to thank my second supervisor Prof. Dr. Klaus Prettner. The joint work on the third chapter was always a pleasure and his input and support truly helped me developing the chapter into what it has become.

I am especially grateful that I could write chapter 2 together with Bohdan Kukharskyy. I feel deep respect for Bo for him being such a brilliant scholar and at the same time a very polite and humble person.

Further, I want to thank Lars Henning Mühlen for sharing an office and an apartment through many years of my thesis. Also, I want to thank him for many great talks as well as many great discussions and also nights out.

I feel grateful for all my friends and colleagues who I have met at the University of Hohenheim and elsewhere during the last years. Especially Martyna, Fabian, Jonas, Dario, Marina, Philip and Timo, thank you for everything.

My final thanks are directed at my family and my partner. Without you, I would not have been able to tackle this challenge and many others. You guys rock.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the preface of one of the most established textbooks "Economics" by Samuelson & Nordhaus (1998), the authors state that the ultimate goal of economics is "to improve the living conditions of people in their everyday life". They further emphasise that economics has increased its scope greatly over the past half-century. The flag of economics flies over its traditional territory of the marketplace, but it also covers the environment, legal studies, statistical and historical methods, gender and racial discrimination, and even family life." In this spirit, this dissertation thesis addresses three heterogenous topics and aims to contribute to this noble goal by utilising recent data which has not commonly been used in economic research yet. While being heterogenous in nature, all included chapters address core determinants directly or indirectly influencing human well-being: gun-related violence and its determinants, the impact of the size of the middle class on educational outcomes as well as the effect of lower costs in gaining access to remote markets on subsequent economic activity and its spatial distribution. In order to derive better answers to those questions, it is of utmost importance, besides the development of a deeper understanding based on theoretical inquiry, to address the underlying issues empirically. Therefore, the subsequent chapters of this dissertation, while also proposing stylised models in order to motivate the empirical assessment, are focused on the analysis of recently available as well as newly constructed data sets. The remainder of this chapter will give a brief overview of the motivation for the included works based on the above defined welfare perspective.

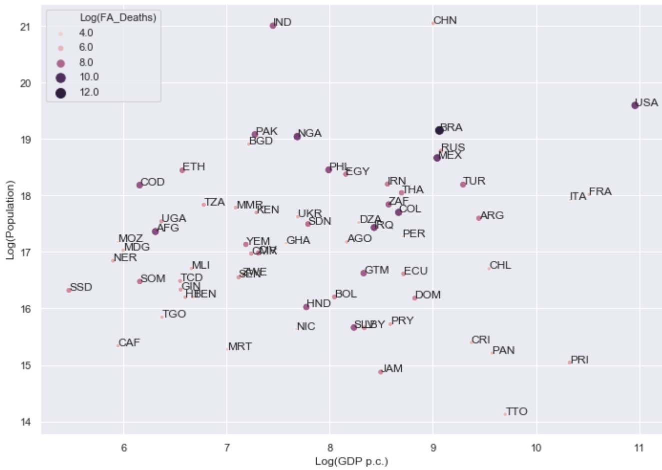


Figure 1.1. *Global Burden of Armed Violence (2016)*[1].

One of the most precious assets of every human being is their physical integrity and health. Being explicitly included in a vast range of national constitutions, the right to physical integrity is also guaranteed by the "Universal Declaration of Human Rights" (1948) issued by the General Assembly of the United Nations. Nevertheless, this elementary human right is constantly being neglected by private individuals as well as representatives of state authorities. In many cases this is being done under utilisation of small arms like handguns or rifles. According to Richmond, Cheney & Schwab (2005), the burden of global non-conflict related firearm deaths are estimated to range from 196,000 to 229,000, adjusted to the year 2000. This is about one fifth of the number of annual car accident fatalities. The number of non-conflict firearm-related injuries and disabilities is suspected

[1] The data for the figure was published, among others, in Mc Evoy & Hideg (2017). It has been obtained from http://www.smallarmssurvey.org/fileadmin/docs/M-files/Armed_violence/Small-Arms-Survey-DB-violent-deaths.xlsx on 04.03.2019.

to be substantially higher. Figure 1.1 depicts per country non-conflict related firearm casualties for the year 2016 in relation to the respective population and GDP per capita (PPP in current $) figures[2]. The larger and darker a bubble is, the higher is the associated firearm death count. The United States of America, with a death count of 10.147, are ranking third in the international comparison in 2016. Only Brazil and Mexico show higher numbers. By relating this number to population size and per capita income, it is evident that the burden of firearm-related violence is especially problematic in the US. As the figure further illustrates, countries which exhibit similar casualty rates are usually also characterised by lower per capita incomes. It appears that in the US case, there must be some mechanisms at play which are not related to economic development and the associated higher rates of social conflict. In the first two months of 2019 alone, the non profit corporation "Gun Violence Archive" recorded 2,342 deaths and 4,019 injuries[3]. In monetary terms, Gani, Sakran & Canner (2017) estimate the annual financial burden of firearm-injury related healthcare expenditure faced by the American citizens to amount to 2.8 billion US dollars for emergency department visits alone regarding the period from 2004 - 2014. While the estimated monetary costs for urgent care alone already amount to a sizeable figure, the impact of social well-being by the harm done is far higher. The purpose of Chapter 2 (joined work with Bohdan Kukharskyy) of this thesis contributes to a better understanding of the issue at hand and especially the US case in two ways. First to improve the structure of the analysis and to provide a guiding framework for the empirical part, a stylised theoretical model of gun-related crime is proposed. The model relates to the economics of crime literature spearheaded by Becker (1968). Modelling crime as a function of economic costs and benefits, Becker (1968) has received substantial criticism. Critics argue in particular that crime can only be understood by accounting for social factors as well[4]. This criticism is supported empirically by Glaeser, Sacerdote &

[2] The underlying set of countries has been trimmed down for ease of presentability. It contains all countries with more than 200 non-conflict related firearm deaths in 2016. Further countries with populations less than 1.2 million and GDP per capita lower than 250 USD have been excluded. All numbers are in their logarithmic form. Population and GDP per capita data have been obtained from https://data.worldbank.org on 04.03.2019.

[3] These figures have been accessed on 04.03.2019 on https://www.gunviolencearchive.org/.

[4] See among others: Hirschi (1969, 1986).

Scheinkman (1996) and Glaeser & Sacerdote (1999). Among others, those studies stress that differences in the levels of criminal behaviour are also rooted in differences in social norms and civic interactions. By including social variables into the model, we address this issue. From our model, we derive the following core hypotheses: Gun-related offenses are increasing in the number of illegal guns. Firearm offenses are decreasing in the level of social capital. Thirdly, gun-related offenses are decreasing in police intensity. In the second part of the chapter, we test our core hypotheses by confronting the model with the data. This is achieved using recently available county-level data in order to construct a novel panel data set. This data set contains detailed information on the number of (gun-related) offenses, police intensity, proxies for the availability of illegal guns, and a vast array of socioeconomic variables. The underlying data covers about 90% of all US counties. Accordingly, we are able to draw a representative picture of the US Further, we introduce a new proxy for the prevalence of illegal guns by exploiting annual information of guns reported as stolen. Building on the works of Putnam, Leonardi & Nanetti (1994), Putnam (1995, 2000), we construct a proxy for social capital based on the prevalence of religious, social and civic organisations in a given county. The empirical analyses of chapter 2 relate to the contributions of Duggan, Hjalmarsson & Jacob (2011) and Cook & Ludwig (2006) which both find a positive relationship between gun prevalence and (gun) homicide rates. Chapter 2 contributes to the empirical literature in three ways: It introduces a novel proxy for gun prevalence based on gun thefts. The empirical analyses are extended to an almost nation-wide set of counties. And thirdly, exploiting time variation in illegal guns we approach a causal inference with respect to the effect of illegal guns on gun-related violence.

Another key determinant of individual well-being is education. Article 26 of the "Universal Declaration of Human Rights" (1948) states that "Everyone has the right to education. Education shall be free, at least in the elementary and fundamental stages." The article further emphasises that "Education shall be compulsory". This claim has been solidified by UN General Assembly (1966). The universal right to education has been reaffirmed by both the UN General Assembly (1981) and the UN General Assembly (1990). While the former explicitly calls for the access to universal education for women, the latter does so for children. The previous examples illustrate the importance of education from a human rights perspective which has the ultimate goal to ensure well-being. From an economic viewpoint there exists a multitude of channels via which the effects of education on well-being are mitigated. For example, education increases productivity and with it individual as well as aggregate income (Lucas, 1988) and it enhances ingenuity (Romer, 1990; Strulik, Prettner, & Prskawetz, 2013). Besides that, it further has various positive effects on social outcomes. Barro (1999) and Glaeser, La Porta, Lopez-de-Silanes & Shleifer (2004) have found it to benefit the quality of institutions and democratic processes. Further, higher educational outcomes help avoiding social conflict[5].

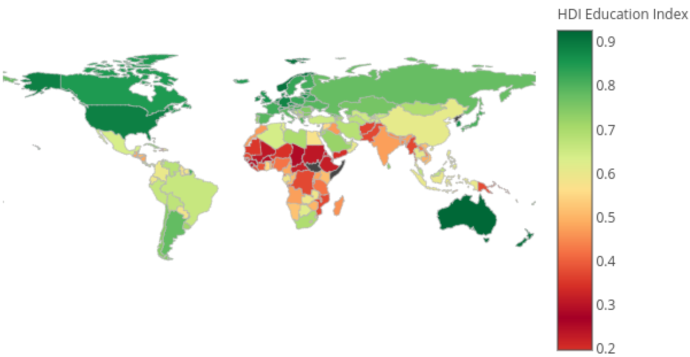


Figure 1.2. *Human Development Report Education Index, 2013. Source: Human Development Report*[6].

[5] See Ostby & Urdal (2011) for a review of the literature.
[6] The data for the underlying figure has been obtained from http://hdr.undp.org/en/content/education-index.

Alesina & Perotti (1996) find that higher levels of educational attainment are associated with significantly higher socio-political stability. This nexus is especially pronounced in the context of developing countries.

While there is obvious consent on the importance of education from both, the human rights as well as the economic perspective, the global community - while having made noticeable progress over the last decades - is still far from guaranteeing adequate access to it. Figure 1.2 gives an overview on the global situation with respect to educational attainment. It illustrates the per country score in the UN Education Index[7] which is one of the three components of the Human Development Index. Individual values lie within the closed unit interval. As we can see, a considerable fraction of nations receive scores below 0.75. The global mean - being 0.646 - reflects the fact that large parts of the global population still do not have appropriate access to education.

One of the factors which have been found helpful in explaining differential educational outcomes in previous research is a society's share of middle class households. Chapter 3 (joint work with Klaus Prettner) addresses the question whether a larger share of middle class households increases educational outcomes in India. In order to give a structured picture of the argument and to guide the empirical analysis, we propose a stylised model of the demand for education in a setting where the population is divided into three income groups. From the model we derive the hypothesis that larger shares of middle class households contribute to higher levels of average educational attainment. In order to test our empirically our hypothesis, we utilise household, individual and village level data from the Indian Household Survey 2005.

The empirical part of the chapter is based on Indian data for various reasons. While being one of the world's most populous nations, India is at the same time one of the most heterogenous nation states in the modern world (Vannemann & Dubey, 2013). The main reason is that (modern) the social hierarchy of the Indian society is based on a unique caste system. This caste system has a lasting impact on social life and has persisted over thousands of years. It gives the Indian society a clear and hierarchical structure

[7] The index is computed based on the average adult years of schooling combined with the expected years of schooling for children. For more detailed information, see .

and divides society into five main groups[8]. In order to causally identify the impact of the middle class share on subsequent educational outcomes, we propose to use the middle three groups as an instrumental variable for the share of the middle class per administrative district. With this novel approach, we are able to empirically prove that larger shares of middle class households causally increase the average educational attainment in said district.

Another factor often found to be of elementary importance when trying to explain (local) differences in economic development, and with it economic well-being, are differences in local transport costs. Figure 1.3 represents a global cost-distance model of local travel times in hours to reach the closest urban area[9]. While travel times in many densely populated areas of the world are of neglectable magnitude, considerable parts of the world are still characterised by extensive degrees of remoteness. This is especially true for nation states which are large in terms of surface area. As indicated by figure 1.3, this is also true for the most spacious country, the Russian Federation. As one can easily observe, vast parts of eastern Russia exhibit travel-times of more than eight hours for reaching the closest urban area. Focussing on the more southern areas of said region, we see that travel-times are considerably lower. One the one hand, this is simply due to the fact that this region is more densely populated. On the other hand, a factor which might also be important for both, the higher population density, but also lower travel times, is the fact that this area is traversed by the Transsiberian Railway. Its construction is held accountable by many for both the existence, as well as the economic viability of settlements in eastern Russia. This impact can largely be attributed to the reduction in transport costs stemming from it.

Transport costs are one, if not the most important determinant of access to remote markets for material goods. Scholars attributable to the theoretical branch of economic analysis

8 The underlying definition is the structure of so-called *varnas*.

9 Urban area in this context is defined as a contiguously populated area with more than 50,000 inhabitants. Please refer to https://forobs.jrc.ec.europa.eu/products/gam/description.php for more detailed information about the methodology.

10 The shading indicates the travel time in hours needed to reach the closest urban area. Urban is defined as a contiguous settlement with more than 50.000 inhabitants. The data for the underlying figure was published by the European Commission's Joint Research Centre and has been obtained from https://forobs.jrc.ec.europa.eu/products/gam/ on 04.03.2019.

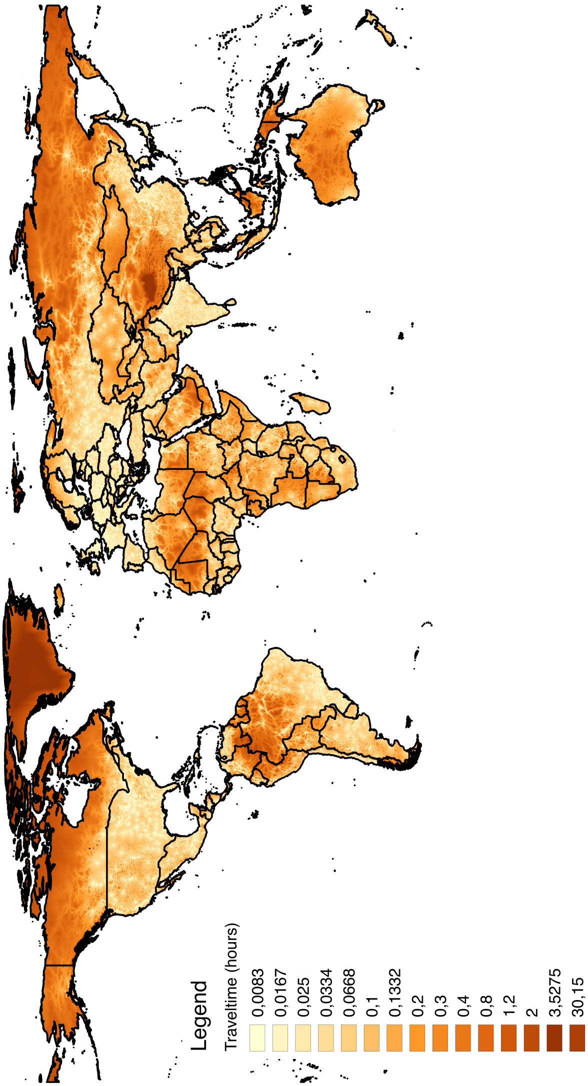


Figure 1.3. *Global Travel Time Map (2009)*[10].

appear to agree on the fact that reductions in said costs result in higher incomes and lower susceptibility with regard to economic shocks (Donaldson, 2018). While there is consensus in the theoretical literature, the question at hand has received little attention with respect to empirical analysis. Chapter 4 contributes to this scarce landscape in multiple ways. First, I construct a novel and highly localised data set of economic activity for eastern Russia[11]. This data set is based on satellite pictures of local nightlight emissions. The presence of which has been established to be a suitable proxy for economic activity by a number of influential studies (e.g. Pinkovskiy & Sala-i-Martin (2014), Michalopoulos & Papaioannou (2013) and Storeygard (2016)). Due to the lack of sub-national accounts data for the region in question, I rely on nocturnal lights emissions in order to approximate local economic activity and its spatial distribution. This data allows me to build my analysis on geo-localised observational units with a size of approximately 11x11km. In doing so, I am able to propose a novel cross-section, mirroring economic activity in a part of Russia for which - to my knowledge - there has not been any reliable data, nor any empirical analyses so far. Locations which are characterised by low transport costs often exhibit favourable natural endowments, as well as are benefitting from the existence of returns to scale. Therefore it is detrimental for the quantification of the causal effect of transport cost advantages that one is able to empirically disentangle said effects. The historical context of the planning and construction of the Transsiberian Railway allows for the implementation of an instrumental variable strategy. This enablers me to solve the underlying endogeneity problem and furthermore, demonstrate the causal positive long-run effect of the railway on economic development in its vicinity. Further, I am able to show that lower distances to the Railway have a causal and positive effect on the spatial agglomeration of economic activity.

11 The underlying definition of eastern Russia in this chapter is based on the units of analysis being situated east of the 60.5 longitude line. This roughly demarcates being east of the Ural mountains.

Bibliography

Alesina, A. & Perotti, R. (1996). Income distribution, political instability, and investment. *European Economic Review*, *40*(6), 1203–1228.

Barro, R. J. (1999). Determinants of democracy. *Journal of Political Economy*, *Vol. 107*(No. 6), S158–S183.

Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, *76*, 169–217.

Cook, P. & Ludwig, J. (2006). The social costs of gun ownership. *Journal of Public Economics*, *90*, 379–391.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, *108*(4-5), 899–934.

Duggan, M., Hjalmarsson, R., & Jacob, B. A. (2011). The short-term and localized effects of gun shows: evidence from California and Texas. *Review of Economics and Statistics*, *93*(3), 786–799.

Gani, F., Sakran, J. V., & Canner, J. K. (2017). Emergency Department Visits For Firearm-Related Injuries In The United States, 2006 â“ 14. *Health Affairs*, *36*(10), 1729–1738.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). {D}o {I}nstitutions {C}ause {G}rowth? *Journal of Economic Growth*, *Vol. 9*, 271–303.

Glaeser, E. L. & Sacerdote, B. (1999). Why Is There More Crime in Cities? *Journal of Political Economy*, *107*(6), 225–258.

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and Social Interactions. *Quarterly Journal of Economics*, *111*(2), 507–548.

Hirschi, T. (1969). *Causes of Delinquency*. University of California Press, Berkley, CA.

Hirschi, T. (1986). On the Compatibility of Rational Choice and Social Control Theories of Crime. In D. Cornish & R. Clarke (Eds.), *The reasoning criminal: Rational choice perspectives on offending*. New York: Springer-Verlag.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*(1), 3–42.

Mc Evoy, C. & Hideg, G. (2017). *Global Violent Deaths 2017*. Small Arms Survey, Graduate Institute of International and Development Studies. Geneva.

Michalopoulos, S. & Papaioannou, E. (2013). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica*, *81*(1), 113–152.

Ostby, G. & Urdal, H. (2011). *Education and Civil Conflict: A Review of the Quantitative, Empirical Literature*, UNESCO.

Pinkovskiy, M. & Sala-i-Martin, X. (2014). Lights, Camera,... Income!: Estimating Poverty Using National Accounts, Survey Means, and Lights.

Putnam, R. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, *6*(1), 65–78.

Putnam, R. (2000). *Bowling alone: The collapse and revival of American community.* Simon&Schuster, NY.

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1994). *Making Democracy Work: Civic Traditions in Modern Italy.* Princeton University Press.

Richmond, T. S., Cheney, R., & Schwab, C. W. (2005). The global burden of non-conflict related firearm mortality. *Injury Prevention*, *11*(6), 348–352.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, *98*(5), S71–S102.

Samuelson, P. A. & Nordhaus, W. D. (1998). *Economics* (16th ed.). McGraw-Hill Higher Education. McGraw-Hill Inc.

Storeygard, A. (2016). Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *The Review of Economic Studies*, *83*(3), 1263–1295.

Strulik, H., Prettner, K., & Prskawetz, A. (2013). The past and future of knowledge-based growth. *Journal of Economic Growth*, *18*(4), 411–437.

UN General Assembly. (1966). International Covenant on Civil and Political Rights. New York: UN General Assembly.

UN General Assembly. (1981). Convention on the Elimination of All Forms of Discrimination Against Women. New York: United Nations.

UN General Assembly. (1990). *Convention on the Rights of the Child* (1577th ed.). New York: United Nations.

Universal Declaration of Human Rights. (1948). New York: United Nations General Assembly.

Vannemann, R. & Dubey, A. (2013). Horizontal and Vertical Inequalities in India. In J. C. Gornick & M. Jäntii (Eds.), *Income inequality - economic disparitites and the middle class in affluent countries* (1st, Chap. 16, pp. 439–458). Stanford: Stanford University Press.

Chapter 2

Gun Violence in the US: Correlates and Causes[1]

2.1 Introduction

It is difficult to overestimate the severity of gun violence in the United States. In the period between 2001 to 2014, the Center for Disease and Control Prevention (CDC) recorded 164,089 firearm homicides. Over the same period of time, the number of non-fatal injuries caused by gunshots is estimated to be more than sixfold – a total of 1,002,647.[2] While these numbers are striking in themselves, the extent of gun violence in the US becomes even more blatant in international comparisons. According to the United Nations Office on Drugs and Crime (UNODC), the number of gun murders per capita in the US in 2012 was nearly 30 times higher compared to the U.K.[3] Not surprisingly, the issue of gun violence has become one of the most pertinent topics in the political and public discourse of the United States. Unfortunately, this debate is still seldomly based on scientific analysis of facts and empirical evidence. The current paper contributes to this discussion by providing a large-scale investigation of the explanatory factors of gun-related offenses using novel county-level data. Moreover, our aim is to go beyond conditional correlations and come closer towards a causal inference of the sources of gun violence in the United States.

[1] Joint work with Bohdan Kukharskyy. Published as Kukharskyy & Seiffert (2017).
[2] Source: https://1.usa.gov/1plXBux and https://1.usa.gov/1qo12RL.
[3] See https://data.unodc.org.

Figures 2.1 and 2.2 provide a first glance at the distribution of gun violence across US counties over the period 2000-2010.[4] More specifically, Fig. 2.1 depicts the average per capita number of gun-caused homicides, while Fig. 2.2 displays the average per capita number of gun-related robberies. Notably, the prevalence of gun violence varies substantially, even within individual states. The average standard deviation of gun-caused homicides (sd = 0.020) and gun-related robberies (sd = 0.261) among counties *within* a given state are comparable in size to standard deviations of the respective offense type *across* all US counties (sd = 0.025 and sd = 0.327, respectively).

What are the factors that can explain this variation? Although the media and press are ripe with anecdotes on potential explanatory factors, there is no consensus on this topic in the literature. To lend structure to this complex debate and to guide our empirical investigation, we develop a novel theoretical model of gun-related crime. In our model, individuals differ with respect to criminal inclinations, defined as the willingness and ability to extract a booty from law-abiding citizens through unlawful behavior (e.g., robbery). Depending on their criminal inclinations, agents decide whether to become law-abiding

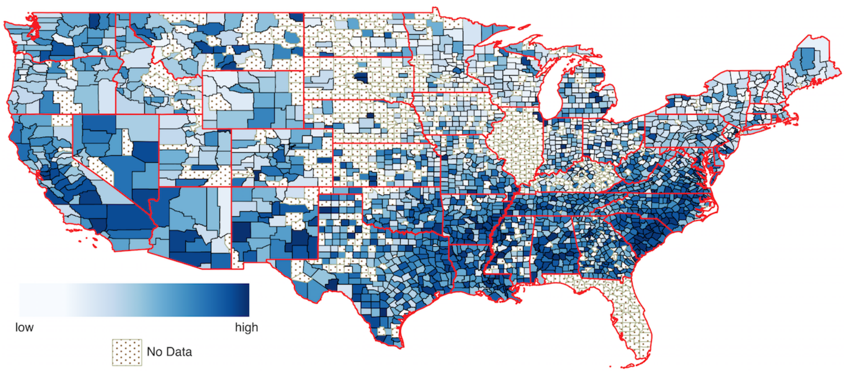


Figure 2.1. *Per capita number of gun-caused homicides, 2000-2010. Data source: Uniform Crime Reporting.*

[4] These figures are constructed using Uniform Crime Reporting data by the Federal Bureau of Investigation (FBI), drawn from https://icpsr.umich.edu/icpsrweb/NACJD/series/57. See section 2.3.1 for data description.

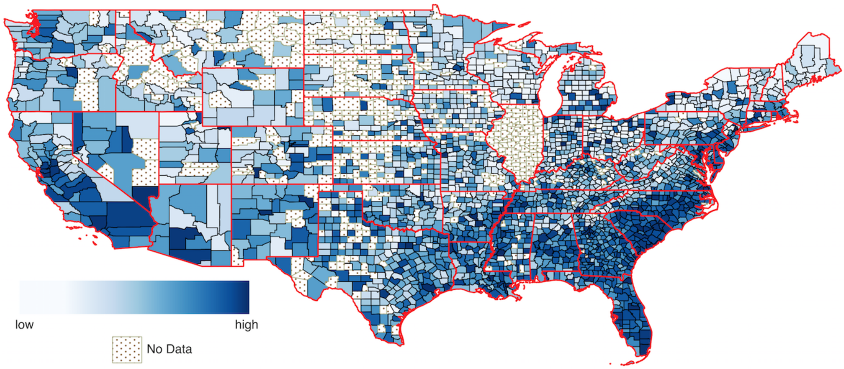


Figure 2.2. *Per capita number of gun-related robberies, 2000-2010. Data source: Uniform Crime Reporting.*

citizens employed in the legal sector or, alternatively, become criminals and earn a living via illegal activities. Individuals who engage in criminal activities choose whether to stay unarmed or acquire a gun and commit firearm-related felonies. Gun acquisition is costly, but possession of a gun has a threatening effect on a victim and allows a felon to reap a higher booty. In equilibrium, only the most criminally inclined individuals commit armed crimes, whereas agents with low criminal inclinations act unarmed.

This simple framework allows us to analyze the effects of various factors on the (per capita) number of firearm offenses in a given county. In particular, we derive the following three key hypotheses: First, gun-related offenses increase with the number of illegal guns. Intuitively, a larger number of illegal guns in circulation decreases the costs of obtaining an illegal weapon and, thereby, increases the expected payoff from gun-related offenses. Second, firearm offenses decrease with the level of social capital, broadly defined as shared beliefs and values that contribute to a well-functioning society. In our model, social capital shapes the distribution of criminal inclinations in a given region: Counties with a high level of social capital have more individuals with low criminal inclinations and fewer individuals with high criminal inclinations. Given that only the most criminally inclined individuals commit a firearm-related crime, gun violence decreases with the level of social capital. Third, gun-related offenses decrease with police intensity. Intuitively, a higher police presence increases the probability of detection and, thereby, decreases the expected payoff from gun-related offenses.

Although the focus of our analysis lies on explaining the causes of *gun-related* offenses, our theoretical framework suggests that the identified key explanatory factors – illegal guns, social capital, and police intensity – drive the variation in *total* (i.e., armed and unarmed) offenses. More specifically, the model predicts that the (per capita) number of offenses in a given county increases with the number of illegal guns and decreases with social capital and police intensity. The intuition behind these predictions draws on the theoretical results that an armed felon commits ceteris paribus more offenses compared to an unarmed one. Hence, even though a lower number of illegal guns, a higher level of social capital, and a higher police intensity may induce some criminals to switch from armed to unarmed offenses, the overall number of offenses in a given region decreases.

To bring our hypotheses to the data, we construct a novel county-level panel dataset which contains information on the number of (gun-related) offenses, police intensity, proxies for the availability of illegal guns, and a wide range of socioeconomic factors. Crime-related information is drawn from the Uniform Crime Reporting (UCR) database for the period 1986-2014. This data is collected by the Federal Bureau of Investigation (FBI) from more than 18,000 local law enforcement agencies and provides detailed county-level information on the incidence of crime known to the police. With more than 90% of US counties represented in this dataset, it adequately serves our goal of giving a comprehensive account of crime in the United States. To the best of our knowledge, it is the only publicly available source of information on gun violence at such a high level of disaggregation.[5] Throughout the analysis, we consider four alternative outcome variables – gun-related robberies, gun-caused homicides, total robberies, and total homicides. We further draw from the UCR annual information on police officers and police employees to measure police intensity in

[5] Apart from a few county-level studies discussed below, the vast majority of research on this topic has been conducted using state-level data, see, e.g., Azrael, Cook & Miller (2004), Fleegler, Lee, Monuteaux, Hemenway & Mannix (2013), Gius (2013), Kalesan, Mobily, Keiser, Fagan & Galea (2016), Lanza (2014) and Siegel, Ross & King (2013, 2014b, 2014a). Clearly, such an approach cannot account for substantial within-state variation in gun violence documented in Fig. 2.1 and 2.2. Our county-level analysis allows us to explore this variation, while effectively controlling for unobserved heterogeneity across states using state fixed effects.

a given county.

This paper suggests a novel proxy for the prevalence of illegal guns.[6] More specifically, we exploit annual UCR information on gun thefts reported to police departments. Given that stolen guns are by definition available to criminals, our proxy provides a *direct* measure for the variation in the number of illegal guns in a given region. A further advantage of our measure lies in its availability for the vast majority of counties over the entire period of 1986-2014.

Following the seminal work by Putnam, Leonardi & Nanetti (1994) and Putnam (1995, 2000), we approximate the level of social capital with the associational density in a given county. To obtain a time-varying measure of associational activism, we exploit annual data on the prevalence of religious, social and civic organizations (such as community, parent-teacher, students', scouting, retirement, or ethnic associations), reported by the US Census Bureau's County Business Patterns (CBP) for the period 1986-2014. The idea behind this proxy is that voluntary participation in (non-profit) associational activities boosts social interaction and cooperation and, thereby, promotes the norms of reciprocity and trust.

We start our empirical analysis by exploring conditional correlations in a cross-section of counties. Controlling for more than a dozen alternative explanations of gun violence (such as organized crime, criminal networks, urbanization, education, fractionalization, poverty), as well as state fixed effects, we find the per capita number of gun-related offenses to be positively correlated with the number of illegal guns and negatively correlated with social capital and police intensity. Although these correlations are in line with our theoretical predictions, they do not allow causal interpretation for at least two reasons: First, the relationships may be confounded by omitted variables (such as history, political preferences, etc.). Second, the results obtained from cross-sectional regressions are prone to the issue of reverse causality: A large number of illegal guns may be the outcome (rather than the source) of a higher prevalence of firearm offenses. Similarly, social capital may

[6] Previous studies used subscriptions to the Guns & Ammo magazine (Duggan, Hjalmarsson & Jacob (2011)) or the percentage of suicides committed with a firearm (Cook & Ludwig (2006)) as indirect proxies for the gun prevalence in a given county.

'deteriorate' whereas police presence may increase in regions where gun-related offenses are frequent. To address both issues, we then turn to panel data analysis. This approach allows us to account for unobservable county-specific factors using county fixed effects. Moreover, by exploiting time-lagged variation in illegal guns, social capital, and police intensity, we move closer towards a causal inference.

Using UCR panel data for the period 1986-2014, we find a positive effect of lagged gun thefts in a given county, and negative effects of lagged associational density and lagged police intensity on the per capita number of gun-related and total offenses, controlling for state-year and county fixed effects. We further document that gun thefts, associational activism, and police intensity from any of the previous three years have a significant impact on the contemporaneous extent of gun violence in a given county. Although this evidence suggests that a high number of illegal guns is not merely a 'byproduct' of firearm offenses, it does not preclude the possibility that criminals steal a weapon in a given year to use it in a future period. In other words, past gun thefts may still be endogenous to current gun violence. We account for this endogeneity problem by constructing an alternative measure of illegal guns based on gun thefts in the neighboring states. More specifically, we calculate for each county the total value of guns stolen in all states adjacent to the one in which a given county is located, weighted by bilateral distances and other relevant factors. The idea behind this proxy builds on the fact that illegal guns are frequently transported across state borders, and a higher number of gun thefts in the neighboring states is likely to increase the number of illegal guns in a given county.[7] The identifying assumption behind this approach is that an individual county is too small to drive the variation in gun thefts across all neighboring states over time. In other words, the total incidence of past gun thefts across all adjacent states is plausibly exogenous to firearm offenses in a single county of the neighboring state.[8] Using this alternative measure, we provide robust evidence for the positive causal effect of illegal guns on the number of gun

[7] According to Mayors Against Illegal Guns (2010), 30% of guns recovered in 2009 from a crime scene in a given state were originally purchased in a different state. Adjacent states constitute the major source of illegal guns, see https://www.atf.gov/about/firearms-trace-data-2015.

[8] We conduct a wide range of robustness checks to preclude possible violations of this identifying assumption.

offenses.

Our theoretical model relates to the economics of crime literature, originating with the seminal contribution by Becker (1968).[9] At the heart of this literature lies the so-called 'deterrence hypothesis', which states that the expected utility of crime ceteris paribus decreases in the probability of detection and in the associated penalty. Our theoretical framework corroborates this hypothesis and contributes to the literature in three major ways. First, it explicitly introduces gun-related illegal activities – alongside unarmed felony – into the model. Second, assuming heterogeneity across individuals with respect to their criminal inclinations, our framework provides for the coexistence of unarmed *and* armed crime in equilibrium. Third, by linking the distribution of criminal inclinations to the level of social capital in a given region, we derive a testable prediction regarding the effect of social norms and values on gun violence.

The latter contribution deserves further attention in light of the literature debate. The Becker (1968) approach of modelling crime solely in terms of economic costs and benefits has invoked some criticism from sociologists and criminologists, who argue that illegal behavior is generally socialized and that crime cannot be fully understood without knowledge of the social background from which it originates, see, e.g., Hirschi (1969, 1986). The latter view is reinforced by the empirical evidence provided by Glaeser, Sacerdote & Scheinkman (1996), who find that no more than 30% of the variation in crime rates within New York City can be explained by pecuniary factors and observable local area characteristics and assert that a major share of differences in crime rates must arise from social norms and civic interactions, cf. also Glaeser & Sacerdote (1999).[10] Our theoretical framework aims to build a bridge between the economic view of crime along the lines of Becker (1968) and alternative conceptions of criminal behavior suggested by sociologists.

From the empirical perspective, our paper relates to two seminal studies that use UCR

[9] See Freeman (1999) and Draca & Machin (2015) for reviews of this literature.

[10] Several studies establish a negative correlation between social capital (as measured by voter turnouts or membership in civic organizations) and crime at the US state level, see Galea, Karpati & Kennedy (2002), Kennedy, Kawachi, Prothrow-Stith, Lochner & Gupta (1998), Messner, Baumer & Rosenfeld (2004), Rosenfeld, Messner & Baumer (2001), Saegert & Winkel (2004). Using instrumental variables approach, recent empirical contributions report a negative causal impact of social capital on crime in Italy (Buonanno, Montolio & Vanin (2009)), Netherlands (Akçomak & ter Weel (2012)), and a cross-section of countries (Lederman, Loayza & Menendez (2002)).

county-level data to investigate the effect of guns on (gun-related) crime. In a panel of the 444 largest counties over the period 1980-1998, Duggan et al. (2011) find a positive relationship between subscriptions to *Guns & Ammo* – one of the nation's largest gun magazines – and homicide rates. Using panel data for the 200 largest counties in the period 1980-1999, Cook & Ludwig (2006) find a positive correlation between the percentage of suicides committed with a firearm – their proxy for the prevalence of guns in the population – and a county's homicide rate. Our contribution to this literature is threefold. First, we suggest a novel, more direct proxy for gun prevalence based on gun thefts. Second, we implement our analysis in a larger sample of (more than 2,500) US counties over a longer period of time. Third, and most importantly, by exploring time variation in illegal guns due to gun thefts in neighboring states, we move closer towards a causal inference regarding the effect of guns on gun violence.

The remainder of the paper is structured as follows. In section 2.2, we develop a simple theoretical model of crime and derive our testable hypotheses. Section 2.3 describes our dataset and presents the empirical results from the cross-section of counties (section 2.3.1) and the panel data analysis (section 2.3.2). In section 2.4, we discuss the policy implications of our work. Section 4.5 concludes.

2.2 The Model

Consider a region (county) populated by a unit measure of individuals who differ with respect to their criminal abilities $c \in (0, 1]$.[11] Individuals with a higher c can ceteris paribus extract a larger booty from law-abiding citizens. Criminal inclinations are distributed according to the cumulative distribution function $F(c)$, with a continuous density function $f(c)$.

Each individual decides whether to become a law-abiding citizen and earn his or her living by legal employment or become a criminal and engage in illegal activities. The compensation of law-abiding citizens is given by a constant wage rate, $w > 0$. Criminals can expropriate wages from law-abiding citizens (for instance, via a robbery). Each felon

[11] Throughout the paper, we use the terms 'criminal ability' and 'criminal inclination' interchangeably.

decides upon the number of offenses (robberies) x, and chooses whether to act unarmed or to buy a gun in order to increase his booty.

Consider first the maximization problem of an unarmed criminal. The booty (b) of an unarmed (u) felon is proportional to the number of committed offenses, his criminal ability, and the victim's wage (income) level, i.e., $b_u = xcw$. This booty can only be reaped with probability $(1-\delta)$, since with the inverse probability $\delta \in (0,1)$ a criminal is detected and caught. In the latter case, a felon is charged with a monetary penalty px, which is proportional to the number of committed offenses (robbed individuals).[12] For simplicity, we assume a constant penalty rate $p > 0$, which can be thought of as a fine or an imprisonment sentence imposed for a given offense.[13] The expected payoff of an unarmed felon can thus be expressed as:

$$\max_x E(\pi_u) = (1-\delta)(xcw)^\alpha - \delta px, \tag{2.1}$$

whereby $\alpha \in (0,1)$ is a constant that governs diminishing marginal utility from a monetary booty. This optimization problem yields the maximum number of offenses committed by an unarmed felon with a criminal ability c:

$$x_u = (cw)^{\frac{\alpha}{1-\alpha}} \left(\frac{1-\delta}{\delta}\frac{\alpha}{p}\right)^{\frac{1}{1-\alpha}}. \tag{2.2}$$

Substituting for x in equation (2.1), we obtain the expected payoff of an unarmed felon:

$$E(\pi_u) = \left(\frac{cw}{p}\right)^{\frac{\alpha}{1-\alpha}} B(\delta), \tag{2.3}$$

whereby

$$B(\delta) \equiv \left(\frac{1-\delta}{\delta^\alpha}\right)^{\frac{1}{1-\alpha}} (1-\alpha)\alpha^{\frac{\alpha}{1-\alpha}} \tag{2.4}$$

is defined for notational simplicity. Note that $B'(\delta) < 0$ for all $\delta, \alpha \in (0,1)$. A simple

12 Our definition of a penalty includes, but is not limited to, imprisonment or unpaid community service, since both punishments deprive an individual of monetary earnings.

13 Assuming non-linear penalties significantly overcomplicates our analysis without changing the main predictions.

inspection of equations (2.2) and (2.3) reveals that both the number of unarmed offenses and the associated expected payoff increase in the felon's criminal ability (c) and in the wage rate of law-abiding citizens (w), and decrease in the probability of detection (δ) and in the associated penalty (p).

Consider now the maximization problem of an armed (a) criminal. Let $g > 0$ denote the costs of obtaining a gun and assume that these costs are the same across all felons in a given region. For any given number of offenses, x, the booty of an armed felon with a criminal ability c is given by $b_a = \lambda xcw$, whereby a constant $\lambda > 1$ reflects an increase in the payoff due to the fact that victims are threatened with a gun. The maximization problem of an armed felon can thus be expressed as

$$\max_x E(\pi_a) = (1 - \delta)(\lambda xcw)^\alpha - \delta px - g. \tag{2.5}$$

This optimization problem yields the maximum number of offenses committed by an armed felon:

$$x_a = (\lambda cw)^{\frac{\alpha}{1-\alpha}} \left(\frac{1-\delta}{\delta} \frac{\alpha}{p} \right)^{\frac{1}{1-\alpha}}, \tag{2.6}$$

and the associated expected payoff:

$$E(\pi_a) = \left(\frac{\lambda cw}{p} \right)^{\frac{\alpha}{1-\alpha}} B(\delta) - g, \tag{2.7}$$

whereby $B(\delta)$ is given by equation (2.4). As before, the number of offenses and the expected payoff increase in a felon's criminal ability and in the wage rate of law-abiding citizens, and decrease in the probability of detection and the associated penalty. It is also evident from the comparison of equations (2.2) and (2.6) that $x_a > x_u$, i.e., an armed felon commits ceteris paribus a larger number of offenses. Yet, the expected payoff of an armed criminal is not necessarily higher than the expected payoff of an unarmed felon because the gain in the booty due to the gun-threatening effect has to be weighted against the costs of obtaining a gun. This tradeoff can be illustrated in a diagram with $c^{\frac{\alpha}{1-\alpha}}$ – a monotonically transformed measure of an individual's criminal inclination – on the horizontal axis, see Figure 2.3. Both $E(\pi_u)$ and $E(\pi_a)$ linearly increase in $c^{\frac{\alpha}{1-\alpha}}$, cf. equations (2.3) and (2.7).

Yet, $E(\pi_a)$ has a negative vertical intercept (due to $g > 0$) and is steeper than $E(\pi_u)$ due to the gun-threatening effect ($\lambda > 1$). Figure 2.3 thus suggests the following sorting pattern: Most criminally inclined individuals engage in armed offenses, since their expect payoff is high enough to compensate the costs of acquiring a gun; individuals with intermediate criminal abilities commit unarmed felonies; the least criminally inclined individuals – whose expected payoff from an unarmed felony $E(\pi_u)$ is smaller than the wage rate w – become law-abiding citizens.

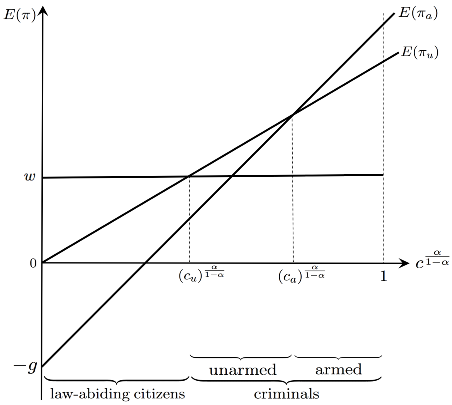


Figure 2.3. *Sorting into legal and illegal activities.*

Using equations (2.3) and (2.7), one can easily derive cutoff criminal inclinations for engaging in unarmed and armed offenses. More specifically, equating the expected payoff from an unarmed felony with the wage rate, $E(\pi_u(c_u)) = w$, one obtains a cutoff criminal inclination, c_u, for which an individual is indifferent between becoming a law-abiding citizen or committing an unarmed offense. All individuals with $c \leq c_u$ are employed in the legal sector while those with $c > c_u$ engage in illegal behavior. From $E(\pi_a(c_a)) = E(\pi_u(c_a))$, we obtain the second threshold, c_a, such that a felon with this criminal inclination is just indifferent between being armed or not, and all individuals with $c > c_a$ commit an armed (rather than unarmed) crime. Using equations (2.3) and (2.7), we obtain:

$$c_u = w^{\frac{1-2\alpha}{\alpha}} pB(\delta)^{-\frac{1-\alpha}{\alpha}} \quad , \quad c_a = \left(\frac{g}{B(\delta)(\lambda^{\frac{\alpha}{1-\alpha}} - 1)}\right)^{\frac{1}{1-\alpha}} \frac{p}{w}. \tag{2.8}$$

Before we discuss the determinants of (armed) offenses, a few remarks are in order. If the $E(\pi_a)$-line is sufficiently flat, the equilibrium cutoff c_a may lie outside of the unit interval, in which case *no* individual has an incentive to commit an armed offense. Conversely, a sufficiently steep $E(\pi_a)$-line may lead to $c_a < c_u$, in which case *all* offenses are firearm-related. In order to ensure that a firearm-related felony is neither a strictly dominated nor a strictly dominant strategy of all criminals, we impose parameter restrictions on exogenous parameters α, δ, p, and w that fulfill

Assumption 1. $0 \leq c_u \leq c_a \leq 1$.

Bearing in mind that the measure of individuals has been normalized to unity, the per capita number of armed offenses in a given region can be expressed as:

$$N_a = \int_{c_a}^{1} x_a f(c) dc, \tag{2.9}$$

whereby x_a and c_a are given by equations (2.6) and (2.8), respectively. Notice that, for any combination of x_a and c_a, the per capita number of firearm offenses depends on the distribution of criminal capabilities in a given region, $f(c)$. To investigate the effect of a society's criminal inclination on the prevalence of firearm offenses, we impose a functional form for $F(c)$. In what follows, we assume that criminal inclinations are distributed according to the bounded (upper-truncated) Pareto function:

$$F(c) = \frac{1 - \left(\frac{c_{min}}{c}\right)^{\kappa}}{1 - c_{min}^{\kappa}}, \tag{2.10}$$

whereby $\kappa > 0$ is the shape parameter of this distribution function, $c_{min} > 0$ represents the lower bound of the support, and the upper bound of c has been set equal to one. Figure 2.4 depicts the Pareto density function $f(c)$ associated with the cumulative distribution function from equation (2.10) for two values of κ – a high and a low one. Lower values of κ reflect a more criminally inclined society and vice versa. The reason for assuming that criminal inclinations are distributed Pareto is twofold. First, as shown in the Online Appendix 2.5, this functional form provides a good fit to the actual distribution of criminal activities within US states and counties. Second, given that the behavior of this distribution function is fully characterized by a single parameter (κ), it allows us to derive our testable predictions in the simplest possible manner.

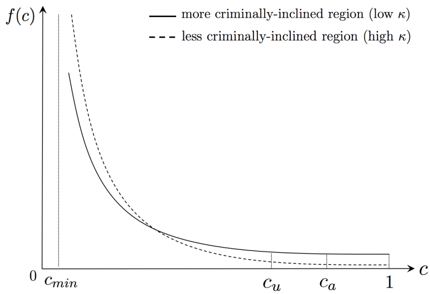


Figure 2.4. *Distribution of criminal inclinations.*

Using equations (2.6), (2.8), (2.9), and (2.10), we establish

Proposition 1. *The per capita number of armed offenses, N_a*

(i) decreases in the costs of obtaining a gun, g

(ii) decreases as the society becomes less criminally inclined, i.e., as κ increases

(iii) decreases in the probability of detection, δ.

Proof. See Online Appendix 2.5.

The intuition behind Proposition 1*(i)* can be easily inferred from Fig. 2.5. An increase in the costs of obtaining a gun, g decreases the expected payoff from an armed felony and the $E(\pi_a)$-line shifts downwards. As a result, the cutoff c_a – above which criminals are willing to engage in a firearm-related crime – rises and the per capita number of gun-related crimes ceteris paribus decreases. The logic behind Proposition 1*(ii)* is illustrated in Fig. 2.4. An increase in κ decreases the density of the distribution function for any $c \geq c_a$ – where criminals commit firearm offenses. Hence, the per capita number of gun-related offenses decreases as the society becomes less criminally inclined. Part *(iii)* of Proposition 1 results from the interplay of two effects. First, an increase in the probability of detection δ reduces the expected benefits from criminal activities for any given c, which can be illustrated as a clockwise pivoting of $E(\pi_u)$ and $E(\pi_a)$ in Fig. 2.6. Yet, given that $\lambda > 1$, the $E(\pi_a)$-line decreases at a higher rate (cf. equations (2.3) and (2.7)). As a result, the equilibrium cutoff c_a increases (cf. equation (2.8)) and the number of *individuals* engaged in armed felonies decreases. Second, a higher probability of detection implies a lower number of *offenses* x_a per armed individual (cf. equation (2.6)). The latter effect reinforces the former and implies a lower per capita number of firearm-related offenses due to an increase in the probability of detection δ.

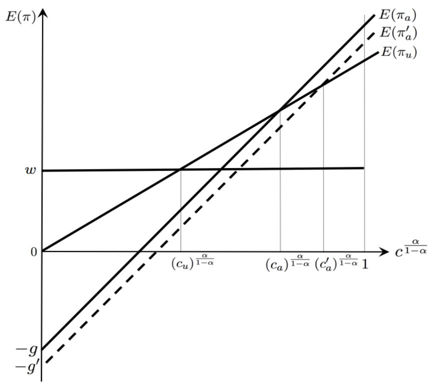


Figure 2.5. *The effect of an increase in gun costs, $g' > g$.*

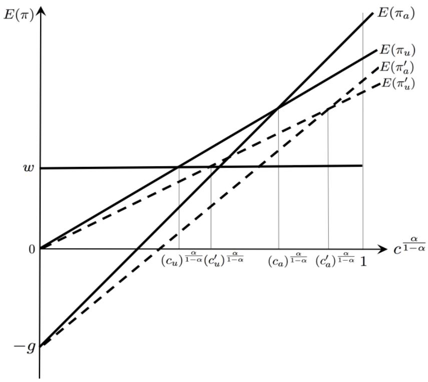


Figure 2.6. *The effect of the probability of detection, $\delta' > \delta$.*

Before turning to the derivation of further results, it is worth pausing to briefly discuss the generality of Proposition 1. First, it should be noted that parts *(i)* and *(iii)* hold for any distribution of criminal inclinations and do not hinge on the specific distributional assumption from equation (2.10). Second, assuming that $F(c)$ is distributed Pareto, the criminal inclination of the society can be alternatively captured as an increase in c_{min}

(rather than a decrease in κ). We verify in the Online Appendix 2.5 that N_a rises in c_{min}. This result reinforces Proposition 1*(ii)* and suggest that the per capita number of armed offenses decreases as the society becomes less criminally inclined.

Our model can be further used to study the effect of crime-related penalties p and the wage rate w on N_a. As shown in the Online Appendix 2.5, the per capita number of armed offenses decreases in p. The logic behind this result can be easily inferred from Fig. 2.3. Due to an increase in p, both $E(\pi_u)$ and $E(\pi_a)$ pivot clockwise, yet the $E(\pi_a)$-line does so at a higher rate (since $\lambda > 1$, cf. equations (2.3) and (2.7)). As a result, the cutoff c_a shifts to the right and fewer criminals commit armed offenses. Moreover, given that x_a decreases in p (see equation (2.6)), the number of offenses committed by an armed criminal ceteris paribus decreases. Both effects imply a lower per capita number of armed offenses due to an increase in p. We further show in the Online Appendix 2.5 that N_a increases in w. The mechanism behind this result can once again be illustrated using Fig. 2.3. Due to an increase in w, both $E(\pi_u)$ and $E(\pi_a)$ pivot counter-clockwise, yet the $E(\pi_a)$-line does so at a higher rate (since $\lambda > 1$, cf. equations (2.3) and (2.7)). Hence, the equilibrium cutoff c_a decreases and more individuals commit armed offenses. Moreover, a higher wage rate of law-abiding citizens induces armed felons to commit a larger number of offenses x_a (cf. equation (2.6)).[14] Hence, the per capita number of armed offenses increases in w. Since we do not explicitly model the legal sector of the economy and follow a very reductionist approach in modeling the penalties, we do not formulate propositions regarding the effects of w and p on N_a. Nevertheless, we account for these factors in our empirical analysis.

Thus far, we have focused on studying the determinants of *armed* offenses. Yet, our model can also be used to derive predictions regarding the number of *total* (i.e., armed and unarmed) offenses. Bearing in mind that that the measure of individuals has been

[14] Note that an increase in the wage rate also raises the opportunity costs of illegal behavior, which can be illustrated as an upward shift of the w-line. Yet, in our simple model, the decision of a criminal whether to commit an armed vs. unarmed offense is unaffected by the criminalâ™s opportunity costs but rather depends on the value of the booty, the probability of detection, and the associated punishment.

normalized to unity, the per capita number of offenses in a given region reads:

$$N = \int_{c_u}^{c_a} x_u f(c)dc + \int_{c_a}^{1} x_a f(c)dc, \tag{2.11}$$

whereby x_u and x_a are given by equations (2.2) and (2.6), respectively, while c_u and c_a are given by equation (2.8). Analyzing this expression, we establish

Proposition 2. *The per capita number of offenses, N*

(i) decreases in the costs of obtaining a gun, g

(ii) decreases as the society becomes less criminally inclined, i.e. as κ increases

(iii) decreases in the probability of detection, δ.

Proof. See Online Appendix 2.5.

Note that g, κ, and δ affect N in the same direction as they impact N_a in Proposition 1. The intuition behind Proposition 2*(i)* can be inferred from Fig. 2.5. Individuals with criminal inclinations $c \in (c_a, c'_a)$ – who would have committed armed offenses before an increase in g – decide to engage in unarmed crime instead. Given that armed felons commit ceteris paribus a higher number of offenses compared to unarmed ones (cf. equations (2.2) and (2.6)), the per capita number of offenses decreases in the costs of obtaining a gun, g. The logic behind Proposition 2*(ii)* is illustrated in Fig. 2.4. An increase in κ decreases the density of the distribution function for any $c \geq c_u$ – where criminals engage in crime – and the per capita number of offenses decreases.[15] Lastly, one can use Fig. 2.6 to infer the intuition behind Proposition 2*(iii)*. Since both c_u and c_a increase in δ (see equation (2.8)), fewer individuals engage in criminal activities. Moreover, individuals with $c \in (c_a, c'_a)$ – who would have previously engaged in armed felonies – switch to unarmed crime, which further reduces the per capita number of offenses due to the fact that $x_a > x_u$ (cf. equations (2.2) and (2.6)).

[15] As before, this result is qualitatively unchanged if we capture an increase in the criminal inclination of a society via an increase in c_{min} (rather than a decrease in κ).

As in the case of Proposition 1, it should be noted that parts *(i)* and *(iii)* of Proposition 2 do not hinge on the assumption of Pareto-distributed criminal inclinations and are established for a general distribution function $F(c)$. One can further show that the per capita number of total offenses N decreases in the penalty rate p. Yet, the effect of the wage rate w on N is no longer unambiguously positive. The reason behind this ambiguity depends on the interplay of two effects. On one hand, a higher income of law-abiding citizens ceteris paribus raises the monetary booty and increases the number of offenses. On the other hand, an increase in w raises the opportunity costs of unarmed crime and induces some unarmed felons to become law-abiding citizens. Without imposing further restriction on model parameters, the overall effect of w on N is ambiguous.

2.2.1 Hypotheses

In this section, we draw insights from the economics, sociology, and criminology literature to map key model parameters to observable factors and, thereby, formulate our testable hypotheses.

What determines the costs of obtaining a gun, g? According to the recent report by the US Department of Justice (Planty & Truman (2013)), the primary source of firearms for criminals is an illegal market (see also Cook, Parker & Pollack (2015)). Cook, Ludwig, Venkatesh & Braga (2007) provide some insight into the underground gun market by conducting interviews with gang members and gun dealers in the city of Chicago. One of the key insights of this study is that the underground gun market is 'thin', and that the acquisition of an illegal firearm is associated with substantial transaction (search) costs and large mark-ups over legal prices. A standard economic analysis of such a market would imply that the costs of obtaining an illegal gun are decreasing in the supply of illegal guns. We thus maintain the following functional relationship:

$$g = \mathrm{f}(\underset{-}{\text{illegal}}\ \text{guns}).$$

How do we map the criminal inclination of a given county $(1/\kappa)$ to the data? Philosophers such as David Hume, Immanuel Kant and John Stuart Mill have long emphasized the role of moral sentiments such as guilt, shame, and remorse in shaping moral behavior and, in particular, an individual's willingness to commit a crime.[16] Recent theoretical contributions by Bénabou & Tirole (2006, 2011), Funk (2006) and Weibull & Villa (2006) study these aspects by explicitly introducing social norms into the models of crime, see McAdams & Rasmusen (2007) and van der Weele (2012) for reviews of this literature. Since the seminal contributions by Coleman (1988), Coleman & Coleman (1994) and Putnam et al. (1994), Putnam (1995, 2000), sociologists and political scientists generally refer to the shared values and effective norms that evoke those sentiments and, thereby, prevent a person from committing a crime as 'social capital'.[17] As discussed in the introduction, ample empirical evidence suggests that social capital has a crime-deterring effect. Based on this evidence, we assert that κ – an inverse measure of a society's criminal inclination – is a positive function of social capital:

$$\kappa = \mathrm{f}(\underset{+}{\text{social capital}}).$$

Next, consider the probability of detection, δ. Arguably, this probability is primarily a function of police intensity. Since the seminal contribution by Levitt (1997), economists have suggested several strategies to identify the causal effect of policing on crime deterrence, see Nagin (2013) and Draca & Machin (2015) for reviews of this literature. Among the most convincing approaches, is the usage of terrorist attacks or alerts as an instrument for exogenous (re-)allocations of police resources. In such a quasi-experimental setting, several contributions find a robust positive effect of police intensity on crime deterrence in many cities, including Buenos Aires (Di Tella & Schargrodsky (2004)), the District of Columbia (Klick & Tabarrok (2005)), London (Draca, Machin & Witt (2011)), and Stockholm (Poutvaara & Priks (2006)). In view of this evidence, we treat δ as a positive

[16] The role of remorse and mental anguish in a criminal's moral dilemma is succinctly summarized in Dostoevsky's "Crime and Punishment": "If [a thief] has a conscience, he will suffer for his delinquency. That will be his punishment – as well as the prison."

[17] According to Coleman & Coleman (1994), social capital is the set of relationships that support effective norms "[...] that inhibit crimes in a city, make it possible for women to walk freely outside at night and for old people to leave their homes without fear."

function of police intensity:

$$\delta = \text{f}(\underset{+}{\text{police intensity}}).$$

Above-mentioned inquiries merely suggest functional dependencies of the model parameters, g, κ, and δ. Combining these relationships with our results derived in Propositions 1 and 2, we expect a positive effect of illegal guns and a negative effect of social capital and police intensity on the per capita number of armed (N_a) and total (N) offenses (henceforth, summarized as $N_{(a)}$):

$$N_{(a)} = \text{f}(\underset{+}{\text{illegal guns}}, \underset{-}{\text{social capital}}, \underset{-}{\text{police intensity}}). \tag{2.12}$$

Before turning to the empirical implementation of our hypotheses, it is worth pausing to discuss some potential concerns with our analysis. First, our model is admittedly very simple. In particular, it does not allow law-abiding citizens to (legally) acquire firearms in order to protect themselves from offenders.[18] Given that official county-level data on legal gun ownership are, to the best of our knowledge, not available, we do not formulate a hypothesis regarding the impact of legal guns on the relative prevalence of firearm offenses in the first place.[19] Nevertheless, our empirical analysis considers indirect proxies for legal gun ownership suggested in the literature (see footnote 6). Moreover, to the extent that the stock of legal guns in a given county is determined by state-specific gun control laws, we account for this potential confounding factor using state fixed effects.

Second, one can rightly argue that illegal guns, social capital, and police intensity affect $N_{(a)}$ via more than one model parameter. What are the potential alternative channels? For instance, one might assert that social capital has a positive effect on the probability of detection, δ. Intuitively, members of communities with pronounced civic participation are more likely to report crimes to the police, bring disputes to the attention of courts

[18] The effect of legal gun ownership on crime is highly debated in the literature. Lott & Mustard (1997) and Bronars & Lott (1998) argue that a higher prevalence of firearms among law-abiding citizens might reduce crime. Yet, several more recent empirical studies have shown that the "more guns, less crime" hypothesis does not hold empirically, see, e.g., Duggan et al. (2011) and Ayres & Donohue (2003).

[19] In Kukharskyy & Seiffert (2016), we study the effect of legal gun ownership on crime using novel state-level data.

and law enforcement agencies, and engage in public surveillance. Yet, given that the prevalence of firearm offenses, is decreasing both in κ and δ (see Propositions 1 and 2), this alternative channel reinforces the predicted negative effect of social capital on $N_{(a)}$. Furthermore, one can argue that police intensity is associated with a higher cost of obtaining a gun, g.[20] Given that the relationship between g and $N_{(a)}$ is inversely proportional, the predicted effect of police intensity on the per capita number of (gun-related) offenses remains negative. One might also hypothesize a negative relationship between the prevalence of illegal guns and the probability of detection and/or deterrence, δ. Intuitively, if a civilian observes a suspicious activity or an act of violence, he or she is generally less likely to intervene the higher the chances of encountering an armed felon. Yet, once again, given that δ negatively effects $N_{(a)}$, this alternative channel would only reinforce our predictions.

Third, one can certainly envision arguments for why the above-mentioned explanatory factors may affect $N_{(a)}$ in the opposite direction to the one predicted by equation (2.12). For instance, one can argue that a higher level of social capital increases trust among felons, advances the emergence of criminal networks, and, therefore, increases gun violence in a given region. We take these (and other) objections seriously and include proxies for criminal networks, organized crime, as well as a wide range of alternative explanatory factors into our regressions. On balance, we believe that our theoretical model provides a helpful roadmap for the directionality of the effects and proceed with the empirical analysis.

2.3 Empirical Implementation

The structure of our empirical investigation is as follows. In section 2.3.1, we study in a cross-section of counties conditional correlations between the per capita number of offenses and the key explanatory variables – illegal guns, social capital, and police intensity. To come closer towards a causal inference of these effects, we turn to panel data analysis in section 2.3.2. In each section, the main focus lies on studying the determinants of *gun-*

[20] Cook et al. (2015) provide some anecdotal evidence for this claim.

related offenses, i.e., testing Proposition 1. However, we also consider the effects of illegal guns, social capital, and police intensity on *total* (i.e., armed and unarmed) offenses, as suggested by our Proposition 2.

2.3.1 Cross-Section Analysis

Data and Econometric Specification

Our primary source of information on (gun-related) crime in the US is the Uniform Crime Reporting (UCR) data by the United States Department of Justice and Federal Bureau of Investigation (n.d.) (FBI). This database provides detailed information on crime known to the police, collected from more than 18,000 local law enforcement agencies (LEAs). With more than 90% of counties represented in the database, UCR meets fairly well its goal of providing an overall view of criminal activities in the US[21] Due to the fact that this database is publicly available, it has become the workhorse tool in empirical studies of crime, see, e.g., Glaeser & Sacerdote (1999), Duggan et al. (2011), Cook & Ludwig (2006), Cook et al. (2007).[22] In the following, we provide a brief description of the key variables of interest and relegate the detailed discussion of the (step-by-step) construction of these variables to the Online Appendix 2.5. Summary statistics for the main estimation samples are provided in Table 2.A.1.

The UCR database is structured under the following four key categories: (a) Offenses Known and Clearances by Arrest (OKCA), (b) Supplementary Homicide Reports (SHR), (c) Law Enforcement Officers Killed or Assaulted (LEOKA), and (d) Property Stolen and Recovered (PSR). We use the first two datasets to construct our dependent variables and draw a range of right-hand side variables from the latter ones. All four datasets are available on an annual basis for the period 1986-2014. We exploit the entire timespan in the panel analysis and consider annual averages over the period 2000-2010 in the cross-

[21] Due to diverging data collection methodologies, information for Florida, Illinois (except for Cook county, Chicago), and a few individual counties from other US states is oftentimes missing, see Fig. 2.1 and 2.2.

[22] See, however, Maltz (1999) for a detailed discussion of the limitations of this data. We summarize the main caveats of the UCR data further below and suggest adequate empirical strategies to account for these limitations.

section. Using the correspondence provided by the US Department of Justice, we map the LEA-level data to individual counties – the unit of observation in our analysis.[23]

At the highest level of abstraction, the issue of gun violence has two dimensions – non-lethal and lethal. We approximate the former aspect using information on gun-related robberies from the UCR's OKCA database. More specifically, we take (the log of) the per capita number of gun-related robberies in a given county as our first key dependent variable (henceforth, *GunRobberies*). This outcome variable is well-suited for the analysis of the predictions of our economic model of crime.[24] Using OKCA, we further construct a measure of *TotalRobberies*, defined as the per capita number of total (i.e., armed and unarmed) robberies in a given county.

To capture the second, lethal dimension of gun violence, we use UCR's SHR data. This database reports, among other things, the type of weapon and the circumstance under which a homicide was committed. During the construction of our baseline measure of homicides, we exclude all circumstances indicating an accident (such as 'gun cleaning', 'child playing with gun', etc.), negligence (e.g., 'child killed by babysitter'), or law enforcement killings ('felon killed by police', 'suspected felony', etc.).[25] We then calculate the (log of the) per capita number of firearm-caused homicide incidents in a given county (henceforth, *GunHomicides*) and the (log of the) per capita number of total homicide incidents (henceforth, *TotalHomicides*).[26] To be clear, our theoretical framework does not explicitly encompass (gun-caused) homicides. Yet, one can envision a simple extension of the model in which a gun-related robbery results in the (probabilistic) discharge of the firearm. In such a model, the number of (gun-caused) homicides in a given county would be a positive function of illegal guns and a negative function of social capital and police intensity. However, due to the fact that, in reality, some murders are committed by ordinary citizens for non-economic reasons (such as hatred and animosity), we expect

[23] We choose a slightly higher level of aggregation due to unavailability of control variables at the LEA-level.

[24] Information on usage of guns in other 'economic' offenses (such as burglary or larceny) is unavailable.

[25] See Online Data Appendix 2.5 for the full list of excluded categories.

[26] A homicide incident is an event in which one or more persons are killed at the same place and time. Measures of *GunHomicides* and *TotalHomicides* based on the victim count yield similar results, available upon request.

a weaker effect of factors such as probability of detection or the prevalence of illegal guns on (gun-caused) homicides compared to (gun-related) robberies.

Our baseline econometric specification for the cross-section of counties (c) reads:

$$N_{(a)c} = \beta_1 IllegalGuns_c + \beta_2 SocialCapital_c + \beta_3 PoliceIntensity_c + \chi \boldsymbol{X}_c + \rho_s + \varepsilon_c, \quad (2.13)$$

whereby $N_{(a)c} \in \{GunRobberies_c, GunHomicides_c, TotalRobberies_c, TotalHomicides_c\}$ is the (log of the) average per capita number of a given offense type in 2000-2010, $\boldsymbol{X}_c$ is a vector of county-level controls, ρ_s denotes state fixed effects, and ε_c is the error term.[27] Our theoretical model predicts a positive estimate $\hat{\beta}_1 > 0$, and negative estimates $\hat{\beta}_2 < 0$ and $\hat{\beta}_3 < 0$, see equation (2.12).

We suggest a novel measure for the number of illegal guns based on gun thefts reported in the UCR's PSR database. More specifically, we utilize the annual information on the value of firearms stolen in a given county and take (the log of) the average value in 2000-2010 as our cross-sectional proxy for the prevalence of *IllegalGuns*. Unfortunately, this database does not provide information on the quantity or type of stolen guns. However, it is known from the National Crime Victimization Survey that the vast majority of stolen guns are handguns, see Langton (2012) and Zawitz (1995). Given that the price range for revolvers and pistols is fairly narrow, we believe that our value-based measure provides a good approximation for the number of illegal guns.

We approximate the level of social capital with the associational density, calculated using annual data from the US Census Bureau's County Business Patterns (CBP) for the period 1986-2014. More specifically, we draw from the CBP information on the number of and employment by "religious, grantmaking, civic, professional, and similar organizations", classified according to the 813 code of the North American Industry Classification System (NAICS).[28] Examples of establishments falling into this category are community and ethnic organizations, parent-teacher associations, human rights organizations, and

[27] To simplify the notation, we drop the county-subscript c henceforth.

[28] In 1998, the CBP changed the industry classification from the Standard Industrial Classification (SIC) to North American Industry Classification System (NAICS), whereby religious, social, and civic organizations were classified under the SIC code "86" in the period 1986-1997.

religious and charitable organizations. More than 80% of employment associated with the NAICS code 813 is accounted for by the two more narrowly defined NAICS codes: 8131 ("religious organizations") and 8134 ("social and civic organizations"). Since information on the NAICS code 813 is available for a larger number of counties, we use it for the construction of our baseline proxy, but consider the two more disaggregated codes in the robustness checks. We construct four alternative measures of *SocialCapital* (all expressed in terms of natural logarithms): *(i) employment* by the organizations classified under the NAICS code 813 over the total employment in a given a county, *(ii) employment* by the organizations classified under the NAICS code 813 per capita, *(iii)* the number of *establishments* classified under the NAICS code 813 over the total number of establishments in a given county, *(iv)* the number *establishments* classified under the NAICS code 813 per capita. We use the first measure as our baseline proxy for social capital and consider the other three measures in the robustness checks. The idea behind approximating social capital with the associational density builds on the seminal work by Putnam et al. (1994), Putnam (1995, 2000), who shows that participation in associational activities boosts interaction and cooperation between community members and promotes the norms of reciprocity and trust. The advantage of our measure compared to alternative proxies suggested in the literature (such as voluntary blood donations or voter turnouts) is that it exploits official data from the US Census and is therefore characterized by a high degree of validity and consistency. It is additionally well-suited for the ensuing panel data analysis since this measure is available on an annual basis for the vast majority of US counties over the entire period of 1986-2014. In the cross-sectional analysis, we take the (log of the) associational employment density averaged over 2000-2010 as our measure of *SocialCapital*.

Information on police intensity is drawn from the UCR's LEOKA database. For each LEA, the LEOKA database reports, among other things, the number of police officers and police employees per 1,000 population. To construct our baseline measure of police intensity, we calculate for each year the weighted average of the police officers rate across all LEAs of a given county with weights being the fraction of a county's population served by a given LEA.[29] In the cross-sectional analysis, we take (the log of) the police officers rate averaged over 2000-2010 as our proxy for *PoliceIntensity*.

The choice of variables for the vector of controls is motivated by our theoretical model, the public debate on this issue, and related empirical findings. Our model suggests that the per capita number of (gun-related) offenses depends positively on the wage rate of law-abiding citizens w. As a proxy for w, we use (the log of) a county's per capita *Income* averaged over 2000-2010, collected from the US Census' Small Area Income and Poverty Estimates (SAIPE) database.[30] Poverty may force citizens into illegal behavior and, potentially, compel them to acquire guns in order to raise the associated booty. To account for this potential confounding factor, we draw from the SAIPE database information on the percentage of a county's population living below the poverty line and take (the log of) this value averaged over 2000-2010 as a measure of *Poverty*. We further control for *IncomeInequality*, measured as (the log of) a county's Gini coefficient, as reported by the 2006-2010 American Community Survey (ACS).

To control for the overall level of crime, we draw from the UCR's OKCA information on the total number of offenses across all crime categories and take (the log of) this per capita number averaged over 2000-2010 as our measure of *CrimeRate*. As mentioned in the previous section, one might be concerned that the level of social capital merely reflects the prevalence of criminal networks and organized crime. To account for this alternative explanation, we construct the following two control variables using the UCR's SHR data. *OrganizedCrime* is calculated as (the log of 0.001 plus) the average share of 'gangland killings' and 'juvenile gang killings' in the total number of homicide incidents by county in 2000-2010. *CriminalNetworks* is constructed as (the log of 0.001 plus) the average share of homicides committed by more than one person in total homicides in 2000-2010.

Several recent contributions suggest a positive link between a society's fractionalization and conflict, see, e.g., Arbatli, Ashraf & Galor (2015) and reference therein. This re-

[29] The reason for using weighted averages derives from the fact that some small LEAs may have high police officers rates due to the surveillance of correctional facilities, and simple averages would potentially overstate the police intensity in a given county. However, the results are very similar when we consider non-weighted averages.

[30] This data is drawn on an annual basis from https://www.census.gov/did/www/saipe/data/statecounty/data/.

lationship might be particularly pronounced in one of the most diverse countries in the world – the United States. We consider two different measures of fractionalization – ethnic (*EthnicFrac*) and racial (*RacialFrac*). The former measure is constructed as follows. Using 2006-2010 ACS information on the country of birth of the foreign-born US population, we calculate for each county the share s of ethnic group e stemming from one of the 108 distinct countries of origin. We then aggregate these shares to a Herfindahl index, $EthnicFrac = \ln\left(1 - \sum_{i=e}^{E} s_i^2\right)$, whereby higher values of this index represent a higher ethnic fractionalization in a given county.[31] Furthermore, using information on racial composition of counties from the US Decennial Census 2010, we calculate for each county the share (s) of a racial group (r) – 'Black or African American', 'White American', 'Hispanics', 'American Indian or Native Alaskan American', 'Asian American' and 'Native Hawaiian and other Pacific Islander' – in a county's population and aggregate these shares to a Herfindahl index, $RacialFrac = \ln\left(1 - \sum_{i=r}^{R} s_i^2\right)$, whereby higher values of this index represent a higher racial fractionalization in a given county. We also include the (log of the) percentage of *AfricanAmerican* population in a given county as an additional control variable and verify that our results are robust to controlling for the prevalences of other racial groups.

To account for a possible effect of educational attainment on the willingness of individuals to commit a (gun-related) offense, we control for *Education*, constructed as the (log of the) percentage of over-25 years old citizens with at least a high school degree, as reported by the 2006-2010 ACS. To control for the potential impact of urbanization on the costs of obtaining a gun (g) and the probability of detection (δ), we draw from the 2010 US Census Urban and Rural Classification information on the fraction of a county's population living in urban areas and take the log of this variable as our measure of *Urbanization*. We further control for the (log of the) percentage of children (6-17 years old) living in a *SingleParent* household, drawn from the 2006-2010 ACS.

Administrative information on legal gun ownership at the county level is, unfortunately,

[31] In using the Herfindahl method to construct a measure of fractionalization, we follow Alesina, Devleeschauwer, Easterly, Kurlat & Wacziarg (2003) and Fearon (2003). Our results are virtually unchanged if we capture fractionalization using standard deviations.

unavailable. Azrael et al. (2004) and Cook & Ludwig (2006) approximate the access to guns with the percentage of suicides committed with a firearm. The idea behind this measure is that, if the willingness to commit a suicide is equally distributed across regions, a higher fraction of firearm suicides in total suicides reveals a higher gun ownership in a given region. Following this approach, we use data on suicides from the Center for Disease and Control Prevention (CDC) to control for *LegalGuns*, constructed as the (log of the) share of suicides committed with a firearm in 2004-2010.[32]

Recall from the previous section that the number of (gun-related) offenses depends negatively on the penalty rate, p. Given that the responsibility for criminal law and criminal justice in the US is shared between the federal and state governments, we control for state-specific differences in criminal laws using state fixed effects, included in all regressions.

OLS Estimations

Table 2.1 reports the results of Ordinary Least Squares (OLS) regressions specified in equation (2.13) with *GunRobberies* as a dependent variable. As can be seen from columns (1) and (2), *GunRobberies* are positively correlated with the number of *IllegalGuns* and negatively correlated with the level of *SocialCapital*, respectively. The coefficient of *PoliceIntensity* in column (3) is negative but not significant. However, it becomes significant after controlling for a county's per capita income, poverty rate and income inequality in column (4). All three key explanatory variables – illegal guns, social capital, and police intensity – remain fairly robust in size and significance after including a range of additional control variables in columns (5)-(7). In line with the model's predictions, *GunRobberies* are positively correlated with the number of *IllegalGuns* and negatively correlated with *SocialCapital* and *PoliceIntensity*. The number of gun robberies per capita also tends to be higher in richer and more unequal counties, which have a high (organized) crime

[32] This data is drawn from https://wisqars.cdc.gov:8443/cdcMapFramework/. We also verify that our results are robust to controlling for subscriptions to Guns&Ammo magazine – an alternative proxy for gun prevalence suggested by Duggan et al. (2011). Given that information on Guns&Ammo subscriptions is available only for a small subset of counties, we do not include this proxy in our baseline regressions but provide the results upon request.

rate and a strong prevalence of criminal networks, are racially fragmented, and have a high fraction of African American population and single-parent households. In contrast, counties with a high level of urbanization and education seem to have a lower number of gun robberies per capita. The coefficient of determination in our preferred specification in column (7) suggests that our main explanatory variables, the extensive list of controls, and state fixed effects jointly explain about two-thirds of the cross-sectional variation in gun-related robberies in the US In column (8), we further control for the prevalence of legal guns, which reduces our sample by half. All three key explanatory variables remain robust and highly significant. The positive coefficient of *LegalGuns* suggests that the number of per capita gun robberies is higher in counties with a higher prevalence of legal guns.

Next, we rerun the above-mentioned regressions using *GunHomicides* as a dependent variable, see Table 2.A.2 in Appendix. Throughout specifications, *GunHomicides* are positively and highly significantly correlated with the number of *IllegalGuns*. Apart from column (8), in which the sample is reduced by half, the negative coefficient of *SocialCapital* is also highly significant. The coefficient of *PoliceIntensity* is throughout negative but not significant after including the full set of controls. The lack of significance can be rationalized by the above-mentioned fact that homicide crimes are oftentimes perpetrated "in the heat of moment" and may not be affected by the probability of detection. The coefficients of control variables are comparable to the ones reported in Table 2.1.

Having explored the correlates of *GunRobberies* and *GunHomicides*, we now rerun our regressions using *TotalRobberies* and *TotalHomicides* as dependent variables. Table 2.2 presents the results of our preferred specification with state fixed effects and the full set of controls from column (7) of Table 2.1.[33] In line with our theoretical predictions, per capita robberies and homicides are positively correlated with *IllegalGuns* and negatively associated with *SocialCapital* and *PoliceIntensity*.

In summary, the evidence presented so far is generally consistent with our theoretical predictions: The number of (gun-related) offenses is positively associated with the number

33 Since the estimates of control variables are similar to the ones from Table 2.1, we do not report them for brevity.

Table 2.1. *Cross-section estimates: Correlates of gun robberies.*

Dep.variable: *GunRobberies*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IllegalGuns	0.226***	0.255***	0.266***	0.198***	0.054**	0.081***	0.058***	0.072***
	(0.013)	(0.013)	(0.014)	(0.023)	(0.023)	(0.022)	(0.022)	(0.025)
SocialCapital		-0.164***	-0.163***	-0.126***	-0.136***	-0.162***	-0.134***	-0.149***
		(0.030)	(0.030)	(0.029)	(0.027)	(0.025)	(0.025)	(0.035)
PoliceIntensity			-0.036	-0.075***	-0.051**	-0.067***	-0.051***	-0.070***
			(0.022)	(0.023)	(0.021)	(0.019)	(0.019)	(0.019)
Income				-0.122***	0.082***	0.153***	0.096***	-0.205***
				(0.030)	(0.030)	(0.028)	(0.030)	(0.038)
Poverty				0.424***	0.176***	0.086*	0.068	-0.125*
				(0.055)	(0.050)	(0.048)	(0.060)	(0.073)
Inequality				1.499***	1.196***	1.062***	0.643***	0.395
				(0.239)	(0.221)	(0.207)	(0.221)	(0.276)
CrimeRate					0.560***	0.437***	0.517***	0.678***
					(0.033)	(0.031)	(0.033)	(0.045)
OrganizedCrime					0.157***	0.142***	0.129***	0.065***
					(0.008)	(0.008)	(0.008)	(0.007)
CriminalNetworks					0.026***	0.019***	0.015***	0.012*
					(0.006)	(0.005)	(0.005)	(0.007)
EthnicFrac						0.007	-0.013	-0.013
						(0.035)	(0.034)	(0.049)
RacialFrac						0.112***	0.085**	0.072
						(0.036)	(0.036)	(0.046)
AfricanAmerican						0.213***	0.214***	0.303***
						(0.018)	(0.018)	(0.023)
Education							-0.370***	-0.269**
							(0.094)	(0.113)
Urbanization							-0.048***	-0.046***
							(0.004)	(0.011)
SingleParent							0.140**	0.659***
							(0.061)	(0.089)
LegalGuns								0.139**
								(0.055)
Observations	2,499	2,479	2,477	2,477	2,284	2,264	2,264	1,221
R-squared	0.344	0.366	0.369	0.423	0.575	0.642	0.663	0.860

Note: The table reports estimates of equation (2.13) with *GunRobberies* as a dependent variable. All specifications include state fixed effects. Standard errors are reported in parentheses. *, **, *** indicate significance at 1, 5, 10%-level, respectively.

of illegal guns and negatively related to the level of social capital and police intensity in a given county. Yet, these conditional correlations should not be interpreted as indicative of causal relationships for two main reasons. First, even though we control for a wide range of alternative explanations and state fixed effects, there may be other (unobservable) country-specific factors that confound this relationship. For instance, the historical

Table 2.2. *Cross-section estimates: Correlates of total robberies and total homicides.*

	Dependent variable	
	TotalRobberies	*TotalHomicides*
IllegalGuns	0.034*	0.178***
	(0.018)	(0.019)
SocialCapital	-0.079***	-0.041**
	(0.020)	(0.021)
PoliceIntensity	-0.027*	-0.005
	(0.016)	(0.017)
Observations	2,383	2,448
R-squared	0.868	0.619

Note: The table reports estimates of equation (2.13) with *TotalRobberies* and *TotalHomicides* as dependent variables. All specifications include state fixed effects and full set of controls from column (7) of Table 2.1. Standard errors are reported in parentheses. *, **, *** indicate significance at 1, 5, 10%-level, respectively.

incidence of slavery in a given county might explain both a low level of social capital (see Nunn & Wantchekon (2011)) and a high prevalence of gun-related crime. Second, the results presented above are prone to the issue of reverse causality. Consider for instance the link between illegal guns and gun violence. It is possible that criminals seize a gun from an armed victim in the course of a firearm offense, making gun thefts merely a 'byproduct' of gun-related offenses. Moreover, criminals may undertake armed offenses in order to steal additional guns, in which case a high prevalence of gun thefts is the outcome (rather than the source) of frequent firearm offenses. Likewise, a low level of social capital may be both the cause and the outcome of gun violence: Individuals in counties with a low level of trust may be more likely to pull the trigger, but the level of social capital itself may deteriorate due to frequent firearm offenses. Lastly, police intensity is likely to increase as (gun-related) offenses in a given region become more frequent. This type of endogeneity works against the predicted negative effect of police intensity and might provide a further potential explanation behind the weak statistical significance of *PoliceIntensity* in Tables 2.2 and 2.A.2.

To address the concerns related to the omitted variable bias and reverse causality, we turn to panel data analysis. This approach allows us to account for unobservable time-invariant characteristics of a county using county fixed effects. Moreover, by exploiting time-lagged variation in illegal guns, social capital, and police intensity, we move closer towards a causal inference.

2.3.2 Panel Data Analysis

The baseline econometric specification in this section takes the following form:

$$N_{(a)ct} = \beta_1 \mathit{IllegalGuns}_{c,t-1} + \beta_2 \mathit{SocialCapital}_{c,t-1} + \beta_3 \mathit{PoliceIntensity}_{c,t-1} + \rho_c + \rho_{st} + \chi \boldsymbol{X}_{ct} + \varepsilon_{ct}, \tag{2.14}$$

where $N_{(a)ct} \in \{\mathit{GunRobberies}_{ct}, \mathit{GunHomicides}_{ct}, \mathit{TotalRobberies}_{ct}, \mathit{TotalHomicides}_{ct}\}$ in county c and year t, and $\mathit{IllegalGuns}_{c,t-1}$, $\mathit{SocialCapital}_{c,t-1}$, and $\mathit{PoliceIntensity}_{c,t-1}$ capture, respectively, illegal guns, associational density, and police intensity from the previous period $t-1$.[34] We conduct our analysis for the period 1986-2014, whereby the starting year of the panel is determined by the availability of data on associational density from the CBP. County-specific fixed effects ρ_c account for time-invariant characteristics of a county (such as geography or history) as well as factors that are relatively stable over time (e.g., urbanization). Year fixed effects ρ_t control for aggregate time-specific shocks. In an even more stringent specification, we include state-year fixed effects ρ_{st}, which effectively control for all time-varying state-specific factors, such as gun legislation or criminal laws. Our vector of time-varying county-level controls, $\boldsymbol{X}_{ct}$ includes $\mathit{CrimeRate}_{ct}$, Income_{ct}, and $\mathit{Poverty}_{ct}$, whereby all variables are defined by analogy to section 2.3.1.[35] In all regressions, standard errors are clustered at the county level to adjust for within-county correlation over time. To simplify the notation, we drop the county-subscript c

[34] All variables are defined as in section 2.3.1, apart from $\mathit{GunHomicides}_{ct}$ and $\mathit{TotalHomicides}_{ct}$, which are constructed as $\ln(0.001 +$ per capita number of armed offenses) and $\ln(0.001 +$ per capita number of total offenses), respectively. The reason for adding a small constant (0.001) lies in the fact that most counties feature zero (gun-related) homicides in a given year and these observations would be omitted in the logarithmic specification.

[35] Data on $\mathit{CrimeRate}_{ct}$ in 1993 is missing in the UCR database. In our baseline analysis, we replace $\mathit{CrimeRate}_{c1993}$ by an average of $\mathit{CrimeRate}_{c1992}$ and $\mathit{CrimeRate}_{c1994}$. Our results are robust to dropping this year.

henceforth.

Table 2.3 reports the panel estimates from equation (2.14) with $GunRobberies_t$ as a dependent variable. The effects of the key explanatory factors are in line with our theoretical predictions: Smaller number of gun thefts ($IllegalGuns_{t-1}$), higher associational density ($SocialCapital_{t-1}$), and higher police intensity ($PoliceIntensity_{t-1}$) in period $t-1$ are associated with fewer gun robberies in period t. As can be seen from column (3), these effects are robust to controlling for crime rate, per capita income, and poverty in a given period.[36] The sign of the coefficients of control variables can be well rationalized in terms of our theoretical model. If one were to interpret the crime rate in a given county as a measure for this county's criminal inclination (the inverse of the parameter κ in the model), the negative coefficient of $CrimeRate_t$ is in line with our Proposition 1*(ii)*. The positive coefficient of $Income_t$ is consistent with the positive effect of w on N_a predicted by our model. Lastly, the positive coefficient of $Poverty_t$ suggests that poverty may force citizens into illegal behavior and, potentially, compel them to acquire guns in order to raise the associated booty (parameter $\lambda > 1$ in our model). Controlling for state-year (rather than year) fixed effects in column (4) slightly reduces the size of the coefficients of control variables but leaves the estimates of our key explanatory variables virtually unchanged.

Before introducing the robustness checks from columns (5) and (6), it is worth pausing to briefly discuss the limitations of the UCR data (see Maltz (1999) for a detailed discussion). The main caveat of the UCR panel data is its unbalanced nature. For instance, consecutive observations on *GunRobberies* for the period 1986-2014 are available only for one-third of the counties. The reason for missing values is twofold. First, states may have offense definitions that are incompatible with UCR definitions, leading to data being submitted but not accepted.[37] Second, some law enforcement agencies (LEAs) may withdraw from the UCR program for a certain period of time. If LEAs discontinue reporting to the UCR due to factors related to the explanatory variables, our estimates presented so far may be

[36] Our results are fairly unchanged if we include lagged values of the control variables into the regressions.

[37] For instance, complete data for Illinois have not been included in the UCR since 1985 because the Illinois statutory definition of sexual assault is inconsistent with the UCR definition of rape.

Table 2.3. *Panel estimates: Gun robberies.*

Dep.variable:	OLS					WLS
$GunRobberies_t$	(1)	(2)	(3)	(4)	(5)	(6)
$IllegalGuns_{t-1}$	0.090***	0.030***	0.032***	0.030***	0.028***	0.030***
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$SocialCapital_{t-1}$	-0.042***	-0.058***	-0.057***	-0.071***	-0.048***	-0.051***
	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)	(0.015)
$PoliceIntensity_{t-1}$		-0.084***	-0.090***	-0.092***	-0.073***	-0.068***
		(0.025)	(0.025)	(0.025)	(0.026)	(0.026)
$CrimeRate_t$		0.607***	0.600***	0.579***	0.601***	0.562***
		(0.019)	(0.020)	(0.021)	(0.021)	(0.020)
$Income_t$			0.235***	0.170***	0.245***	0.264***
			(0.052)	(0.056)	(0.062)	(0.062)
$Poverty_t$			0.227***	0.174***	0.101**	0.106**
			(0.038)	(0.041)	(0.044)	(0.044)
County FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	no	no	no
State-year FE	no	no	no	yes	yes	yes
IMR	no	no	no	no	yes	yes
Observations	43,009	42,907	42,773	42,761	40,268	40,268
R-squared	0.766	0.784	0.785	0.800	0.799	0.804

Note: The table reports panel estimates of (variations of) equation (2.14) with $GunRobberies_t$ as a dependent variable. Standard errors in parentheses are clustered at the county level. IMR represents inverse Mills ratios. *, **, *** indicate significance at 1, 5, 10%-level, respectively.

prone to the sample selection bias. To account for this potential bias, one has to bear in mind two possible ways through which non-reporting by LEAs can manifest itself in our county-level analysis. First, if *none* of the LEAs in a given county submits reports to the UCR in a given year, information on gun-related offenses in this county-year is missing. Second, if *some* of the LEAs of a given county fail to submit their reports to the UCR, our measure of gun-related offenses – constructed as the sum of gun-related offenses across all reporting LEAs – understates the actual prevalence of gun violence in this county. We deal with the above-mentioned data limitations by implementing the following two adjustments of our baseline empirical specification.

First, we correct for a potential sample selection bias due to missing county-year observations by testing the sample selection model, cf. Wooldridge (2010). More specifically, for each year in the period 1986-2014, we estimate the following Probit model: $\Pr(y = 1|\boldsymbol{x}) = \Phi(\boldsymbol{x}\psi)$, whereby the binary dependent variable y is equal to one if *Gun-Robberies*$_t$ in a given county is positive and zero otherwise, and $\boldsymbol{x}$ is a vector of controls containing state fixed effects and the following list of county-level variables. To account for the fact that non-reporting to the UCR is most pronounced for smaller and rural counties (see Lynch & Jarvis (2008) and Maltz (1999)), we control for (the logs of) a county's population and per capita income in a given year, as well as the degree of urbanization in 2000-2010.[38] To account for the possibility that missing county-level observations might arise due to a high level of crime in a given period, we further control for the (log of) per capita arrests in a given year, constructed using UCR County-Level Detailed Arrest and Offense Data.[39] From these Probit regressions, we obtain county-year-specific inverse Mills ratios (IMRs), $\hat{\lambda}_{ct}$ and include them, as well as their interaction with year dummies, into our econometric specification from equation (2.14). As can be seen from column (5) of Table 2.3, this robustness check does not materially affect the estimates of our key variables of interest.

Second, to account for potentially endogenous sampling of LEAs and to correct for heteroskedasticity in county-year error terms, we rerun our regressions using weighted least squares (WLS), see Solon, Haider & Wooldridge (2015). More specifically, we exploit UCR information on the number of citizens under the jurisdiction of a given LEA to calculate for each county-year the fraction of the population served by reporting LEAs and use these population shares as weights in the WLS regressions. As can be seen from column (6) of Table 2.3, the WLS estimates of our key explanatory variables remain highly significant and are virtually unchanged in size compared to the OLS coefficients. The estimates from columns (4)-(6) suggest that a one percent decrease in illegal guns, a one percent increase in social capital, or a one percent increase in police intensity in period $t-1$ decreases gun-related robberies in period t by roughly 0.03, 0.05-0.07, and 0.07-0.09 percentage points, respectively.

Next, we rerun the regressions reported in Table 2.3 using $GunHomicides_t$ as a dependent variable. As can be seen from column (3) of Table 2.A.3, all three coefficients of interest are in line with our theoretical predictions and are highly significant, controlling for year and

[38] Yearly estimates of urbanization are, unfortunately, not available.

[39] This data is drawn from https://www.icpsr.umich.edu/icpsrweb/NACJD/studies/35019 and it is available for almost the entire set of counties in the period of 1986-2014.

county fixed effects, as well as a county's crime rate, per capita income, and poverty rate. However, the negative coefficient of $PoliceIntensity_{t-1}$ loses significance after controlling for state-year (rather than year) fixed effects in column (4). In column (5), we correct for the potential sample selection bias by including inverse Mills ratios (as well as their interaction with year dummies) into our specification. Following the approach described above, we obtain these IMRs from the Probit model: $\Pr(y = 1|\boldsymbol{x}) = \Phi(\boldsymbol{x}\psi)$, whereby the binary dependent variable y is equal to one if $GunHomicides_t$ in a given county is positive and zero otherwise, and x is a vector containing state fixed effects and controls for a county's population, per capita income, urbanization, and per capita arrests. The coefficients of $IllegalGuns_{t-1}$ and $SocialCapital_{t-1}$ remain highly robust (both in terms of size and significance) to this sample selection correction, cf. column (5). Moreover, these estimates are virtually unchanged if we rerun our regressions using WLS instead of OLS, cf. column (6) of Table 2.A.3.

In what follows, we conduct further robustness checks of our econometric specification from equation (2.14) using $GunRobberies_t$ as a dependent variable.[40] Recall that our baseline measure of social capital is constructed as the fraction of a county's employment by religious, civic, and social organizations (classified under the NAICS code 813) in the total employment of a given county. In columns (1)-(5) of Table 2.4, we rerun regressions from column (4) of Table 2.3 using alternative measures of *SocialCapital*. In columns (1) and (2), we zoom into this measure by considering the fraction of a county's workforce employed by religious organizations (NAICS code 8131), and by civic and social organizations (NAICS code 8134), respectively.[41] In contrast to the previously used measures, constructed as the *ratio* of associational employment in total employment, the proxy for social capital in column (3) is defined as the *per capita* employment by religious, civic, and social organizations. Instead of *employment*-based proxies utilized so far, columns (4) and (5) consider two *establishment*-based measures: The former is constructed as the ratio

40 We focus henceforth on $GunRobberies_t$ as a dependent variable since it is most suitable to test the predictions of our theoretical model of economic crime. However, all robustness checks yield similar results for $GunHomicides_t$ as an outcome variable.

41 In the period 1986-1997, religious organizations are classified by the CBP under the SIC code 866, while civic and social organizations correspond to the SIC code 864.

of NAICS 813 establishments in the total number of establishments in a given county, while the latter is defined as the per capita number of NAICS 813 establishments in a given county. Regardless of the employed definition, the coefficients of $SocialCapital_{t-1}$ are negative and significant at least at the 5% level. In column (6) of Table 2.4, we utilize an alternative definition of police intensity. Instead of measuring *PoliceIntensity* as the per capita number of police *officers*, this column employs a broader proxy based on the per capita number of police *employees*. The coefficient of $PoliceIntensity_{t-1}$ is negative, highly significant, and comparable in size to the one reported in Table 2.3.

Table 2.4. *Panel estimates: Gun robberies, alternative measures for explanatory variables.*

Dep.variable: $GunRobberies_t$	(1)	(2)	(3)	(4)	(5)	(6)
$SocialCapital_{t-1}$ (empl., religious)	-0.072***					
	(0.020)					
$SocialCapital_{t-1}$ (empl., social&civic)		-0.024**				
		(0.011)				
$SocialCapital_{t-1}$ (empl., per capita)			-0.061***			
			(0.015)			
$SocialCapital_{t-1}$ (est., ratio)				-0.123***		
				(0.033)		
$SocialCapital_{t-1}$ (est., per capita)					-0.138***	
					(0.031)	
$PoliceIntensity_{t-1}$ (employees)						-0.084***
						(0.024)
Observations	40,105	25,278	42,792	43,820	43,820	42,761
R-squared	0.808	0.856	0.800	0.797	0.797	0.800

Note: The table reports panel estimates of equation (2.14) with $GunRobberies_t$ as a dependent variable. All specifications include state-year and county fixed effects, as well as the full set of covariates from Table 2.3. Standard errors in parentheses are clustered at the county level. *, **, *** indicate significance at 1, 5, 10%-level, respectively.

Next, we return to our baseline measures of social capital and police intensity and consider longer lags of the key explanatory variables. As can be seen from Table 2.3, *IllegalGuns*, *SocialCapital*, and *PoliceIntensity* from period $t-3$ continue to have a significant effect on *GunRobberies* in period t. The significance of *IllegalGuns* and *PoliceIntensity* eventually vanishes as one increases the lags to four or five years, yet *SocialCapital* continues to have a significant effect on $GunRobberies_t$ even after five years. The latter finding is line with a large body of literature suggesting a long-lasting impact of social capital on various socio-economic outcomes, cf., e.g., Algan & Cahuc (2010, 2014) and Guiso, Spienza & Zingales (2010).

Table 2.5. *Panel estimates: Gun Robberies, longer lags.*

Dep.variable:	Lags			
$GunRobberies_t$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
$IllegalGuns_{t-n}$	0.025***	0.012**	0.011*	0.008
	(0.005)	(0.006)	(0.006)	(0.006)
$SocialCapital_{t-n}$	-0.039**	-0.053***	-0.038**	-0.033**
	(0.016)	(0.016)	(0.016)	(0.017)
$PoliceIntensity_{t-n}$	-0.085***	-0.055**	-0.030	-0.051*
	(0.025)	(0.026)	(0.027)	(0.026)
Observations	40,895	39,288	37,615	35,827
R-squared	0.803	0.807	0.810	0.814

Note: The table reports panel estimates of equation (2.14) with $GunRobberies_t$ as a dependent variable. $n = 2, ..., 5$ represents the number of lagged periods. All specifications include state-year and county fixed effects, as well as the full set of controls from Table 2.3. Standard errors in parentheses are clustered at the county level. *, **, *** indicate significance at 1, 5, 10%-level, respectively.

In summary, we have established robust relationships between gun-related offenses and lagged values of gun thefts, social capital, and police intensity in line with our theoretical predictions. Do these relationships allow for causal inference? Consider first the effect of social capital. Since it is unlikely that associational density in period $t - n$ ($n = 1, .., 5$) increases in expectation of lower gun robberies in period t, it is reasonable to assert that $SocialCapital_{t-n}$ is exogenous to $GunRobberies_t$. The issue of reverse causality is potentially more pronounced in case of police intensity since employment of police officers in a given year may be driven by the anticipation of higher gun robberies in subsequent years. However, this potential endogeneity would introduce a positive comovement of $PoliceIntensity_{t-n}$ and $GunRobberies_t$, which would work against the predictions of our model. Thus, if we find a strong negative association between $PoliceIntensity_{t-n}$ and $GunRobberies_t$ in our estimates, the true effect of police intensity may be even stronger. Lastly, consider the effect of gun thefts. Regressing *GunRobberies* on the lagged values of *IllegalGuns*, we exclude the possibility that firearm thefts in a given period are merely a byproduct of firearm offenses in this period. However, the evidence presented so far does not yet imply a causal effect of illegal guns since criminals may steal a gun in a given year with an intention to use it at some future time. In other words, $IllegalGuns_{t-n}$ may be endogenous to $GunRobberies_t$. To account for the potential issue of reverse causality, one needs a time-varying measure of illegal guns that is exogenous to gun offenses in a given county and year.

We suggest that firearms stolen in neighboring states are likely to provide this sort of variation. More specifically, to approximate the prevalence of illegal guns in year t and county c from state i, we construct the following alternative county-level measure of

$$IllegalGuns_{ct}^{A} \equiv \ln\left(\sum_j IllegalGuns_t^j \cdot \ell_c^j\right), \tag{2.15}$$

whereby $IllegalGuns_t^j$ is the value of firearms stolen in state $j \neq i$ adjacent (A) to state i, and ℓ_c^j denotes the likelihood that a stolen gun from state j reaches county c.

The idea behind this measure is illustrated in Figure 2.7, using Jefferson county (c) from the Pennsylvania state (PA) as an example. According to tracing reports of the Bureau of Alcohol, Tobacco, Firearms and Explosives, among those guns that were originally purchased in a different state than the one in which they were recovered, the vast majority stems from contiguous states.[42] Hence, county c from state i (PA) is likely to receive a fraction of guns stolen in adjacent states j (in Fig. 2.7: Ohio (OH), West Virginia (WV), Virginia (VA), Maryland (MD), Delaware (DE), New Jersey (NJ), and New York (NY)).

Is this alternative measure of illegal guns exogenous to gun offenses in a given county? Clearly, a criminal from county c may steal a gun from a neighboring state in period $t-1$ to conduct a firearm offense in this county in period t. However, our identifying assumption is that (the mass of criminals from) a single county is too small to drive the variation in gun thefts across *all* adjacent states over time, $\sum_j IllegalGuns_t^j$. We thus assert that $IllegalGuns_{c,t-n}^{A}$ is plausibly exogenous to $GunRobberies_t$. Nevertheless, we conduct a range of robustness checks to preclude possible violations of our identifying assumption.

How do we approximate the likelihood of county c from state i to 'import' a stolen

[42] See, e.g., https://www.atf.gov/about/firearms-trace-data-2015.

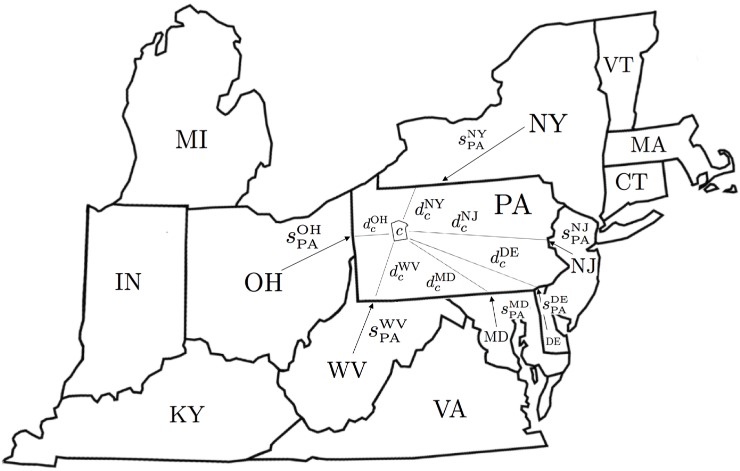


Figure 2.7. *Illegal gun flows from contiguous states to a given county (c).*

gun from a contiguous state j? Since county-level tracing data are, to the best of our knowledge, unavailable, we resort thereby to findings from state-level studies. In a recent contribution, Knight (2013) uses gun tracing data for the year 2009 from the Bureau of Alcohol, Tobacco, Firearms and Explosives (ATF) to estimate the determinants of gun trafficking between states in a gravity-like setting.[43] He finds that distance between a pair of states decreases the likelihood of illegal gun imports from a given source state, with the estimated elasticity of -0.514. For our county-level analysis, we calculate the nearest distance between the borders of county c and state j, d_c^j.[44] Using data from Knight (2013), we further calculate the share of guns s_i^j, originally purchased in state j and recovered in state i, in the nationwide amount of illegal guns traced back to state j. Arguably, a higher s_i^j reflects a higher likelihood of county c from state i to import an illegal gun from state j. Furthermore, Knight (2013) shows that stricter gun regulations in a source state, as

43 For a given destination state, these data report the number of guns recovered in 2009 from crime scenes that were successfully traced to a given source state. Data for other calendar years are, unfortunately, unavailable.

44 Our results are fairly unchanged if we consider distance measures based on centroids (rather than borders).

measured by a unit increase of the Mayors Against Illegal Guns (2010) index (henceforth MAIG), reduce the likelihood of illegal gun 'exports' from this state by an average of -0.102.[45] Based on this information, we construct for each county the following score:

$$\ell_c^j \equiv (\mathrm{MAIG}^j)^{-0.102} \cdot s_i^j \cdot (d_c^j)^{-0.514}.$$

Figure 2.7 illustrates the logic behind this measure, using the afore-mentioned Jefferson county (c). Consider the volume of guns stolen in the Ohio (OH) state in year t, $\mathit{IllegalGuns}_t^{OH}$. These firearms are less likely to be 'exported' to other states the stricter the gun laws in Ohio, MAIG^{OH}. Among those firearms that travel across state borders, fraction $s_{\mathrm{PA}}^{\mathrm{OH}}$ goes to Pennsylvania, on average. Due to the risk associated with transportation of illegal firearms, the amount of guns imported from Ohio is less likely to reach a given county c, the greater distance between this county and OH, d_c^{OH}.[46]

Table 2.6 presents the results of the econometric specification from equation (2.14) with lagged values of gun thefts in the contiguous states, $\mathit{IllegalGuns}_{t-1}^A$ as an additional explanatory variable. As can be seen from column (1), $\mathit{IllegalGuns}_{t-1}^A$ has a positive and highly significant effect on $\mathit{GunRobberies}_t$, controlling for state-year and county fixed effects, as well as $\mathit{CrimeRate}_t$, Income_t, and $\mathit{Poverty}_t$. Adding $\mathit{IllegalGuns}_{t-1}$, $\mathit{SocialCapital}_{t-1}$, and $\mathit{PoliceIntensity}_{t-1}$ in column (2), the coefficient of $\mathit{IllegalGuns}_{t-1}^A$ marginally decreases in size but remains significant at the 5% level. The estimated elasticity of $\mathit{GunRobberies}_t$ with respect to $\mathit{IllegalGuns}_{t-1}^A$ suggests that a one percent increase of gun thefts in adjacent states in the previous period increases a county's gun robberies in the current period by roughly 0.05 percentage points.

Our identification strategy regarding the effect of $\mathit{IllegalGuns}_{t-1}^A$ is built upon the assumption that an individual county is too small to drive the (lagged) variation in gun

[45] This index varies between 0 and 10, whereby each point represents one the following ten gun regulations: 'Straw purchase liability', 'Falsifying purchaser information liability', 'Background check failure liability', 'Gun show checks', 'Required purchaser permit', 'Local discretion to deny carry permits, 'Misdemeanor restrictions', 'Required reporting of lost or stolen guns', 'Local discretion over gun regulations', 'Dealer inspections'.

[46] Our definition of ℓ_c^j does not include gun regulations specific solely to the recipient state since they do not affect the elasticity estimates in our log-log specification. However, we verify that our results are robust to constructing the ℓ_c^j measure based on bilateral differences in gun laws across states.

Table 2.6. *Panel estimates: Gun robberies, illegal guns from adjacent states.*

Dep.variable: $GunRobberies_t$	Full sample (1)	Full sample (2)	Exclude 10% of counties with the largest: *Population* (3)	*CrimeRate* (4)	*Urbanization* (5)	*Income* (6)	Excl. all above (7)
$IllegalGuns^A_{t-1}$	0.059***	0.050**	0.056**	0.046**	0.056***	0.052**	0.056**
	(0.019)	(0.020)	(0.024)	(0.021)	(0.020)	(0.020)	(0.023)
$IllegalGuns_{t-1}$		0.029***	0.025***	0.026***	0.024***	0.029***	0.023***
		(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
$SocialCapital_{t-1}$		-0.072***	-0.067***	-0.067***	-0.067***	-0.068***	-0.057***
		(0.015)	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)
$PoliceIntensity_{t-1}$		-0.086***	-0.071***	-0.075***	-0.064**	-0.090***	-0.062**
		(0.025)	(0.026)	(0.027)	(0.026)	(0.026)	(0.028)
Observations	56,131	42,397	35,346	35,913	35,924	41,595	30,117
R-squared	0.783	0.800	0.750	0.752	0.755	0.802	0.723

Note: The table reports panel estimates of equation (2.14) with $GunRobberies_t$ as a dependent variable. $IllegalGuns^A_{t-1}$ is defined in equation (2.15). All specifications include state-year and county fixed effects, as well as the full set of controls from Table 2.3. Standard errors in parentheses are clustered at the county level. *, **, *** indicate significance at 1, 5, 10%-level, respectively.

thefts across all adjacent states over time. Although this condition is likely to hold in general, one cannot rule out the existence of a few counties that violate our identifying assumption. Arguably, these are populous counties with a high degree of criminal activity, urbanization, and per capita income for which the identifying assumption may not be fulfilled. In columns (3)-(7), we conduct a range of robustness checks to ensure that our results are not driven by those counties. More specifically, in column (3), we exclude the top decile of counties with the largest population.[47] In column (4), we exclude the top decile of counties with the highest *CrimeRate*. In column (5), we exclude the top decile of counties with the largest degree of *Urbanization*. To ensure that a high level of potential booty in a given county does not attract armed criminals from neighboring states, we exclude the top decile of counties with the highest per capita *Income* in column (6). Finally, in column (7), we exclude all of the above. Throughout specifications, the coefficient of $IllegalGuns^A_{t-1}$ remains positive and significant. In summary, the evidence presented above suggests that a higher number of illegal guns in a given period (originating either from a given county or from adjacent states) has a robust positive effect on a county's gun robberies in the subsequent period.

[47] We verify that our results are robust to consideration of alternative thresholds.

Having explored the causes of gun-related offenses, we now rerun our baseline regressions using $\mathit{TotalRobberies}_t$ and $\mathit{TotalHomicides}_t$ as dependent variables. Table 2.7 reports the estimates of equation (2.13) with state-year and county fixed effects, as well as controls for $\mathit{CrimeRate}_t$, Income_t, and $\mathit{Poverty}_t$. In line with our theoretical predictions, $\mathit{IllegalGuns}_{t-1}$ increases while $\mathit{SocialCapital}_{t-1}$ and $\mathit{PoliceIntensity}_{t-1}$ decrease $\mathit{TotalRobberies}_t$. In case of $\mathit{TotalHomicides}_t$, all coefficients have the predicted sign but only $\mathit{IllegalGuns}_{t-1}$ and $\mathit{SocialCapital}_{t-1}$ are significant. Overall, the evidence provides strong support for our theoretical predictions.

Table 2.7. *Panel estimates: Total robberies and total homicides.*

	Dependent variable	
	$\mathit{TotalRobberies}_t$	$\mathit{TotalHomicides}_t$
$\mathit{IllegalGuns}_{t-1}$	0.013***	0.042***
	(0.003)	(0.009)
$\mathit{SocialCapital}_{t-1}$	-0.025**	-0.062**
	(0.012)	(0.029)
$\mathit{PoliceIntensity}_{t-1}$	-0.102***	-0.068
	(0.022)	(0.056)
Observations	52,785	64,556
R-squared	0.850	0.426

Note: The table reports estimates of equation (2.13) with $\mathit{TotalRobberies}_t$ and $\mathit{TotalHomicides}_t$ as dependent variables. All specifications include state-year and county fixed effects, as well as the full set of controls from Table 2.3. Standard errors in parentheses are clustered at the county level. *, **, *** indicate significance at 1, 5, 10%-level, respectively.

2.4 Policy Implications

To formulate policy implications of our work, it is instructive to recall the optimization problem of an unarmed (u) and armed (a) criminal presented, respectively:

$$\max_x E(\pi_u) = (1-\delta)(xcw)^\alpha - \delta p_u x \quad , \quad \max_x E(\pi_a) = (1-\delta)(\lambda xcw)^\alpha - \delta p_a x - g. \quad (2.16)$$

In what follows, we discuss a range of mechanisms that can be applied by policymakers to reduce the number of gun-related, N_a and total offenses, N, given by equations (2.9) and (2.11), respectively.
Consider first the penalty rate p. While in our baseline model p was assumed to be the same for armed and unarmed criminals (cf. equations (2.1) and (2.5)), policymakers can potentially impose larger penalties for an armed crime, $p_a > p_u$, see equation (2.16). In fact, the clause of $p_a > p_u$ is already enshrined in the 18 US Code Â§924(c) of the US federal criminal law: "[...] any person who, during and in relation to any crime of violence or drug trafficking crime [...], uses or carries a firearm, or who, in furtherance of any such crime, possesses a firearm, shall, in addition to the punishment provided for such crime of violence or drug trafficking crime (i) be sentenced to a term of imprisonment of not less than 5 years; (ii) if the firearm is brandished, be sentenced to a term of imprisonment of not less than 7 years; and (iii) if the firearm is discharged, be sentenced to a term of imprisonment of not less than 10 years."[48] Nevertheless, the implementation of this clause constitutes a major challenge for legal authorities since a "few statutes have proven as enigmatic as 18 US Code Â§924(c)", cf. judge Gorsuch (2015). To illustrate the effects of an introduction (and implementation) of a higher gun-related penalty rate, let $p_a \equiv \gamma p_u$, whereby $\gamma > 1$ denotes an increase in the punishment for any given offense due to the fact that a criminal is armed. Figure 2.8 depicts the predicted effect of an increase in γ on the equilibrium sorting of criminals into armed and unarmed activities. A larger cutoff c_a – above which individuals engage in gun-related crime – immediately implies a lower per capita number of gun-related offenses, N_a. It should be noted that an increase in the punishment for a firearm-related crime, γ is not a 'free lunch', since some criminals – those with $c \in (c_a, c'_a)$ in Fig. 2.8 – may either switch from a gun-related to unarmed offenses or substitute guns with another type of weapon (such as knives, brass knuckles, etc.). However, as long the 'threatening effect' of these alternative weapons (parameter λ in our model) is smaller compared to guns, the number of offenses conducted by those criminals will be smaller (cf. equations (2.2) and (2.6)). Hence, the overall number of

[48] See https://www.law.cornell.edu/uscode/text/18/924.

offenses N is expected to decrease in γ. Moreover, given that these alternative weapons are associated with a significantly lower risk of a fatal injury, it is reasonable to assert that the overall number of homicides will decrease due to an increase in the firearm-related punishment γ.

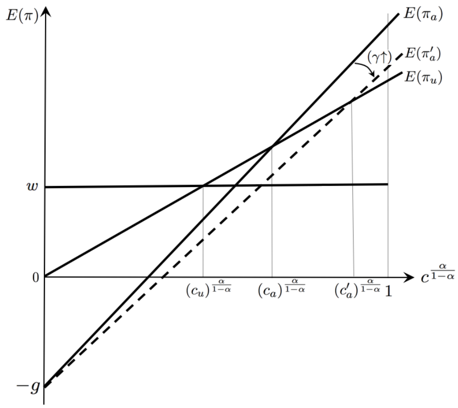


Figure 2.8. *The effect of an increase in gun-related penalties, $\gamma' > \gamma$.*

Second, and related, lawmakers should consider increasing the penalties for possessing and/or carrying an illegal (loaded) gun, even if this gun has not yet been used in furtherance of a crime. From a theoretical perspective, this sanction implies the same effects as an increase in γ, with an additional benefit of the prevention of potential (lethal) crimes. Currently, penalties for an illegal possession of firearms differ widely across states.[49] For instance, possession of a firearm without a permit in the state of New York is punishable by up to one year in prison, a fine of up to $1,000, or both.[50] On the other side of the spectrum, illegal possession of firearms in Arkansas is generally punishable by a fine of up to $500 and a jail sentence of up to 90 days (see Arkansas Statutes Â§5-73). In view of substantial personal and social costs of illegal guns (Cook & Ludwig (2006)), policy-

49 See, e.g., https://www.cga.ct.gov/2012/rpt/2012-R-0345.htm.

50 Possession of a *loaded* firearm without a permit outside of a person's home is punishable by up to 15 years imprisonment, with a mandatory minimum of 3.5 years (see N.Y. Penal Law Â§Â§ 265.01, 265.03, 265.20).

makers are urged to reconsider whether penalties like the latter constitute an appropriate punishment for the possession of an illegal firearm.

Our theoretical model predicts a negative effect of the costs of obtaining a gun (g) on the per capita number of armed (N_a) and total (N) offenses. Using the number of illegal guns as (an inverse) proxy for the costs of obtaining a gun, our empirical analysis provides strong evidence for these predictions. Before formulating recommendations concerning containment of illegal weapons, it is worth pausing to delineate the pervasiveness of illegal guns in the US According to the recent report by the US Department of Justice (Langton (2012)), roughly 1.4 million firearms (an annual average of 232,400) were stolen during burglaries and other property crimes over the period of 2005-2010. At least 80% of these stolen firearms had not been recovered at the time the National Crime Victimization Survey was conducted. Clearly, these numbers only provide a sense of the lower bound of illegal guns in circulation, since a significant fraction of weapons enter the illegal gun market via straw purchasing, falsifying purchaser information, failing to conduct background checks, etc., see Mayors Against Illegal Guns (2010). What can be done to increase a criminal's costs of obtaining an illegal gun, g? First, policymakers can increase g by targeting the major source of illegal weapons – gun traffickers and illegal gun dealers. A negative incentive in the form of higher penalties for the sale and transportation of illegal weapons might be a viable option in this context. Second, one may also consider designing positive incentives (e.g., monetary rewards) for whistle-blowers of illegal gun dealers. This mechanism is likely to decrease trust between sellers and buyers of illegal firearms and, thereby, increase the costs of obtaining an illegal weapon. Third, by tightening the laws on storage of legal weapons, policymakers may prevent some firearms from being stolen and, thereby, reduce the number of illegal guns in circulation. A pioneering policy recently established in the District of Columbia (D.C. Code Ann. Â§7-2507.02(a)) might serve as an example in this context: "[...] each registrant should keep any firearm in his or her possession unloaded and either disassembled or secured by a trigger lock, gun safe, locked box, or other secure device". Fourth, policymakers should consider introducing a nationwide law which would require individual gun owners to report lost or stolen firearms to law enforcement agencies.[51] This law plays a crucial role in combatting straw purchasing since, if a straw buyer is identified through gun tracing, such a requirement would prevent him from evading responsibility by claiming that the crime gun was stolen from him in the first place.

Lastly, according to our model, gun violence can be reduced by decreasing criminal inclinations in a given society. In view of our empirical findings of a robust negative effect of associational density (civic, social and religions organizations) on the prevalence of gun-related offenses, governmental support of associational activism may serve as a tool in combatting gun violence. Yet, a close collaboration between policymakers, sociologists, and criminologists is required in developing further concrete strategies for building social capital. Social programs like *Cure Violence* (*Ceasefire-Chicago*) or *Boston Gun Project* (*Operation Ceasefire*) are suitable case studies in this context.[52] The objective of these programs is to prevent shootings involving youth by changing social norms and 'codes of the street' with the help of social workers specifically trained for this goal. Several evaluations of these projects report statistically significant reductions in gun-related killings and provide anecdotal evidence for the change in gun-related social norms (such as using a gun to settle a dispute) in program sites.[53] Yet, further empirical assessments of these and other programs, as well as further research on the matter of social capital accumulation, is needed to better understand the effect of social capital on gun violence.

2.5 Concluding Comments

We present a simple model of crime in which criminals decide whether to act unarmed or commit firearm-related felonies. This model suggests that gun-related offenses in a given county increase with the number of illegal guns and decrease with social capital and

[51] In 2016, only 10 states and the District of Columbia have such regulations in place, see https://smartgunlaws.org/gun-laws/policy-areas/gun-owner-responsibilities/reporting-lost-or-stolen-firearms/.

[52] See Kennedy, Braga & Piehl (2001), Slutkin, Ransford & Decker (2015), and https://cureviolence.org/.

[53] See Braga, Kennedy, Piehl & Waring (2001), Braga, Kennedy, Waring & Piehl (2001), Braga & Pierce (2004), Butts, Roman, Bostwick & Porter (2015), Delgado, Blount-Hill, Mandala & Butts (2015), Henry, Knoblauch & Sigurvinsdottir (2014), Skogan, Hartnett, Bump & Dubois (2008), Picard-Fritsche & Cerniglia (2013), and Webster, Whitehill, Vernick & Curriero (2012).

police intensity. Combining detailed panel data from the Federal Bureau of Investigation with various socioeconomic variables, we find empirical support for these predictions. To identify the effect of illegal guns, we explore plausibly exogenous variation in illegal gun supplies due to gun thefts in adjacent states. The evidence provided in this paper suggests that illegal guns increase while social capital and police decrease firearm offenses.

To approximate the number of illegal guns, this paper exploits variation in gun thefts. Clearly, a firearm can only be stolen if it was acquired in the first place. Consideration of legal and illegal guns in a unified framework and empirical implementation of its predictions will certainly enhance our understanding of the issue of gun violence. Given that such an investigation would go beyond the scope of the current paper, we relegate it to future research.

Appendix

2.A Appendix Tables

Table 2.A.1. *Summary statistics for main estimation sample.*

Variables	N	mean	sd	min	max
Cross-section:					
GunRobberies	2,264	-2.090	0.938	-4.685	1.508
IllegalGuns	2,264	9.753	1.298	5.345	15.732
SocialCapital	2,264	-3.831	0.553	-8.513	-2.267
PoliceIntensity	2,264	0.677	0.682	-2.501	3.813
Income	2,264	6.841	1.095	1.595	10.865
Poverty	2,264	2.627	0.391	1.043	3.661
IncomeInequality	2,264	-0.838	0.079	-1.082	-0.468
CrimeRate	2,264	3.461	0.568	0.087	5.418
OrganizedCrime	2,264	-6.097	2.063	-6.908	5.902
CriminalNetworks	2,264	-3.042	3.122	-6.908	4.754
EthnicFrac	2,264	-0.412	0.478	-4.044	0.000
RacialFrac	2,264	-1.496	0.792	-4.065	-0.349
AfricanAmerican	2,264	-3.398	1.586	-6.873	-0.171
EducationLevel	2,264	3.538	0.223	2.241	3.996
Urbanization	2,264	2.699	3.341	-6.908	4.605
SingleParent	2,264	-1.192	0.304	-2.873	-0.263
GunHomicides	2,222	-4.263	0.927	-7.136	-0.848
TotalRobberies	2,383	-1.610	1.276	-5.899	2.798
TotalHomicides	2,448	-3.761	0.832	-6.827	-0.489
LegalGuns	1,221	2.072	0.468	-0.429	3.452
Panel:					
GunRobberies	42,761	-2.023	1.121	-5.599	3.168
IllegalGuns	42,761	9.910	1.899	0.000	17.113
SocialCapital	42,761	-3.807	0.438	-9.716	-1.646
PoliceIntensity	42,761	0.498	0.356	-2.520	3.121
CrimeRate	42,761	3.633	0.551	-1.029	6.183
Income	42,761	6.335	1.066	1.264	9.670
Poverty	42,761	2.648	0.431	0.531	4.045
SocialCapital (empl., religious)	40,105	-4.176	0.465	-8.223	-1.743
SocialCapital (empl., social&civic)	25,278	-5.724	0.784	-9.385	-3.076
SocialCapital (empl., per capita)	42,792	1.831	0.560	-3.855	4.255
SocialCapital (est., ratio)	43,820	-3.015	0.370	-5.892	-1.725
SocialCapital (est., per capita)	43,820	0.064	0.387	-2.928	1.723
PoliceIntensity (employees)	42,761	0.847	0.389	-2.146	4.139
GunHomicides	64,556	-5.346	1.864	-6.908	0.010
TotalRobberies	52,785	-1.256	1.158	-4.981	4.211
TotalHomicides	64,556	-4.766	2.011	-6.908	0.262
IllegalGuns^{A}	42,397	11.646	1.282	3.006	16.452

Note: The table reports summary statistics for the main estimation samples used in Tables 2.1, 2.2, 2.3, 2.4, 2.6, 2.7, 2.A.2, and 2.A.3.

Table 2.A.2. *Cross-section estimates: Correlates of gun homicides.*

Dep.variable: *GunHomicides*	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IllegalGuns	0.091***	0.122***	0.152***	0.317***	0.248***	0.242***	0.210***	0.191***
	(0.014)	(0.014)	(0.016)	(0.025)	(0.024)	(0.024)	(0.024)	(0.032)
SocialCapital		-0.198***	-0.198***	-0.108***	-0.091***	-0.098***	-0.077***	-0.006
		(0.032)	(0.032)	(0.029)	(0.027)	(0.027)	(0.027)	(0.045)
PoliceIntensity			-0.094***	-0.027	-0.012	-0.017	-0.008	-0.003
			(0.024)	(0.023)	(0.021)	(0.021)	(0.021)	(0.024)
Income				0.219***	0.457***	0.475***	0.382***	0.137***
				(0.031)	(0.031)	(0.031)	(0.034)	(0.049)
Poverty				0.876***	0.747***	0.695***	0.552***	0.366***
				(0.057)	(0.053)	(0.054)	(0.067)	(0.093)
Inequality				1.095***	0.808***	0.758***	0.734***	0.863**
				(0.248)	(0.228)	(0.228)	(0.247)	(0.353)
CrimeRate					0.191***	0.172***	0.257***	0.445***
					(0.034)	(0.034)	(0.036)	(0.058)
OrganizedCrime					0.136***	0.132***	0.119***	0.075***
					(0.009)	(0.009)	(0.009)	(0.009)
CriminalNetworks					0.061***	0.061***	0.057***	0.055***
					(0.006)	(0.006)	(0.006)	(0.008)
EthnicFrac						-0.037	-0.051	-0.118*
						(0.038)	(0.038)	(0.063)
RacialFrac						-0.067*	-0.053	-0.007
						(0.040)	(0.041)	(0.059)
AfricanAmerican						0.128***	0.114***	0.161***
						(0.020)	(0.020)	(0.029)
Education							0.047	0.220
							(0.106)	(0.145)
Urbanization							-0.042***	-0.068***
							(0.005)	(0.014)
SingleParent							0.168**	0.423***
							(0.067)	(0.115)
LegalGuns								0.270***
								(0.070)
Observations	2,263	2,244	2,244	2,244	2,244	2,222	2,222	1,206
R-squared	0.318	0.337	0.342	0.458	0.547	0.561	0.578	0.722

Note: The table reports estimates of equation (2.13) with *GunHomicides* as a dependent variable. All specifications include state fixed effects. Standard errors are reported in parentheses. *, **, *** indicate significance at 1, 5, 10%-level, respectively.

Table 2.A.3. *Panel estimates: Gun homicides.*

Dep.variable:	OLS					WLS
$GunHomicides_t$	(1)	(2)	(3)	(4)	(5)	(6)
$IllegalGuns_{t-1}$	0.072***	0.044***	0.046***	0.050***	0.048***	0.041***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
$SocialCapital_{t-1}$	-0.160***	-0.181***	-0.182***	-0.078***	-0.075***	-0.079***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.028)	(0.029)
$PoliceIntensity_{t-1}$		-0.134***	-0.138***	-0.047	-0.065	-0.045
		(0.050)	(0.051)	(0.051)	(0.053)	(0.054)
$CrimeRate_t$		0.328***	0.325***	0.355***	0.346***	0.321***
		(0.021)	(0.021)	(0.021)	(0.023)	(0.026)
$Income_t$			0.341***	-0.057	0.270**	0.314***
			(0.083)	(0.088)	(0.105)	(0.105)
$Poverty_t$			0.350***	0.112	0.075	0.094
			(0.066)	(0.076)	(0.080)	(0.080)
County FE	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	no	no	no
State-year FE	no	no	no	yes	yes	yes
IMR	no	no	no	no	yes	yes
Observations	65,806	64,748	64,574	64,556	62,499	62,499
R-squared	0.389	0.393	0.394	0.425	0.419	0.423

Note: The table reports panel estimates of equation (2.14) with $GunHomicides_t$ as a dependent variable. Standard errors in parentheses are clustered at the county level. IMR represents inverse Mills ratios. *, **, *** indicate significance at 1, 5, 10%-level, respectively.

2.B Mathematical Appendix

Proof of Proposition 1

Consider first the proof of parts *(i)* and *(iii)* of Proposition 1. Taking the first-order derivative of (2.9) with respect to g yields $N_a'(g) = -c_a'(g)x_a(c_a)f(c_a) < 0$, whereby the sign of this derivative follows immediately from the fact that $c_a'(g) > 0$. Similarly, differentiating (2.9) with respect to δ, we obtain $N_a'(\delta) = -c_a'(\delta)x_a(c_a)f(c_a) + \int_{c_a}^{1} x_a'(\delta)f(c)dc < 0$, whereby the sign of this derivative results from $c_a'(\delta) > 0$ and $x_a'(\delta) < 0$, cf. equations (2.8) and (2.9).

Consider next the proof of Proposition 1*(ii)*. Plugging the density associated with the cumulative distribution function from equation (2.10) in (2.9) and integrating the resulting expression, we obtain

$$N_a = \frac{\kappa c_{min}^{\kappa}}{1 - c_{min}^{\kappa}} \frac{1-\alpha}{\alpha - \kappa(1-\alpha)} \left(1 - (c_a)^{\frac{\alpha-\kappa(1-\alpha)}{1-\alpha}}\right) (\lambda w)^{\frac{\alpha}{1-\alpha}} \left(\frac{1-\delta}{\delta}\frac{\alpha}{p}\right)^{\frac{1}{1-\alpha}}. \tag{2.5.17}$$

Differentiating N_a from equation (2.5.17) with respect to κ yields after simplification:

$$N_a'(\kappa) = -\frac{c_{min}^{\kappa}(1-\alpha)(\lambda w)^{\frac{\alpha}{1-\alpha}} \left(\frac{1-\delta}{\delta}\frac{\alpha}{p}\right)^{\frac{1}{1-\alpha}}}{((1+\kappa)\alpha - \kappa)^2(1-c_{min}^{\kappa})^2} \cdot X,$$

whereby

$$X \equiv (c_a)^{\frac{\alpha-\kappa(1-\alpha)}{1-\alpha}} \left(\kappa(\alpha - \kappa(1-\alpha))(\ln(c_{min}) - (1-c_{min}^{\kappa})\ln(c_a)) + \alpha(1-c_{min}^{\kappa})\right) - \alpha(1-c_{min}^{\kappa})$$
$$- \ln(c_{min})\kappa(\alpha - \kappa(1-\alpha)).$$

Note that $N_a'(\kappa) \leq 0$ if and only if $X \geq 0$. To assess the sign of X, we take the first-order derivative of X with respect to c_{min} and obtain $X'(c_{min}) = -\frac{\kappa}{c_{min}} \cdot Y$, whereby

$$Y \equiv \alpha - \kappa(1-\alpha) - \alpha c_{min}^{\kappa} - (c_a)^{\frac{\alpha-\kappa(1-\alpha)}{1-\alpha}} \left(c_{min}^{\kappa}((\alpha - \kappa(1-\alpha))\kappa \ln(c_a) - \alpha) + \alpha - \kappa(1-\alpha)\right).$$

To show that $Y \geq 0$, we take the first-order derivative of Y with respect to c_a:

$$\frac{\partial Y}{\partial c_a} = -\frac{(\alpha-\kappa(1-\alpha))^2}{1-\alpha}(c_a)^{\frac{\alpha-\kappa(1-\alpha)}{1-\alpha}-1} \cdot Z \quad , \quad Z \equiv 1 - c_{min}^{\kappa}(1-\kappa\ln(c_a)).$$

Note that Z is (weakly) decreasing in c_{min} for all $c_a \in [0,1]$. That is, if $Z \geq 0$ for the highest possible $c_{min} = c_a$, $Z \geq 0$ holds a fortiori for all $c_{min} \leq c_a$. Evaluating Z at $c_{min} = c_a$ yields $Z|_{c_{min}=c_a} = 1 - c_a^{\kappa}(1-\kappa\ln(c_a))$. Given that $\frac{\partial Z|_{c_{min}=c_a}}{\partial c_a} = \kappa^2 c_a^{\kappa-1}\ln(c_a) < 0$, if $Z|_{c_{min}=c_a} \geq 0$ for the highest possible $c_a = 1$, $Z|_{c_{min}=c_a} \geq 0$ holds a fortiori for all $c_a \leq 1$. Evaluating $Z|_{c_{min}=c_a}$ at $c_a = 1$ yields $Z|_{c_{min}=c_a=1} = 0$. Since $Z \geq 0$ for all permissible parameter values, we have $Y'(c_a) \leq 0$. Hence, if $Y \geq 0$ for the highest possible $c_a = 1$, we have $Y \geq 0$ for all $c_a \leq 1$. Evaluating Y at $c_a = 1$ yields $Y|_{c_a=1} = 0$. Since $Y \geq 0$ for all permissible parameter values, we have $X'(c_{min}) \leq 0$. Hence, if $X \geq 0$ for the highest possible $c_{min} = 1$, $X \geq 0$ holds a fortiori for all $c_{min} \leq 1$. Evaluating X at $c_{min} = 1$ yields $X|_{c_{min}=1} = 0$. We thus have shown that $N_a'(\kappa) \leq 0$ for all parameter values, whereby the sign of this first-order derivative is strict (rather than weak) if $c_{min} < c_a < 1$. This completes the proof of Proposition 1*(ii)*.

Next, we analyze the effect of c_{min} on the per capita number of armed offenses. Differentiating N_a from equation (2.5.17) with respect to c_{min} yields:

$$N_a'(c_{min}) = \frac{\kappa^2 c_{min}^{\kappa-1}}{(1-c_{min}^{\kappa})^2}(\lambda w)^{\frac{\alpha}{1-\alpha}}\left(\frac{1-\delta}{\delta}\frac{\alpha}{p}\right)^{\frac{1}{1-\alpha}} \cdot \Omega, \quad \text{where} \quad \Omega \equiv \frac{1-\alpha}{\alpha-\kappa(1-\alpha)}\left(1-(c_a)^{\frac{\alpha-\kappa(1-\alpha)}{1-\alpha}}\right).$$

Note that the sign of $N_a'(c_{min})$ is determined by the sign of Ω. If $\alpha-\kappa(1-\alpha) > 0$, we have $\Omega > 0$, since $\frac{1-\alpha}{\alpha-\kappa(1-\alpha)} > 0$ and $(c_a)^{\frac{\alpha-\kappa(1-\alpha)}{1-\alpha}} < 1$. Conversely, if $\alpha-\kappa(1-\alpha) < 0$, we have $\Omega > 0$, since $\frac{1-\alpha}{\alpha-\kappa(1-\alpha)} < 0$ and $(c_a)^{\frac{\alpha-\kappa(1-\alpha)}{1-\alpha}} > 1$.[54] We thus have established $N_a'(c_{min}) > 0$.

Differentiating N_a from equation (2.9) with respect to p, we obtain $N_a'(p) = -c_a'(p)x_a(c_a)f(c_a) + \int_{c_a}^1 x_a'(p)f(c)dc < 0$, whereby the sign of this derivative results from $c_a'(p) > 0$ and $x_a'(p) < 0$, cf. equations (2.8) and (2.9). Similarly, taking the first-order derivative of N_a with respect to w yields $N_a'(w) = -c_a'(w)x_a(c_a)f(c_a) + \int_{c_a}^1 x_a'(w)f(c)dc > 0$, whereby

[54] For the 'knife-edge' case of $\alpha-\kappa(1-\alpha) = 0$, the sign of $N_a'(c_{min})$ is undetermined, cf. also N_a from eq. (2.5.17).

the sign of this derivative results from $c_a'(w) < 0$ and $x_a'(w) > 0$, cf. equations (2.8) and (2.9).

Proof of Proposition 2

Differentiating N from equation (2.11) with respect to g yields $N'(g) = -c_a'(g)f(c_a)[x_a(c_a)-x_u(c_a)] < 0$, whereby the sign of this derivative follows from the fact that $c_a'(g) > 0$, see equation (2.8), and $x_a(c) > x_u(c)$ for any given c, cf. equations (2.2) and (2.6). This implies Proposition 2*(i)*. To prove Proposition 2*(iii)*, we differentiate N from equation (2.11) with respect to δ and obtain:

$$N'(\delta) = \int_{c_u}^{c_a} x_u'(\delta)f(c)dc + \int_{c_a}^{1} x_a'(\delta)f(c)dc - c_u'(\delta)x_u(c_u)f(c_u) - c_a'(g)f(c_a)[x_a(c_a)-x_u(c_a)] < 0,$$

whereby the sign of this derivative results from $x_u'(\delta) < 0$, $x_a'(\delta) < 0$, $c_u'(\delta) > 0$, $c_a'(\delta) > 0$, and $x_a(c) > x_u(c)$ for any given c, cf. equations (2.2), (2.6), and (2.8).

Using the definition of Pareto distribution from equation (2.10), the per capita number of total offenses can be expressed as:

$$N = \frac{\kappa c_{min}^{\kappa}}{1-c_{min}^{\kappa}} \frac{1-\alpha}{\alpha-\kappa(1-\alpha)} \left(\lambda^{\frac{\alpha}{1-\alpha}} - (c_a)^{\frac{\alpha-\kappa(1-\alpha)}{1-\alpha}}(\lambda^{\frac{\alpha}{1-\alpha}}-1) - (c_u)^{\frac{\alpha-\kappa(1-\alpha)}{1-\alpha}}\right) w^{\frac{\alpha}{1-\alpha}} \left(\frac{1-\delta}{\delta}\frac{\alpha}{p}\right)^{\frac{1}{1-\alpha}},$$

whereby c_u and c_a are given by equation (2.8). Following the approach described in Appendix 2.5, we prove that $N'(\kappa) < 0$. This implies Proposition 2*(ii)* and completes the proof of Proposition 2.

2.C Distribution of Criminal Activities in the US

To draw assumptions about the behavior and functional form of $f(c)$, we use incident-level data from the National Incident-Based Reporting System (NIBRS) by the UCR.[55] More specifically, we exploit the Property Segment of this data which contains information on the dollar value of property stolen in a given incident. In the most recent year available, 2014, the UCR recorded 3,766,167 incidents of property theft in the US, with the minimum value of \$0, maximum value of \$100,000,350 and the mean of \$1,154. Figure 2.C.1 depicts the density of incidents with stolen property worth less than \$10,000. Apart from the 'spikes' clustered around the round numbers of 500, 1000, 1500, etc., the density in this range appears to be non-increasing in its support.[56]

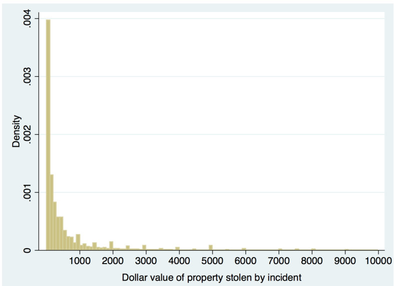


Figure 2.C.1. *Histogram of the value of property stolen by incident in the US in 2014.*

In the following, we show that the actual density of incidents of property theft in the US can be approximated by a Pareto distribution. For a discrete Pareto-distributed random variable, X, the tail distribution (survival) function is given by

$$\Pr[X \geq x] = \left(\frac{x_{min}}{x}\right)^{\kappa} \quad , \quad x \geq x_{min}, \ \kappa > 0,$$

where x_{min} is the lower bound of the support and κ is the shape parameter of this function. If X is indeed distributed Pareto, the relationship between the frequency of theft and the value of stolen property in log-log coordinates should be linear, with the slope equal to

[55] These data are publicly available at https://www.icpsr.umich.edu/icpsrweb/NACJD/series/128.

[56] These spikes can be attributed to the rounding errors in cases of the unknown true value of the stolen property.

$-\kappa$.[57] To assess this relationship, we tabulate the data in fourteen successive bins, having the width increasing by one unit on the logarithmic scale.[58] Figure 2.C.2 plots the log frequency of incidents within each bin against the logarithmized mean dollar value of those incidents. The red line depicts the fitted linear relationship between these variables and the associated OLS results are presented in the top right corner. A high linear fit ($R^2 = 0.979$) suggests that the actual distribution of US crime can be well approximated with a Pareto distribution with a shape parameter of $\kappa = 1.171$.

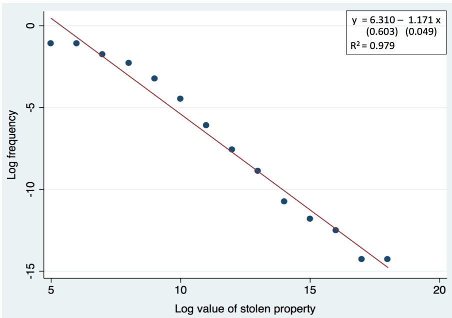


Figure 2.C.2. *Binned distribution of the dollar value of property stolen in the US in 2014.*

We repeat this exercise for all US states available in the NIBRS database (see Table 2.A.4), as well as individual counties (Table 2.A.5 presents exemplary the results for the state of Massachusetts), whereby N, R^2, and κ represent the sample size, linear fit, and the shape parameter, respectively. Generally high R^2 suggest that the Pareto distribution provides a good fit to the actual distribution of criminal activities across US states and counties. Moreover, notice from Table 2.A.5 that the dispersion of criminal activities (as measured by the parameter κ) varies substantially across counties that belong to the same state, despite the shared state-specific criminal law. In the main text, we attribute this variation to differences in social capital.

57 To see this, note from the definition of the Pareto distribution that $d \log f(x)/d \log x = -\kappa$.

58 In our benchmark analysis, we do not consider incidents of stolen property worth less than \$200 (i.e., set $x_{min} \equiv 200$). In most US states, these incidents are classified as misdemeanors or “petty theft” and the associated data entries in this range are likely to be subject to the above-mentioned measurement errors.

Table 2.A.4. *Distribution of criminal activities across US states.*

State	N	R^2	κ	State	N	R^2	κ
Alabama	3,426	0.953	0.778	Montana	38,646	0.938	0.984
Arizona	16,286	0.928	1.063	Nebraska	18,253	0.950	1.034
Arkansas	140,359	0.957	1.158	New Hampshire	38,268	0.960	0.911
Colorado	221,778	0.937	0.980	North Dakota	23,256	0.919	0.996
Connecticut	62,364	0.951	0.964	Ohio	373,845	0.976	1.113
Delaware	49,868	0.946	1.059	Oklahoma	45,230	0.910	1.016
DC	1,756	0.941	1.175	Oregon	81,955	0.955	1.016
Idaho	46,434	0.920	0.816	Pennsylvania	3,055	0.953	0.778
Illinois	12,162	0.949	0.977	Rhode Island	33,265	0.934	1.001
Indiana	2,031	0.858	0.678	South Carolina	267,509	0.955	1.143
Iowa	91,573	0.939	1.057	South Dakota	20,124	0.952	0.884
Kansas	104,144	0.953	1.063	Tennessee	365,586	0.966	1.139
Kentucky	133,787	0.957	1.039	Texas	171,967	0.945	1.058
Louisiana	21,863	0.935	0.996	Utah	114,265	0.950	1.038
Maine	10,670	0.934	1.029	Vermont	12,445	0.942	1.073
Massachusetts	144,788	0.965	1.011	Virginia	281,899	0.968	1.121
Michigan	301,047	0.942	1.166	Washington	306,601	0.930	1.169
Mississippi	11,310	0.943	0.943	West Virginia	42,090	0.965	1.001
Missouri	55,817	0.912	1.112	Wisconsin	94,442	0.933	1.138

Table 2.A.5. *Distribution of criminal activities across counties in Massachusetts.*

County	N	R^2	κ	County	N	R^2	κ
Barnstable	6,048	0.928	0.884	Hampshire	3,322	0.937	0.862
Berkshire	3,237	0.966	0.797	Middlesex	30,114	0.925	0.901
Bristol	16,774	0.950	0.958	Nantucket	384	0.965	0.583
Dukes	156	0.962	0.533	Norfolk	11,511	0.962	0.888
Essex	15,827	0.944	0.965	Plymouth	10,115	0.934	0.820
Franklin	1,530	0.934	0.931	Suffolk	4,381	0.898	0.974
Hampden	18,550	0.942	1.057	Worcester	22,476	0.944	1.035

2.D Data Appendix

All crime-related measures in our paper are constructed using Uniform Crime Reporting (UCR) data by the Federal Bureau of Investigation (FBI).[59] This data is available at the level of law enforcement agencies (LEAs). We map all LEAs to US counties using the 2012 Law Enforcement Agency Identifiers Crosswalk by the US Department of Justice.[60] In the following, we detail the construction of the main dependent and independent variables obtained from the UCR data.

GunHomicides and ***TotalHomicides*** are constructed using the UCR's Supplementary

[59] This data is publicly available at https://icpsr.umich.edu/icpsrweb/NACJD/series/57.

[60] This crosswalk is available at https://www.icpsr.umich.edu/icpsrweb/NACJD/series/00366. In our baseline analysis, we drop observations from Alaska and Hawaii.

Homicide Reports (SHR) database. The SHR lists all known homicide incidents that took place in a given year on the area monitored by a given LEA and provides information on the circumstance under which a given homicide was committed. During the construction of our baseline measure of (gun-caused) homicides, we excluded the following list of circumstances indicating an accident, negligence, or killing of the (suspected) felon: 'victim shot in hunting accident', 'gun-cleaning death - other than self', 'children playing with gun', 'other negligent handling of gun', 'all other manslaughter by negligence', 'felon killed by police', 'felon killed by private citizen', 'all suspected felony type'. We further excluded rare circumstances that are hard to rationalize with our theoretical model, such as 'child killed by babysitter', 'institutional killings', 'sniper attack', and 'abortion'. Using SHR information on the type of the offender's weapon, we identify all homicides that were committed by one of the following firearm types: 'handgun – pistol, revolver', 'rifle', 'shotgun', 'firearm, type not stated', and 'other gun'. We then calculate the yearly sum of gun-caused and total (i.e., gun-caused and gun-unrelated) homicides by LEA and aggregate this information to county-level data using the above-mentioned LEA Identifiers Crosswalk. Using US Census annual county-level population data, we construct our proxies for the per capita number of *GunHomicides* and *TotalHomicides*.

GunRobberies and ***TotalRobberies***. These measures are constructed using the UCR's Offenses Known and Clearances by Arrest (OKCA) database, which reports, among other things, the 'actual number of gun robberies' (ACT NUM GUN ROBBERY) and the 'actual number of total robberies' (ACT NUM ROBBERY TOTL). Both variables are reported at the LEA-level on a monthly basis. The challenge behind aggregating this information to county-level annual measures lies in the fact that OKCA codifies both zero and missing values as "0". Following the methodology delineated in the UCR codebooks, we distinguish missing (gun-related) robberies from "true" zeroes using information on 'grand total of all actual offenses' (ACT # ALL FIELDS). This process involves several steps: First, we exploit information on the *latest* month reported in the yearly return (NUMBER OF MONTHS REPORTED), to replace zero values in the ensuing months with missing values. However, information on the latest reported month (say, November) does *not* necessarily imply that all preceding (eleven) months are included in the reports (see UCR codebooks). To identify missing observations in the preceding months, we calculate in the second step the average monthly number of grand total offenses in a given LEA/year and replace this LEA's "0"-values as missing if the monthly average lies above a certain threshold. During the construction of our baseline measures, we set this threshold equal to 15. That is, if the average monthly number of grand total offenses in a given LEA is larger than fifteen, we treat this LEA's zero values as missing.[61] Third, we identify LEAs that report offenses quarterly, semiannually, or annually and replace "0"-values in the non-reporting months as missing. Having distinguished missing values from true zeroes in grand total offenses, we replace all zero monthly values in gun-related and total robberies with missing values if a LEA's grand total offenses in the respective month is missing. Missing monthly values in gun-related and total robberies are then replaced by the averages in the respective category across all months reported by a LEA in a given year. Monthly gun-related and total robberies are summed up to annual LEA-level data, which, in turn, is aggregated to annual county-level data using the LEA Identifiers Crosswalk. Using yearly population data from the US Census, we construct our proxies for the per capita number of *GunRobberies* and *TotalRobberies* in a given county-year.

IllegalGuns. Our proxy for the number of illegal guns is constructed using the UCR's Property Stolen and Recovered (PSR) database, which reports, among other things, the value of firearms stolen in a given month in the area monitored by a given LEA. In the raw PSR data, both zero and missing values are coded as "0". However, the PSR database contains twelve dummy variables (STATUS) which indicate whether information in a given month was reported or not. Missing monthly values in stolen firearm value are replaced by the average value of stolen firearms in a given year and the annual LEA-level data is aggregated to the county level using the LEA Identifiers Crosswalk.

PoliceIntensity. Our measures of police intensity are constructed using the UCR's Law Enforcement Officers Killed or Assaulted (LEOKA) database. For each LEA/year, the LEOKA reports the number of police officers per 1,000 population (OFFICER RATE

[61] Since the above-mentioned threshold is chosen arbitrarily, we run a wide range of unreported robustness checks to ensure robust to considering alternative thresholds.

PER 1,000 POP) and the number of police employees per 1,000 population (EMPLOYEE RATE PER 1,000 POP).[62] To construct our baseline measure of police intensity, we calculate for each year the weighted average of the police officers rate across all LEAs of a given county with weights being the fraction of a county's population served by a given LEA. In the robustness checks, we consider the weighted average rate of police employees as an alternative proxy for police intensity in a given county.

[62] In years 1981-82, 1985-89, and 1995-96, these rates are reported per 10,000 population. The reported values of police officers and employees in those years are multiplied by 10 for consistency. Due to the fact that the reported values of police officers and employees in 1990 are exceptionally high (oftentimes exceeding the preceding years by the factor of thirty) we replace these values in 1990 by an average of the years 1989 and 1991.

Bibliography

Akçomak, S. & ter Weel, B. (2012). The impact of social capital on crime: Evidence from the Netherlands. *Regional Science and Urban Economics*, *42*, 323–340.

Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic Growth*, *8*(2), 155–194.

Algan, Y. & Cahuc, P. (2010). Inherited Trust and Growth. *American Economic Review*, *100*(5), 2060–2092.

Algan, Y. & Cahuc, P. (2014). Trust, Growth and Happiness: New evidence and policy implications. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of economic growth* (Vol. 1, pp. 49–120). North-Holland, Amsterdam.

Arbatli, C. E., Ashraf, Q. H., & Galor, O. (2015). The Nature of Conflict.

Ayres, I. & Donohue, J. J. (2003). Shooting down the "More Guns, Less Crime" Hypothesis. *Stanford Law Review*, *55*, 1193–1312.

Azrael, D., Cook, P. J., & Miller, M. (2004). State and Local Prevalence of Firearms Ownership Measurement, Structure, and Trends. *Journal of Quantitative Criminology*, *20*(1), 43–62.

Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, *76*, 169–217.

Bénabou, R. & Tirole, J. (2006). Incentives and Prosocial Behavior. *American Economic Review*, *96*(5), 1652–1678.

Bénabou, R. & Tirole, J. (2011). *Laws and Norms.*

Braga, A. A., Kennedy, D. M., Piehl, A. M., & Waring, E. J. (2001). Measuring the impact of Operation Ceasefire. In *Reducing gun violence: The boston gun project's operation ceasefire.* Washington, DC: National Institute of Justice, U.S. Department of Justice.

Braga, A. A., Kennedy, D. M., Waring, E. J., & Piehl, A. M. (2001). Problem-oriented policing, deterrence, and youth violence: An evaluation of Boston's Operation Ceasefire. *Journal of Research in Crime and Delinquency*, *38*(195-225).

Braga, A. A. & Pierce, G. L. (2004). Disrupting illegal firearms markets in Boston: The effects of Operation Ceasefire on the supply of new handguns to criminals. *Journal of Forensic Sciences*, *46*(4), 701–706.

Bronars, S. & Lott, J. R. (1998). Criminal Deterrence, Geographic Spillovers, and the Right to Carry Concealed Handguns. *American Economic Review: Papers & Proceedings*, *88*(2), 475–479.

Buonanno, P., Montolio, D., & Vanin, P. (2009). Does Social Capital Reduce Crime? *Journal of Law and Economics*, *52*(1), 145–170.

Butts, J. A., Roman, C. G., Bostwick, L., & Porter, J. R. (2015). Cure Violence: A Public Health Model to Reduce Gun Violence. *Annual Review of Public Health*, *36*, 39–53.

Coleman, J. S. (1988). Social Capital in the Creation of Human Capital. *American Journal of Sociology*, *94*, 95–120.

Coleman, J. S. & Coleman, J. S. (1994). *Foundations of Social Theory*. Harvard University Press.

Cook, P. & Ludwig, J. (2006). The social costs of gun ownership. *Journal of Public Economics*, *90*, 379–391.

Cook, P., Ludwig, J., Venkatesh, S., & Braga, A. A. (2007). Underground Gun Markets. *Economic Journal*, *117*, 588–618.

Cook, P., Parker, S. T., & Pollack, H. A. (2015). Sources of guns to dangerous people: What we learn by asking them. *Preventive Medicine*, *79*, 28–36.

Delgado, S. A., Blount-Hill, K.-L., Mandala, M., & Butts, J. A. (2015). *Perceptions of violence: Surveying young males in New York City*.

Di Tella, R. & Schargrodsky, E. (2004). Do police reduce crime? Estimates using the allocation of police forces after a terrorist attack. *American Economic Review*, *94*(1), 115–133.

Draca, M. & Machin, S. (2015). Crime and Economic Incentives. *Annual Review of Economics*, *7*, 389–408.

Draca, M., Machin, S., & Witt, R. (2011). Panic on the streets of London: Police, crime, and the July 2005 terror attacks. *American Economic Review*, *101*(5), 2157–2181.

Duggan, M., Hjalmarsson, R., & Jacob, B. A. (2011). The short-term and localized effects of gun shows: evidence from California and Texas. *Review of Economics and Statistics*, *93*(3), 786–799.

Fearon, J. D. (2003). Ethnic and Cultural Diversity by Country. *Journal of Economic Growth*, *8*(2), 195–222.

Fleegler, E., Lee, L., Monuteaux, M., Hemenway, D., & Mannix, R. (2013). Firearm legislation and firearm-related fatalities in the United States. *JAMA Internal Medicine*, *173*(9), 732–740.

Freeman, R. B. (1999). The Economics of Crime. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics,* (Vol. 3). North-Holland, Amsterdam.

Funk, P. (2006). Governmental Action, Social Norms, and Criminal Behavior. *Journal of Institutional and Theoretical Economics*, *161*(3), 522–535.

Galea, S., Karpati, A., & Kennedy, B. (2002). Social capital and violence in the United States, 1974–1993. *Social science & medicine*, *55*, 1373–1383.

Gius, M. (2013). An examination of the effects of concealed weapons laws and assault weapons bans on state-level murder rates. *Applied Economics Letters*, *21*(4), 265–267.

Glaeser, E. L. & Sacerdote, B. (1999). Why Is There More Crime in Cities? *Journal of Political Economy*, *107*(6), 225–258.

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and Social Interactions. *Quarterly Journal of Economics*, *111*(2), 507–548.

Gorsuch, N. M. (2015). United States v. Rentz.

Guiso, L., Spienza, P., & Zingales, L. (2010). Civic Capital as the Missing Link. In J. Benhabib, A. Bisin, & M. Jackson (Eds.), *Handbook of social economics*. North-Holland, Amsterdam.

Henry, D. B., Knoblauch, S., & Sigurvinsdottir, R. (2014). *The effect of intensive CeaseFire intervention on crime in four Chicago police beats: Quantitative assessment*.

Hirschi, T. (1969). *Causes of Delinquency*. University of California Press, Berkley, CA.

Hirschi, T. (1986). On the Compatibility of Rational Choice and Social Control Theories of Crime. In D. Cornish & R. Clarke (Eds.), *The reasoning criminal: Rational choice perspectives on offending*. New York: Springer-Verlag.

Kalesan, B., Mobily, M. E., Keiser, O., Fagan, J. A., & Galea, S. (2016). Firearm legislation and firearm mortality in the USA: A cross-sectional, state-level study. *The Lancet*, *forthcomin*.

Kennedy, B. P., Kawachi, I., Prothrow-Stith, D., Lochner, K., & Gupta, V. (1998). Social capital, income inequality, and firearm violent crime. *Social science & medicine*, *47*(1), 7–17.

Kennedy, D. M., Braga, A. A., & Piehl, A. M. (2001). Developing and implementing Operation Ceasefire. In *Reducing gun violence: The boston gun project's operation ceasefire*. Washington, DC: National Institute of Justice, U.S. Department of Justice.

Klick, J. & Tabarrok, A. (2005). Using Terror Alert Levels to Estimate the Effect of Police on Crime. *Journal of Law and Economics*, *48*(1), 267–279.

Knight, B. (2013). State Gun Policy and Cross-State Externalities: Evidence from Crime Gun Tracing. *American Economic Journal: Economic Policy*, *5*(4), 200–229.

Kukharskyy, B. & Seiffert, S. (2016). *Live by the gun, die by the gun*.

Kukharskyy, B. & Seiffert, S. (2017). *Gun Violence in the U.S. : Correlates and Causes*, University of Hohenheim, Institute of Economics.

Langton, L. (2012). *Firearms Stolen during Household Burglaries and Other Property Crimes, 2005–2010*.

Lanza, S. P. (2014). The effect of firearm restrictions on gun-related homicides across US states. *Applied Economics Letters*.

Lederman, D., Loayza, N., & Menendez, A. M. (2002). Violent Crime: Does Social Capital Matter? *Economic Development and Cultural Change*, *50*(3), 509–39.

Levitt, S. D. (1997). Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime. *American Economic Review*, *87*(3), 270–290.

Lott, J. R. & Mustard, D. B. (1997). Crime, Deterrence, and Right-to-carry concealed handguns. *Journal of Legal Studies*, *26*, 1–68.

Lynch, J. P. & Jarvis, J. P. (2008). Missing Data and Imputation in the Uniform Crime Reports and the Effects on National Estimates. *Journal of Contemporary Criminal Justice*, *24*(1), 69–85.

Maltz, M. D. (1999). *Bridging Gaps in Police Crime Data*.

Mayors Against Illegal Guns. (2010). Trace the Guns: The Link Between Gun Laws and Interstate Gun Trafficking. Available at: www.MayorsAgainstIllegalGuns.org.

McAdams, R. H. & Rasmusen, E. B. (2007). Norms in Law and Economics. In A. M. Polinsky & S. Shavell (Eds.), *Handbook of law and economics*. North-Holland, Amsterdam.

Messner, S. F., Baumer, E. P., & Rosenfeld, R. (2004). Dimensions of Social Capital and Rates of Criminal Homicide. *American Sociological Review*, *69*, 882–903.

Nagin, D. S. (2013). Deterrence: A Review of the Evidence by a Criminologist for Economists. *Annual Review of Economics*, *5*, 83–105.

Nunn, N. & Wantchekon, L. (2011). The Slave Trade and the Origins of Mistrust in Africa. *American Economic Review*, *101*(7), 3221–3252.

Picard-Fritsche, S. & Cerniglia, L. (2013). *Testing a public health approach to gun violence: An evaluation of Crown Heights Save Our Streets, a replication of the CureViolence model*.

Planty, M. & Truman, J. L. (2013). *Firearm violence, 1993-2011*.

Poutvaara, P. & Priks, M. (2006). *Hooliganism in the shadow of a terrorist attack and the tsunami: Do police reduce group violence?*

Putnam, R. (1995). Bowling alone: America's declining social capital. *Journal of Democracy*, *6*(1), 65–78.

Putnam, R. (2000). *Bowling alone: The collapse and revival of American community*. Simon&Schuster, NY.

Putnam, R. D., Leonardi, R., & Nanetti, R. Y. (1994). *Making Democracy Work: Civic Traditions in Modern Italy*. Princeton University Press.

Rosenfeld, R., Messner, S. F., & Baumer, E. P. (2001). Social Capital and Homicide. *Social Forces*, *80*(1), 283–310.

Saegert, S. & Winkel, G. (2004). Crime, Social Capital, and Community Participation. *American Journal of Community Psychology*, *34*(3/4), 219–233.

Siegel, M., Ross, C. S., & King, C. (2013). The relationship between gun ownership and firearm homicide rates in the United States, 1981-2010. *American Journal of Public Health*, *103*(11), 2098–105.

Siegel, M., Ross, C. S., & King, C. (2014a). A new proxy measure for state-level gun ownership in studies of firearm injury prevention. *Injury Prevention*, *20*(3), 204–207.

Siegel, M., Ross, C. S., & King, C. (2014b). Examining the relationship between the prevalence of guns and homicide rates in the USA using a new and improved state-level gun ownership proxy. *Injury Prevention*, *20*(6), 424–426.

Skogan, W. G., Hartnett, S. M., Bump, N., & Dubois, J. (2008). *Evaluation of CeaseFire-Chicago.*

Slutkin, G., Ransford, C., & Decker, B. R. (2015). Cure Violence: Treating Violence As a Contagious Disease. In *Envisioning criminology* (pp. 43–56). Springer International Publishing.

Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What Are We Weighting For? *Journal of Human Resources*, *50*(2), 301–316.

United States Department of Justice and Federal Bureau of Investigation. (n.d.). Uniform Crime Reporting program data.

van der Weele, J. (2012). Beyond the State of Nature: Introducing Social Interactions in the Economic Model of Crime. *Review of Law and Economics*, *8*(2), 401–432.

Webster, D. W., Whitehill, J. M., Vernick, J. S., & Curriero, F. C. (2012). Effects of Baltimore's Safe Streets program on gun violence: A replication of Chicago's CeaseFire program. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, *90*(1), 27–40.

Weibull, J. W. & Villa, E. (2006). *Crime, punishment and social norms.*

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press.

Zawitz, M. W. (1995). *Guns used in crime.*

Chapter 3

The Size of the Middle Class and Educational Attainment: Theory and Evidence from the Indian Subcontinent[1]

3.1 Introduction

The Industrial Revolution in Europe entailed an array of socio-political changes that transformed the basis of material wealth and political influence from landownership towards ingenuity and entrepreneurship. Representatives of this new kind of influential individuals predominantly belonged to the urban middle class. The formerly ruling class of landowners lost most of their influence both economically and politically (Doepke & Zilibotti, 2008; Galor, Moav, & Vollrath, 2009). Doepke & Zilibotti (2008) stress that this fundamental change in the basis of wealth is largely attributable to differences in the preference structures across social classes. The constituents of the land-owning class are seen as exhibiting a poor work ethic, a low preference for saving, and inadequate entrepreneurial and innovative skills. By contrast, the members of the new affluent social class of industrialists are held to be diligent and exhibiting a more future-oriented preference structure.

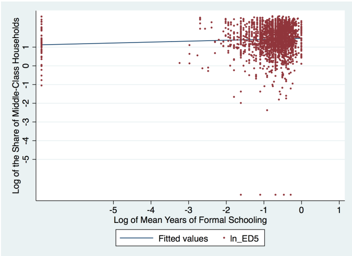
One channel by which differences in time preference rates exert an influence at the macroeconomic level is the savings behaviour of individuals. The more impatient individuals

[1] Joint work with Klaus Prettner. Published as Prettner & Seiffert (2018).

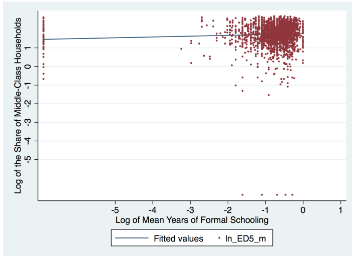
are, the less they save and the weaker is aggregate physical capital accumulation, one of the sources of growth over the medium run (Solow, 1956; Cass, 1965; Diamond, 1965; Koopmans, 1965). Another – probably even more important – channel is education, i.e., foregoing current consumption in favour of accumulating human capital, often of the subsequent generation (Galor & Weil, 2000; Galor, 2005, 2011). Since substantial educational investments are often made early in life, whereas the benefits of these investments accrue later or even to the next generation, societies with relatively large shares of middle class households that are relatively patient should exhibit a high level of average educational attainment.

There are many different pathways by which education exerts positive economic and social effects. Education raises the productivity of workers and has spillover effects in team production (Lucas, 1988); education fosters the creation of new ideas, in the historical context mostly by tinkerers and in modern times by means of targeted research and development of highly educated scientists (Romer, 1990; Strulik, Prettner, & Prskawetz, 2013); and education has numerous beneficial effects on socio-economic development, for example, on institutions and on democratization (Barro, 1999; Glaeser, La Porta, Lopez-de-Silanes, & Shleifer, 2004). Especially in the context of a conflict-prone country such as India, higher levels of average education seem to have a pacifying effect on societies (see Ostby & Urdal, 2011, for an overview of numerous contributions supporting this finding). For example, Alesina & Perotti (1996) show that socio-political instability is significantly reduced by higher levels of education in a panel of 71 countries over the time-span from 1960 to 1985; Tadjoeddin & Murshed (2007) find an inverted U-shaped relationship between average years of education and social conflict in Indonesia; and in a panel-analysis of 125 countries over the years 1960 through 1999, Collier (2004) finds that higher levels of male educational attainment are associated with significantly lower levels of conflict risk. Given the many beneficial effects of education, it is therefore of utmost importance for less developed countries to understand its central determinants.

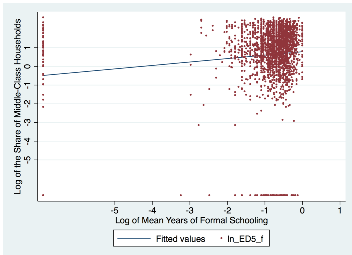
As previous research has found support for the hypothesis that a larger share of the middle class fosters education and industrialisation in Europe, this channel should obviously



(a) *Full Sample*



(b) *Male Sample*



(c) *Female Sample*

Figure 3.1.1. *Cross-Sectional Plots of Educational Attainment versus Share of Middle-Class in Indian villages (2005).*

not be ignored when assessing the evolution of developing countries towards modern, industrialised economies. As Figure 3.1.1 clearly shows, there exists a significantly positive correlation between middle class shares and average educational attainment in Indian villages. We see this as an indication that, especially in the context of educational attainment, the share of middle class households appears to play a similarly important role in developing economies in general – and India in a narrower sense – as it did in the Industrial Revolution in Europe.

In this contribution, we develop a stylised model of the demand for education in a society that is divided into three income strata. Households decide between the consumption of necessary goods, investments in the education of the subsequent generation, and spending on luxury goods. Since poorer households spend a substantial part of their income on subsistence needs, while richer households spend a non-negligible part of their income on luxury goods, the share of income that is spent on education tends to be the highest for members of the middle class. We show that in this setting, a higher proportion of middle class households in the society has the potential to induce higher levels of average educational attainment.

In order to empirically test the impact of middle class shares on average educational attainment, we use Indian household survey data which entails detailed information about education, income, and caste membership. To start off our empirical assessment, we examine the cross-sectional conditional correlation between the share of middle class households and average educational attainment in Indian villages/neighbourhoods (henceforth: villages). While the results from the simple OLS estimations confirm a positive association between middle class shares and education, they do not allow us to make any statements about the direction of causality. The first reason is the potential presence of omitted variable bias. The second reason is that reverse causality might distort the OLS results. More explicitly, observational units that are characterised by higher levels of education might be fertile breeding grounds for middle class households. To overcome this problem, we expand our empirical analysis to the extent that we construct a novel instrument for the share of middle class households. Exploiting India's unique Hinduistic society, we use the share of members of the two castes that lie between the spiritual and intellectual elite of the Hinduistic society, the *Brahmans*, and the untouchable lowest caste, *Dalits* as an instrument for middle class shares per village. Applying this instrument, we find a positive and significant effect of middle class shares on average educational attainment. Furthermore, we show that the effect is especially pronounced in rural areas and when focussing on female education. The strong positive effect of the share of the middle class on the educational attainment of women is particularly interesting because there is widespread consent that the empowerment of women is highly effective in promoting economic development (Diebolt & Perrin, 2013; Prettner & Strulik, 2017; Bloom, Kuhn, & Prettner, 2017).

Our paper is structured as follows. Section 3.2 contains the theoretical motivation and derives the optimal investment of households in the education level of their children. In Section 3.3, we analyse the relation between education and the size of the middle class from an empirical point of view, and in Section 3.4 we draw our conclusions.

3.2 The Size of the Middle Class and Economic Growth: Theoretical Motivation

In the following, we provide a stylised theoretical consideration to motivate the empirical analysis of the potential impact of inequality in terms of the population share of the middle class on educational attainment.

3.2.1 Basic Assumptions

Consider a developing economy populated by households who have the possibility to consume three different types of goods: i) necessary goods such as food, clothing, and basic shelter without which it is impossible to survive, ii) investments in the human capital of the next generation such as spending on education, and iii) luxury goods such as large housing, vacations abroad, expensive jewellery, etc., which might become desirable once that incomes are high enough and the basic needs are largely fulfilled. This structure implies a hierarchy of needs in the sense that households strive to fulfil their basic needs before they start to invest in education. The lowest priority is attached to luxury goods that households will only start to consume if their incomes surpass the threshold above which the diminishing marginal utility from spending on basic consumption goods and on education renders the consumption of luxury goods to be desirable.

3.2.2 Optimal Choices of the Households

We conceptualize the described hierarchy of needs by the following household utility function:

$$U(c_{s,t}, e_t, c_{l,t}) = \omega_s \log(c_{s,t} - \bar{c}_s) + \omega_h \log(h_t) + \omega_l \log(c_{l,t} + \bar{c}_l), \tag{3.2.1}$$

where $U(\cdot)$ is the utility level of the household, $c_{s,t}$ refers to consumption of basic goods at time t, h_t denotes expenditures on educating children, $c_{l,t}$ is consumption of luxury goods, ω_i for $i = s, h, l$ refers to the weights of the corresponding goods in the utility function with $\omega_s > \omega_e > \omega_l$, $\bar{c}_s$ denotes the subsistence consumption level of basic goods that is necessary for survival of period t, and $\log(\bar{c}_l)$ defines a lower bound on felicity derived from luxury goods for the case in which no income is spent on this item, i.e., $c_{l,t} = 0$. A reasonable assumption would be $\log(\bar{c}_l) = 0$ such that, without any spending on luxury goods, the utility derived from this component is zero. The short-cut formulation of altruism in this model is the well-known "warm-glow" motive of giving (Andreoni, 1989) that usually leads to qualitatively similar results as the more complicated formulation of a dynastic utility function. Note that we could also introduce other standard goods that do not belong to the basket of basic goods or luxury goods and would therefore exhibit a similar utility effect as h_t. However, this would merely complicate the exposition without changing the results and without adding new insights.

The budget constraint of a household implies that expenditures on the different types of consumption and on education must not exceed income. Formally, it is given by

$$w_t \geq c_{s,t} + h_t + p_{l,t}c_{l,t}, \tag{3.2.2}$$

where w_t is the income level of the household at time t, the price of the basic consumption goods ($c_{s,t}$) is normalized to unity, the price of education is measured in terms of foregone basic consumption, and the price of luxury goods ($c_{l,t}$) is given by $p_{l,t} > 1$. Note that, due to the local non-satiation implied by our utility function, the budget constraint will always hold with equality at any optimal allocation for $w_t > 0$. We can therefore solve the optimization problem by means of the method of Lagrange (see Appendix 3.4 for the derivations). From now on we assume that the income level is high enough to guarantee that the subsistence consumption needs are fulfilled, i.e., $w_t > \bar{c}_s$ and people do not starve to death.

The set of optimal choices can be split into two parts depending on the level of household income (see Appendix 3.4). For a high income level $w_t > p_{l,t}(\omega_h + \omega_s)\bar{c}_l/\omega_l + \bar{c}_s \equiv \hat{w}_t$, the demand functions are given by

$$c_{s,t} = \frac{\omega_s(w_t + p_{l,t}\bar{c}_l) + (\omega_h + \omega_l)\bar{c}_s}{\omega_s + \omega_h + \omega_l}, \tag{3.2.3}$$

$$h_t = \frac{\omega_h(w_t + p_{l,t}\bar{c}_l - \bar{c}_s)}{\omega_s + \omega_h + \omega_l}, \tag{3.2.4}$$

$$c_{l,t} = \frac{\omega_l(w_t - \bar{c}_s) - p_{l,t}(\omega_h + \omega_s)\bar{c}_l}{p_{l,t}(\omega_s + \omega_h + \omega_l)}, \tag{3.2.5}$$

which are all positive as long as $w_t > \hat{w}_t$. However, for $w_t \leq \hat{w}_t$, households find themselves in the corner solution at which the demand for luxury goods is zero and the other two demand functions are given by

$$c_{s,t} = \frac{\omega_s w_t + \omega_h \bar{c}_s}{\omega_s + \omega_h}, \tag{3.2.6}$$

$$h_t = \frac{\omega_h(w_t - \bar{c}_s)}{\omega_s + \omega_h}. \tag{3.2.7}$$

To summarize, we have the following set of (non-homothetic) demand functions for the whole range of allowed income levels $w_t > \bar{c}_s$

$$c_{s,t} = \begin{cases} \frac{\omega_s w_t + \omega_h \bar{c}_s}{\omega_s + \omega_h} & \text{for } \bar{c}_s < w_t < \hat{w}_t, \\ \frac{\omega_s(w_t + p_{l,t}\bar{c}_l) + (\omega_h + \omega_l)\bar{c}_s}{\omega_s + \omega_h + \omega_l} & \text{for } \hat{w}_t < w_t, \end{cases}$$

$$h_t = \begin{cases} \frac{\omega_h(w_t - \bar{c}_s)}{\omega_s + \omega_h} & \text{for } \bar{c}_s < w_t < \hat{w}_t, \\ \frac{\omega_h(w_t + p_{l,t}\bar{c}_l - \bar{c}_s)}{\omega_s + \omega_h + \omega_l} & \text{for } \hat{w}_t < w_t, \end{cases}$$

$$c_{l,t} = \begin{cases} 0 & \text{for } \bar{c}_s < w_t < \hat{w}_t, \\ \frac{\omega_l(w_t - \bar{c}_s) - p_{l,t}(\omega_h + \omega_s)\bar{c}_l}{p_{l,t}(\omega_s + \omega_h + \omega_l)} & \text{for } \hat{w}_t < w_t. \end{cases}$$

For a comparatively low income level, households spend most of their income on basic consumption goods and do not consume luxury goods at all. As incomes rise, the share of income spent on basic consumption goods decreases, while the share of income spent on education increases. Once that household income surpasses the level of $\hat{w}_t$, consumption of luxury goods becomes positive and the share of income spent on education of the children starts to decrease again.

3.2.3 The Income-Based Stratification of the Society

Now we assume that the society consists of three income groups i) the rich (indexed by r) with an income level of $w_{r,t} > \hat{w}_t$, ii) the middle class (indexed by m) with an income level of $w_{m,t} < w_{r,t}$, and iii) the poor (indexed by p) with an income level of $w_{p,t} < w_{m,t}$. We normalize the total population size to unity and denote the share of the rich by θ_r and the share of the poor by θ_p such that the share of the middle class is given by $1 - \theta_r - \theta_p$. With these assumptions, the share of education expenditures in the economy determines the average human capital stock of the next generation as

$$\begin{aligned}\bar{h} &= \frac{\theta_p h_{p,t} + (1-\theta_p-\theta_r)h_{m,t} + \theta_r h_{r,t}}{\theta_p w_{p,t} + (1-\theta_p-\theta_r)w_{m,t} + \theta_r w_{r,t}} \\ &= \frac{\omega_e\left[\frac{\theta_r(p_l\bar{c}_l-\bar{c}_s+w_{r,t})}{\omega_e+\omega_l+\omega_s} + \frac{(1-\theta_p-\theta_r)(w_{m,t}-\bar{c}_s)}{\omega_e+\omega_s} + \frac{\theta_p(w_{p,t}-\bar{c}_s)}{\omega_e+\omega_s}\right]}{w_{m,t}\left(1-\theta_p-\theta_r\right) + \theta_p w_{p,t} + \theta_r w_{r,t}}. \end{aligned} \tag{3.2.8}$$

The appropriate interpretation of this expression is the following. The cost of education rises with the weighted average of the incomes in an economy which is reflected by the denominator. The reason is that education is labour-intensive such that the salaries of teachers, instructors, and professors rise with the average salary level of a country. Thus, a nominal increase in the expenditures of households on education does not necessarily lead to an increase in the human capital stock of the next generation because the increase could just compensate for a given increase in the nominal wages of teachers, instructors, and professors. What is needed to increase the average human capital stock of the next generation is an increase in the *share* of expenditures that are devoted to education.

It remains to be shown how the average human capital stock depends on the income-specific stratification of the society. To this end, we show how the average human capital stock depends on the population shares of the poor and the rich. Taking the derivatives of Equation (3.2.8) with respect to θ_r and θ_p yields

$$\frac{\partial \bar{h}}{\partial \theta_p} = \frac{\left(w_{m,t} - w_{p,t}\right)\left\{\theta_r\left[p_l \bar{c}_l\left(\omega_e + \omega_s\right) - \omega_l w_{r,t}\right] - \bar{c}_s\left[\omega_e + \omega_l\left(1 - \theta_r\right) + \omega_s\right]\right\}}{\omega_e^{-1}\left(\omega_e + \omega_s\right)\left(\omega_e + \omega_l + \omega_s\right)\left[w_{m,t}\left(1 - \theta_p - \theta_r\right) + \theta_p w_{p,t} + \theta_r w_{r,t}\right]^2}, \tag{3.2.9}$$

$$\begin{aligned}\frac{\partial \bar{h}}{\partial \theta_r} &= \frac{\left[w_{m,t}\left(\theta_p - 1\right) - \theta_p w_{p,t}\right]\left[\omega_l w_{r,t} - p_l \bar{c}_l\left(\omega_e + \omega_s\right)\right]}{\omega_e^{-1}\left(\omega_e + \omega_s\right)\left(\omega_e + \omega_l + \omega_s\right)\left[w_{m,t}\left(1 - \theta_p - \theta_r\right) + \theta_p w_{p,t} + \theta_r w_{r,t}\right]^2} \\ &+ \frac{\bar{c}_s\left[w_{r,t}\left(\omega_e + \omega_l + \omega_s\right) - w_{m,t}\left(\omega_e + \omega_l \theta_p + \omega_s\right) + \omega_l \theta_p w_{p,t}\right]}{\omega_e^{-1}\left(\omega_e + \omega_s\right)\left(\omega_e + \omega_l + \omega_s\right)\left[w_{m,t}\left(1 - \theta_p - \theta_r\right) + \theta_p w_{p,t} + \theta_r w_{r,t}\right]^2}.\end{aligned} \tag{3.2.10}$$

Inspecting Equation (3.2.10), we observe that i) the common denominator of the terms in both lines is always positive; ii) the numerator of the term in the second line is always positive because $w_{m,t}\left(\omega_e + \omega_l \theta_p + \omega_s\right) < w_{r,t}\left(\omega_e + \omega_l + \omega_s\right)$ due to $w_{r,t} > w_{m,t}$ and because $\theta_p < 1$; iii) as a consequence of i) and ii) the term in the second line is always positive; iv) the sign of the term in the first line is ambiguous and depends on the sign of the expression $\omega_l w_{r,t} - p_l \bar{c}_l\left(\omega_e + \omega_s\right)$. If this expression is positive, the term in the first line of Equation (3.2.10) is negative such that a reduction in the population share of the middle class that is due to an increase in the population share of the rich has a negative effect on average human capital. The intuition is that the rich spend a portion of their income on luxury goods such that the share that they spend on educating their children is lower in comparison to the middle class. An overall increase of the population share of the rich could therefore reduce the overall ratio of spending on human capital accumulation. Note that the expression $\omega_l w_{r,t} - p_l \bar{c}_l\left(\omega_e + \omega_s\right)$ is more likely to be positive if the rich have a strong preference for luxury goods, i.e., if ω_l is relatively large in comparison to ω_e and ω_s.

Inspecting Equation (3.2.9), we observe that i) the denominator is again always positive; ii) the sign of the numerator is a priori ambiguous; iii) the term $\theta_r\left[p_l \bar{c}_l\left(\omega_e + \omega_s\right) - \omega_l w_{r,t}\right]$ has a similar effect as in Equation (3.2.10) but it is weighted with the population share of the rich (θ_r); iii) the term $-\bar{c}_s[\omega_e + \omega_l\left(1 - \theta_r\right) + \omega_s]$ is always negative. This implies that a larger subsistence need $\bar{c}_s$ reduces the fraction of income that the poor spend on education because a higher $\bar{c}_s$ means that it is more difficult to fulfil the subsistence consumption needs. Therefore, it is again possible that an increase in the population share of the poor reduces average education expenditures.

Altogether, we have a plausible mechanism by which the population share of the middle class raises the share of resources that a society devotes to education. If the fraction of education spending of the middle class is larger than the corresponding fractions of the poor and the rich, then an increase in the relative size of the middle class raises overall human capital accumulation. This is the implication that we test in Section 3.3.

3.3 The Size of the Middle Class and Educational Attainment: Empirical Results

In the following, we present the empirical investigation of the effects of the relative size of the middle class on education. The subsequent analysis is based on household data from the Indian subcontinent. While we are aware of the usual concerns about accuracy of household income data for developing countries, we also emphasise the advantages of using Indian villages as the unit of analysis for our project. Not only is India the second most populous country in the world. It also displays a vast heterogeneity with respect to ethno-linguistic and other societal criteria. All subject to a relatively homogeneous legislative landscape. Furthermore, India is characterised by substantial cross- as well as within-state income inequality (see Vannemann & Dubey, 2013, for a comprehensive description). In addition, the data published by the Indian Human Development Survey (IHDS) provides a sound basis for empirical analysis both from a coverage and a quality perspective. The strongest argument for basing our choice of the research subject, however, is the Indian caste system. It is unique regarding its impact on social life and its persistence over centuries, which enables us to use it as a novel instrument for the size of the middle class. This in turn allows us to shed light on the causal effect of the size of the middle class on educational outcomes. Many other attempts to empirically investigate the question at hand are hampered by the lack of a valid instrument for the share of the middle class.

3.3.1 Measuring the Size of the Middle Class in India

The term *middle class* in social sciences is of ambivalent nature. There exists a wide array of potential characteristics on which the employed definition could be based on. They range from income-based measures to metrics based on occupational functions such as the tier in management hierarchies (Chauvel, 2013). In economics, scholars mostly rely on some income-based variant, where members of the middle class are those whose incomes lie within some interval including the median/mean. This interval is often symmetric with an early example being the study by Thorow (1987) in which the interval ranges from 75% to 125% of the mean (Ravallion, 2010).

While it appears that Thorow's measure has become somewhat of a standard in the literature on affluent economies, there is less consensus on developing economies. The measures used vary widely with the scope of the respective contribution. According to Ravallion (2010), we can identify the following groups: Birdsall, Graham & Pettinato (2000), among others, stick closely to the widely used relative measure spanning the aforementioned interval in a study assessing potential changes in size and income shares of the middle class and their relationship to increased integration in global markets. The second group of scholars rely on absolute measures to quantify the middle class. A prominent example is Banerjee & Duflo (2008) who rely on a measure that defines the middle class as households with daily PPP per capita expenditures between \$2 and \$10. In justifying their measure, they argue that it produces similar results with respect to population shares as the income quantiles covered by Thorow's median based relative measure. The third group is represented by Milanovic & Yitzhaki (2002). Those authors' definition of a global middle class includes all persons living on incomes between the mean incomes of Brazil and Italy.

This leaves us with the task of deciding which households we should consider as the middle class in the Indian case. As Figure 3.3.1 shows, India is characterised by an income distribution that is strongly skewed to the left. Daily household incomes per capita in PPP US dollars range from 0.0003 to 303.55 with a mean of 2.91. Clearly, the above mentioned absolute middle class measures are not applicable in the underlying

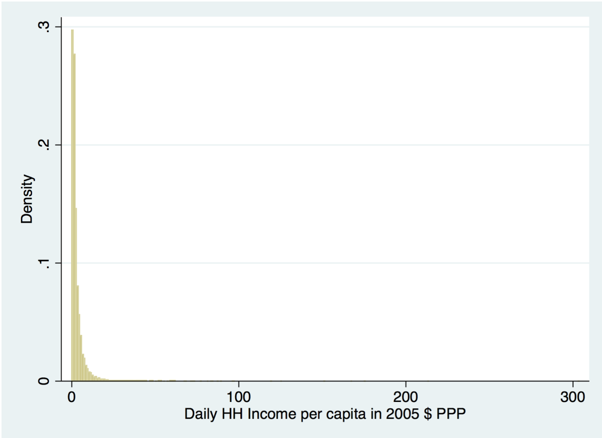


Figure 3.3.1. *Histogram of Indian Household Incomes.*

analysis focussing on Indian villages. Following Vannemann & Dubey (2013), we therefore assign all households with incomes between 75% and 250% of the district median income to the middle class. We base our middle class measure on the district median instead of the country median due to the substantial cross-district income differences illustrated in Figure 3.3.3. In the presence of a standard deviation that is about half of the mean of this measure, taking the national or even state-level median as a reference would lead to a highly biased measure of the middle class.

By following this strategy, we arrive at a mean share of middle class households per village of 36%. This is comparable to the share of middle class households in the United States in 2010 as reported by the OECD. Furthermore, as Figure 3.3.2 indicates, the skewness of the distribution of the middle class shares per village is rather low.

One potential concern with our measure could be that higher shares of middle class households per village coincide with higher mean household incomes in the respective village. Comparing Figures 3.3.3 and 3.3.5, one can see that a high mean income seldom coincides with a large share of middle class households. The unconditional correlation

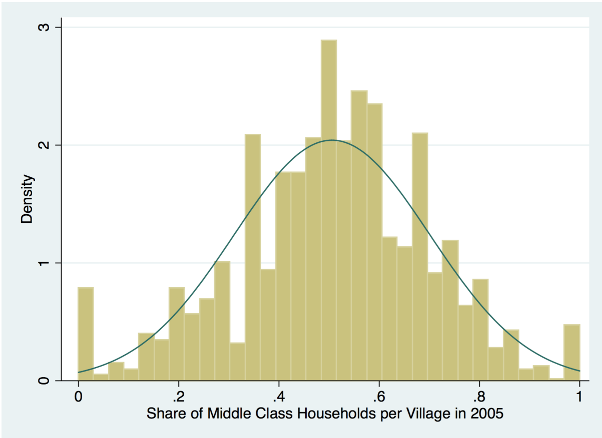


Figure 3.3.2. *Histogram of Middle Class per Village.*

between mean village income and the middle class share is 0.03. This is strong evidence against this potentially problematic correlation.

3.3.2 Data

Our analysis is based on household, individual, and village level data sets from the Indian Household Survey (IHDS) 2005. The IHDS provides nationally representative data on a multitude of topics sampled from 41,554 households in 1503 villages and 971 urban neighbourhoods for the years 2005 and 2011[2]. The village units are derived from the IHDS' primary sampling units (PSU). They are the lowest level of aggregation unit above the household level for which the IHDS data structure allows. With its extensive coverage both in terms of the representativity and the span of covered topics, it is, to our knowledge, the premier survey covering India. In what follows, we will provide a brief discussion of the key variables and their construction.

The 2005 wave of the IHDS is comprised of 8 data sets of which we utilise the individual

[2] The data are available here: https://ihds.umd.edu.

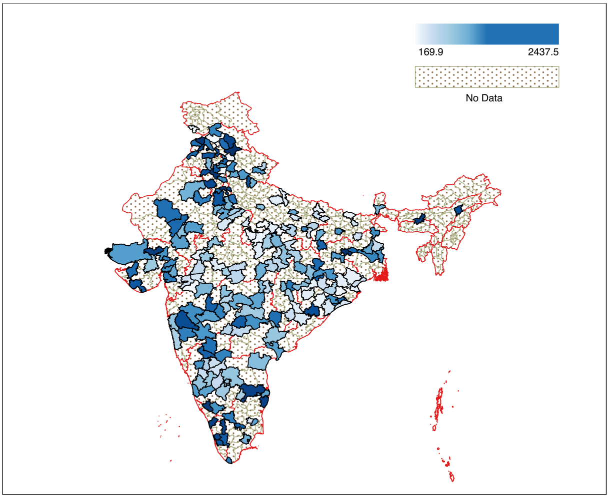


Figure 3.3.3. *Heatmap of Annual Median Household Incomes per District.*

file, the household file, and the village file.[3] Our left-hand side variables are created from the individual file, while we construct our key variable of interest and the employed instrumental variable from the household file and the village file, respectively.

The key dependent variable in our empirical setup is the average educational attainment per PSU. In order to construct this measure, we use the mean years of schooling reported by PSU. Following Castelló-Climent, Chaudhary & Mukhopadhyay (2017), we restrict our sample to the survey population aged 25 years and above to ensure that the upper end of the education distribution is not censored by age. In order to identify potentially different effects on the educational outcomes of men and women, we run the regression using three different dependent variables: the plain average years of formal schooling per PSU as well as mean male and female education separately. While the cross-PSU mean of our main education measure is 4.8 years, it varies from 0 to 13.8 years[4]. A brief look at the

[3] The further 5 files cover: medical facilities, non-resident family members, primary school, birth history, and crop production.

[4] The cross-PSU mean for male education is 6.2 with a range from 0 to 14.7. The cross-PSU mean

Table 3.3.1. *Descriptive Statistics for Sampling Units*

		Mean	Standard Deviation
Log Educational Attainment			
Full Sample	1961	1.733	0.553
Male	1987	1.932	0.467
Female	1987	1.357	1.103
Log Share of Middle-Class Households	1961	-1.017	1.054
Log Household Income	1961	5.303	0.800
Urban dummy	1961	0.324	0.468
Log Distance to Bus Station	1937	-4.573	3.676

respective summary statistics for the split sample justifies our strategy to additionally run the regression on dependent variables based on gender. Indian women on average spend roughly half the time in the formal education system as compared to men. Furthermore, this gender difference is more pronounced in rural PSUs as compared to urban areas[5]. Figure 3.3.4 displays the district means of educational attainment for the full sample. The darker the blue, the higher is average years of schooling. One can easily observe that Indian districts exhibit considerable spatial heterogeneity in education. This holds both across and within states, which are indicated by red outlines. For the female sample, the cross-district heterogeneity is roughly 10% higher[6].

While the within-state heterogeneity with regards to educational attainment is already striking, within district heterogeneity is even more pronounced. We are unable to graphically depict this heterogeneity at the lower aggregation level due to location-censoring in the underlying survey. Table 3.3.1 shows that it appears to be even more pronounced within districts. Hence, in our further analysis, we will empirically assess in how far those differences in educational attainment may be driven by differences in the share of middle class households at the PSU level.

The key variable of interest is the proportion of households per sampling unit that belong

for female education is 3.4 with a range from 0 to 13.6.

5 In rural areas, women report about 44% of the educational attainment relative to males. In urban areas, the fraction is about 67%.

6 Figures 3.B.1 and 3.B.2 in the appendix display the spatial distribution of average years of schooling per district for the male and female samples.

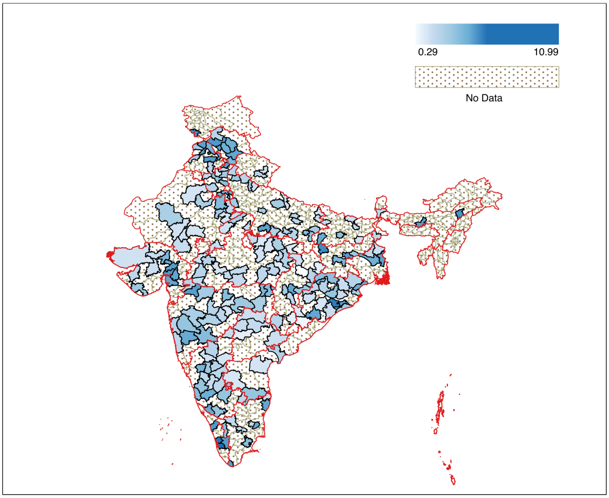


Figure 3.3.4. *District-Level Mean of Education in Years (Full Sample).*

to the middle class. We use it to estimate the effect of the size of the middle class on educational attainment. In so doing, we define a household as belonging to the middle class if household income lies between 75% and 250% of its respective district. We chose the district mean over the state or even country mean due to the vast cross-state and cross-district inequality in household per capita incomes[7]. As figure 3.3.5 illustrates, there is vast heterogeneity in the middle class household shares at the district level. This holds – as with the two previously presented measures – for both the within-state and the cross-state perspective and is also observable comparing adjacent districts.

In Section 4.1 we discussed the presence of a positive unconditional correlation between the share of middle class households per PSU and our three different education measures. Still, it might be the case that this basic relationship is distorted by potentially confounding factors. In order to purge our specification from those effects, we include an array of control variables in our baseline specification. This includes measures for mean income

[7] Figure 3.B.3 in the appendix illustrates this considerable heterogeneity at the district level.

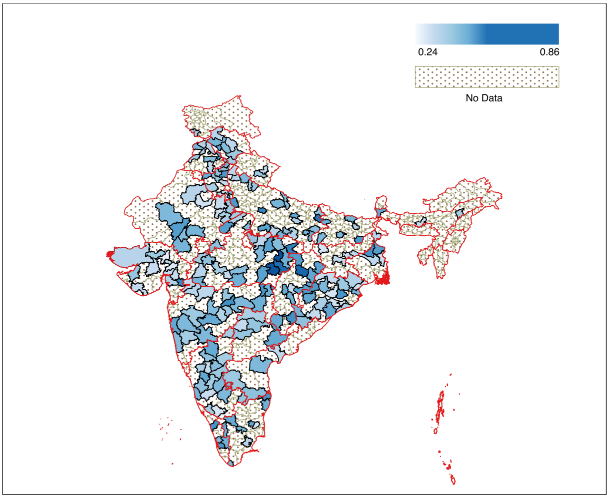


Figure 3.3.5. *District-Level Mean Share of Middle Class Households.*

per PSU (*Income*), the distance to the closest middle school (*Dist. School*), the distance to the closest bus station (*Dist. Bus*), and a binary indicator for urban status (*Urban*). In addition, all specifications include district-level fixed effects.

3.3.3 Baseline Econometric Specification & Empirical Results

We commence our empirical investigation by a simple OLS estimation of the following equation

$$ln\left(Education_c\right) = \beta \times ln(MiddleClass_c) + Controls + \epsilon_c \qquad (3.3.1)$$

where $Education_c$ is average educational attainment, c indexes villages, our independent variable of interest is the share of middle-class households in a PSU ($MiddleClass_c$), ϵ_c is the error term, and β is the parameter that we aim to estimate. We run the regression thrice, using the three different measures of educational attainment as dependent variable.

We take the logarithm of all employed variables except for the urban dummy.[8] Potentially unobservable confounders are being accounted for by the inclusion of district-specific dummy variables. In all cases, we report robust standard errors. The results of this baseline OLS specification are reported in Tables 3.3.3 to 3.3.4.

Table 3.3.2. *Baseline Regression Results (Full Sample)*

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Education (full)	Education (full)	Education (full)	Education (full)	Education (full)
Middle Class	0.081***	0.036**	0.040**	0.039**	0.039**
	(0.020)	(0.017)	(0.017)	(0.016)	(0.016)
Income		0.405***	0.396***	0.328***	0.323***
		(0.026)	(0.026)	(0.028)	(0.028)
Dist. School			-0.021***	-0.018***	-0.016***
			(0.005)	(0.005)	(0.005)
Urban				0.357***	0.335***
				(0.038)	(0.039)
Dist. Bus					-0.010*
					(0.005)
Observations	1,937	1,937	1,937	1,937	1,937
R-squared	0.447	0.534	0.540	0.557	0.558

Note: The table reports estimates of Equation (3.3.1) with *Education (all)* as dependent variable. All specifications include district fixed effects. Robust standard errors are reported in parentheses, where *, **, and *** indicate significance at 1%, 5%, and 10% level, respectively.

Comparing columns (1) through (5), we observe that the magnitude of the coefficient of interest remains fairly constant after adding the logarithm of mean PSU income. This indicates that the mean income level appears to be the most important factor to correct for within district differences in PSU-level average educational attainment. All further controls, while being statistically significant and in cases of considerable magnitude, have hardly any noticeable impact on the strength of the relationship between the size of a PSUs middle class and the associated levels of education. Accordingly, a one standard deviation increase in the share of middle-class households is associated with a 4.2% increase in average years of formal education of the respective PSU's population. Absent the full set of controls, this effect is about twice as large. It is noteworthy that, while being substantially smaller than the strength of the correlation between education and income as well as between education and urbanity, the size of the middle class plays a considerably

[8] Since it is the case that variables take on the value 0, we add a small number (0.001) to all non-binary variables in the regression.

larger role than the two variables that we included to catch PSU-level differences in the indirect cost of schooling caused by higher commuting costs (*Dist. School* and *Dist. Bus*).

Table 3.3.3. *Baseline Regression Results (Male Sample)*

VARIABLES	(1) Education (full)	(2) Education (full)	(3) Education (full)	(4) Education (full)	(5) Education (full)
Middle Class	0.081***	0.036**	0.040**	0.039**	0.039**
	(0.020)	(0.017)	(0.017)	(0.016)	(0.016)
Income		0.405***	0.396***	0.328***	0.323***
		(0.026)	(0.026)	(0.028)	(0.028)
Dist. School			-0.021***	-0.018***	-0.016***
			(0.005)	(0.005)	(0.005)
Urban				0.357***	0.335***
				(0.038)	(0.039)
Dist. Bus					-0.010*
					(0.005)
Observations	1,937	1,937	1,937	1,937	1,937
R-squared	0.447	0.534	0.540	0.557	0.558

Note: The table reports estimates of Equation (3.3.1) with *Education (male)* as a dependent variable. All specifications include district fixed effects. Robust standard errors are reported in parentheses, where *, **, and *** indicate significance at 1%, 5%, and 10% level, respectively.

Comparing the results from the two alternative regression specifications focussing on male and female educational attainment reported in Tables **??** and 3.3.4 yields valuable additional insights. It appears that the magnitude of the relation between the dependent variable and the independent variables is smaller in the male sample, whereas it becomes substantially larger in the female sample. A first conclusion that we can draw from this fact is that, unsurprisingly, women fare much better education-wise in better-off and urban areas. Similarly, female education exhibits a stronger negative relationship with higher travel costs to school than male education. In addition, it is worth noting that, while the coefficients on the controls such as income or urbanity double or triple relative to the male sample, the magnitude of the relationship to the share of the middle class in the female sample is about seven times as large as compared to the male sample. This can be taken as first evidence for the share of the middle class not only being important for male and overall education but that it is especially strongly correlated with higher educational outcomes for Indian women.

The results from our baseline empirical examination so far suggest a sizeable positive correlation between the share of middle class households and educational attainment.

Table 3.3.4. *Baseline Regression Results (Female Sample)*

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Education (fem)	Education (fem)	Education (fem)	Education (fem)	Education (fem)
Middle Class	0.297***	0.215***	0.223***	0.220***	0.220***
	(0.067)	(0.058)	(0.058)	(0.057)	(0.057)
Income		0.742***	0.725***	0.578***	0.571***
		(0.058)	(0.058)	(0.061)	(0.062)
Dist. School			-0.037***	-0.029**	-0.027**
			(0.014)	(0.014)	(0.013)
Urban				0.772***	0.739***
				(0.104)	(0.111)
Dist. Bus					-0.014
					(0.016)
Observations	1,937	1,937	1,937	1,937	1,937
R-squared	0.431	0.479	0.482	0.495	0.496

Note: The table reports estimates of Equation (3.3.1) with *Education (fem)* as a dependent variable. All specifications include district fixed effects. Robust standard errors are reported in parentheses, where *, **, and *** indicate significance at 1%, 5%, and 10% level, respectively.

The relationship is robust to the inclusion of a broad set of controls. Furthermore, the relationship is more pronounced if the sample is restricted to women. Due to the likely existence of endogeneity in the underlying relation, these insights do not yield any information on the direction of causality. To address this issue, we turn to an instrumental variable approach in the subsequent sections.

3.3.4 Using the Share of OBC Households as an Instrument

The Indian caste system is an ancient system of societal stratification that is unique to the Hinduistic culture of the Indian subcontinent. It provides a strictly defined hierarchical order of the society that has been in place for thousands of years and remained largely unchanged until today. This persistence makes it the ideal candidate for a historical instrument for middle class shares. In this section, we describe the historical and sociological characteristics of the Hinduistic caste system that enables the subsequent empirical specification. It is based on Ghurye (1932) which is – until today – one of the most influential contributions on the topic.

The oldest obtainable records about the Indian caste system can be attributed to the Indo-Aryan culture and date back as far as 1500 BCE. While the Indo-Aryan culture was stemming from and in its early times was mostly restricted to the Gangetic Plain,

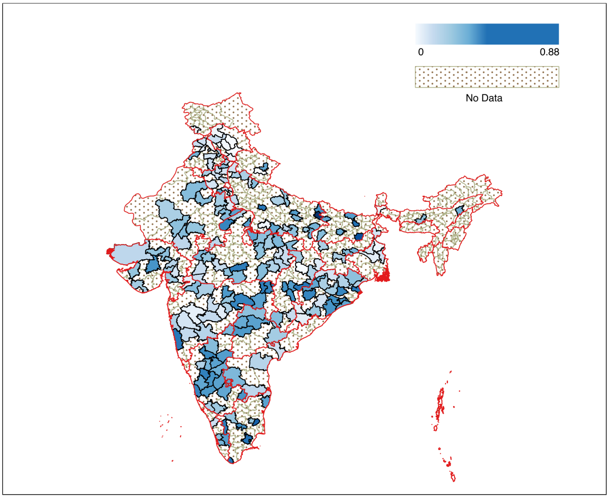


Figure 3.3.6. *District-Level Mean Share of OBC Households.*

followers of Brahmanism diffused it across the Indian subcontinent. It was first mentioned by a non-local source around 300 BCE. The Greek explorer Megasthenes described it in the following words: "it is not permitted to contract marriage with a person of another caste, nor to change from one profession or trade to another [...]".

Ghurye (1932) lists the following basic characteristics that make the Hinduistic culture a perfect candidate for our IV strategy:

Segmentation of Society. The caste system in the Hinduistic culture was omnipresent. It consisted of five main groups, so-called *varnas*, which were again organised in different subgroups.[9] While most of the Western cultures linked societal status mainly to wealth, caste membership was solely determined by birth. Accordingly, two soldiers (a profession open to most of the castes) could exhibit the same rank in the military and similar

[9] The *jati*-system was rather functional than hierarchical. Mainly, it indicated the occupation of its bearer and also is the basis of Indian surnames. So, basically, two different *jatis* could lie within the same *varna*.

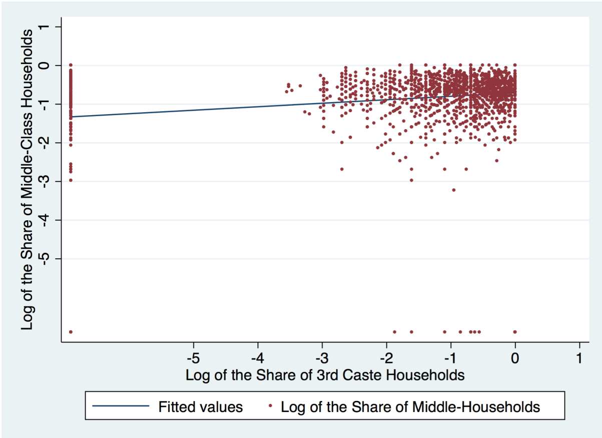


Figure 3.3.7. *Cross Sectional Plot of Share of Middle-Class Households Versus Share of OBC Caste Households.*

personal wealth, while the hierarchy defined by the *varna*-system would still determine their relative social status in civil life. In contrast to classes in the Weberian sense, which do not exhibit standing councils or explicit codes of conduct, the majority of the Indian castes had standing committees. Those were ruling on far more issues than, for example, guilds or similar organisations.

The pronounced division of society as a consequence of this system resulted in millennia-old patterns of caste-wise endogamy as well as a clear separation of them within villages.

Hierarchical Structure of Society. In addition to the clear separation, there was also a strict hierarchical order with the *Brahmans* on top. In the hinduistic context, *dharma* describes the notion that each individual has a certain role to play or a function to fill to ensure the upholding of spiritual and social order. The *brahmans* being born into the highest caste exercise priestly and religious tasks. The *kshatriyas* formed the second class which was tasked with administrative and military duties. The third class, the *vaishyas* can be seen as a caste of commoners. Their tasks included agriculture and trading. The lowest of the four *varnas* are the *shudras*. Their duty is to work in all trades that serve the three superior castes which are considered to be twice-born. They would work as farmhands, be shop-owners, etc. The lowest-status group in the hinduistic society are the *dalits*, often also referred to as *untouchables*. While opinions vary whether they can actually be seen as part of the varnas, they undoubtedly are the absolutely lowest group tasked with everything that is considered as unclean such as tanning or cleaning latrines.

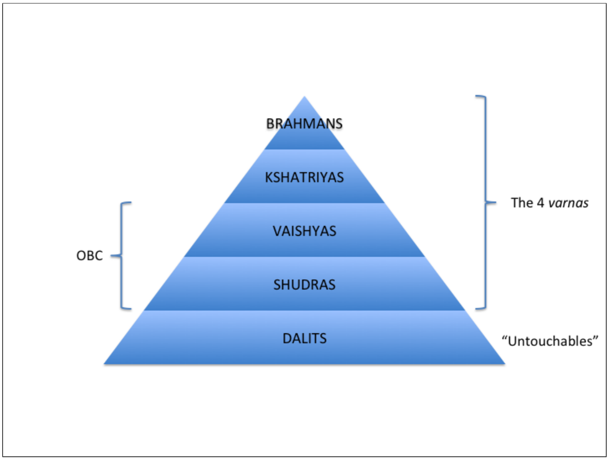


Figure 3.3.8. *Hierarchical Order of the varnas (own depiction based on Ghurye, 1932).*

The hierarchical system as a whole is strongly related with the notion that spiritual pollution is transmitted from lower towards upper castes. In theory, physical contact between members of a superior and an inferior caste soils the body of the member of the relatively high caste. In some regions of modern India this goes so far that even the shadow of a *dalit* overlapping with the body of a member of a higher caste results in the defiling of the superiors body. This further aggravated the lack of social intercourse between members of different castes and with it also the strong caste-endogamous mating patterns, keeping the caste-principle alive and highly present until today.

The IHDS reports caste membership in the following categories: *Brahmin*, *High Caste*, *Other Backward Caste (OBC)*, *Dalit*. Furthermore, there are four other categories comprising the non-hindu members of the modern Indian society.[10] The OBC definition in the IHDS is very closely linked to the castes of *Vaishyas* and *Shudras*. As the historical functions of those castes most closely resembles the modern image of a middle-class household, we suggest the proportion of OBC households per village as an instrument for middle class shares at the village level in the subsequent instrumental variable analysis. Figure 3.3.7 clearly indicates that there exists a positive relationship between the shares of middle-class households and the share of OBC households at the village level.

3.3.5 Instrumental Variable Specification and Empirical Results

In this section, we present the instrumental variable specification and the results from running the cross-section model presented in Subsection 3.3.3. In particular, we modify Equation (3.3.1) such that we use the share of OBC-households per village as an instrument for the share of middle-class households. The set of employed control variables and fixed effects remains the same.

Table 3.3.5 presents the results from the IV regression. Panel A indicates a positive and significant coefficient for the share of OBC households. The magnitude of the coefficient is slightly diminishing with the inclusion of additional controls moving from column (1) to (5). In presence of the full set of control variables displayed in column (5), a one standard deviation increase in the share of OBC households is associated with a 27,7% increase in the share of middle-class households in the respective PSU. This estimate is significant at the 1%-level. As indicated by the Kleibergen-Paap statistic and the F-statistic across panels B through D, our specification is not affected by weak identification issues.[11]

Panel B in Table 3.3.5 reports the second-stage results for average PSU education as

[10] Those categories are: *Adivasi* (mostly animistic tribes), *Muslim*, *Sikh/Jain* and *Christian*. As we exclude all these categories from the empirical analysis because the focus is on hinduistic India, we do not go further into the details at this point.

[11] The critical values of the test provided by Stock & Yogo (2005) are 16.38, 8.96, 6.66, and 5.53 for a 10%, 15%, 20%, and 25% bias of the obtained estimator, respectively. Accordingly, the null hypothesis of the underlying estimators being biased due to weak instrumentation are rejected in all cases.

Table 3.3.5. *Instrumental Variables Results: Different Dependent Variables*

	(1)	(2)	(3)	(4)	(5)
Panel A: First Stage - Share of MC Households on Share of OBC Households					
OBC Share	0.132***	0.125***	0.119***	0.119***	0.122***
	(0.021)	(0.020)	(0.021)	(0.022)	(0.022)
Panel B: Second Stage - Education (full)					
Middle Class	0.411***	0.370***	0.360***	0.363***	0.358***
	(0.102)	(0.102)	(0.099)	(0.098)	(0.098)
Kleibergen-Papp F Stat.	30.32	30.41	30.94	30.94	30.70
ARW F Stat.	20.87	17.13	17.10	17.68	17.38
Observations	1,937	1,937	1,937	1,937	1,937
Panel C: Second Stage - Education (male)					
Middle Class	0.374***	0.342***	0.333***	0.335***	0.330***
	(0.098)	(0.100)	(0.097)	(0.096)	(0.096)
Kleibergen-Papp F Stat.	29.89	29.98	30.50	30.50	30.26
ARW F Stat.	17.67	14.23	14.18	14.51	14.30
Observations	1,965	1,965	1,965	1,965	1,965
Panel D: Second Stage - Education (female)					
Middle Class	0.947***	0.874***	0.857***	0.863***	0.855***
	(0.229)	(0.227)	(0.222)	(0.222)	(0.222)
Kleibergen-Papp F Stat.	29.89	29.98	30.50	30.50	30.26
ARW F Stat.	19.48	18.19	17.98	18.30	17.94
Observations	1,965	1,965	1,965	1,965	1,965
Controls					
District FE	YES	YES	YES	YES	YES
Income	NO	YES	YES	YES	YES
Dist. School	NO	NO	YES	YES	YES
Urban	NO	NO	NO	YES	YES
Dist. Bus	NO	NO	NO	NO	YES

Note: The table reports IV estimates of Equation (3.3.1) with alternating education variables as a dependent variable. Robust standard errors are reported in parentheses, where *, **, and *** indicate significance at 1%, 5%, and 10% level, respectively.

dependent variable. The IV coefficient on the share of middle-class households as reported in column (5) is 0.358 and it is significant at the 1%-level. This implies that a one percent increase in the share of the middle class induces an 0.358% increase in average years of schooling. Accordingly, the results obtained from the IV specification support the findings of the above OLS specification. Moreover, in the IV specification it appears that the obtained coefficients on the share of middle-class households are far less sensitive to the inclusion of additional controls as compared to the baseline OLS results.

Turning to panels C and D, we now discuss the insights obtained from the sample split into men and women. As with the overall education measure as dependent variable, there is no indication of weak instrument bias in any of the specifications. While the coefficient on the middle-class share is slightly smaller for the male sample as compared to the overall sample, with a magnitude of 0.855 in column (5), the impact of an increase in the share of middle-class households is more than twice as high in the female sample. This suggests that the effect of a larger middle class on female education is far higher as compared to the effect on male education. One potential explanation is that middle class households not only put a special emphasis on education in general but on female education specifically. Keeping in mind the paternalistic structure of the Indian society, another potentially important channel comes to mind: assortative mating, a well-researched regularity found throughout most of the western developed countries. The basic finding of this strand of the literature is that individuals are likely to marry spouses who carry similar characteristics across a wide variety of aspects (Lefgren & McIntyre, 2006). With regards to labour market characteristics and educational attainment, Hout (1982) finds a strong association between husbands' and wifes' occupational statuses. In addition, Cancian, Danziger & Gottschalk (1993) and Juhn & Murphy (1997) find that women with promising labour market characteristics are more likely to marry men with high wages, while Pencavel (1999) shows that those trends appear to have intensified over the second half of the last century in the developed world. This notion is supported with updated data for the time span from 1962 to 2003 in a more recent contribution by Schwartz & Mare (2005).

While the assortative mating channel is of large importance in the western world, the evidence regarding the Indian sub-continent remains relatively scarce. In a case study focussing on the Tamil Brahman subcaste of the "Eighteen Village Vattimas", Fuller & Narasimhan (2008) find that education and employment have become the crucial criteria for the arrangement of marriages in recent decades. They claim that for men who usually marry in their late twenties to early thirties, education and current or prospective labour market outcomes are by far the most crucial factors determining the assortative mating potential. With regards to women they find that the factors impacting the prestige ranking are basically the same. The main differences the authors report are that this is less a question of potentially higher expected incomes than rather focussed on matching the spouse's education. Namely, women with higher educational attainment are seen as "more congenial partners for educated men" (Fuller & Narasimhan, 2008). Also, the authors conclude that women may actively participate in the labour market until they give birth to children and thereby at least temporarily contribute to household (market) income. In addition, more educated women are seen as better qualified to assist in their children's education. To sum up, Fuller & Narasimhan (2008) conclude the every Vattima interviewee inevitably discusses individual education and career perspectives when discussing grown-up children and their marriage prospects. Pache-Huber (2004) describes similar findings for the middle-class Maheshwaris in Rajasthan. Our results support all these findings.

In order to provide some insights on how our results might differ once we compare rural to urban India, we present the results obtained from running the IV regression of Equation (3.3.1) on rural and urban sub-samples in Tables 3.3.6 and 3.3.7, respectively.

The first-stage results for the rural sample reported in Panel A of Table 3.3.6 indicate that the impact of the share of OBC households on the share of middle-class households is of similar magnitude as compared to the full sample. As before, it appears that there is no weak instrument bias in this specification as both the Kleibergen-Paap statistic as well as the F-statistic are above the critical values.

Comparing the second-stage results as reported in columns (1) through (3) to their counterparts in Table 3.3.5, we observe that the causal effect of increasing the size of the middle

Table 3.3.6. *Instrumental Variables Results: Rural Sample*

Dep. Var.:	(1) Education (full)	(2) Education (male)	(3) Education (female)
Panel A: First Stage - Share of MC Households on Share of OBC Households			
OBC Share	0.105***	0.154***	0.154***
	(0.023)	(0.056)	(0.056)
Panel B: Second Stage - Education			
Middle Class	0.517***	0.486***	1.171***
	(0.169)	(0.167)	(0.369)
Kleibergen-Papp F Stat.	19.97	19.97	19.97
ARW F Stat.	13.05	11.06	12.34
Observations	1,320	1,320	1,320
Controls			
District FE	YES	YES	YES
Income	YES	YES	YES
Dist. School	YES	YES	YES
Dist. Bus	YES	YES	YES

Note: The table reports IV estimates of Equation (3.3.1) with alternating education variables as a dependent variable. Urban districts are excluded. Robust standard errors are reported in parentheses, Robust standard errors are reported in parentheses, where *, **, and *** indicate significance at 1%, 5%, and 10% level, respectively.

class by 1% is to increase average years of formal schooling in rural India by 0.517%. This implies that the effect of middle-class shares in a rural setting is about 44% larger as compared to the full sample. Comparing the coefficients on middle class size obtained from running the rural regression taking the male and female education as dependent variable as reported in columns (2) and (3), we observe a similar pattern as on the full sample. While the coefficient obtained when focussing on male education is slightly smaller than in the average education case, the coefficient in the female education scenario is more than twice as large in magnitude as the coefficient obtained from the male sample. We take those findings as evidence that the motives that drive the differences between men and women are stronger in rural parts of India. This is hardly surprising taking the realistic assumption that urban areas are characterised by a less traditional and a more egalitarian society relative to their rural counterparts.

Table 3.3.7. *Instrumental Variables Results: Urban Sample*

	(1)	(2)	(3)
Dep. Var.:	Education (full)	Education (male)	Education (female)
First Stage - Share of MC Households on Share of OBC Households			
OBC Share	0.154***	0.154***	0.154***
	(0.056)	(0.056)	(0.056)
Middle Class	0.137**	0.124**	0.194*
	(0.059)	(0.051)	(0.102)
Kleibergen-Papp F Stat.	7.645	7.645	7.645
ARW F Stat.	3.917	4.271	2.842
Observations	645	645	645
Controls			
District FE	YES	YES	YES
Income	YES	YES	YES
Dist. School	YES	YES	YES
Dist. Bus	YES	YES	YES

Note: The table reports IV estimates of Equation (3.3.1) with alternating education variables as a dependent variable. Rural districts are excluded. Robust standard errors are reported in parentheses, where *, **, and *** indicate significance at 1%, 5%, and 10% level, respectively.

Summing up our findings from the different instrumental variable estimations, we find that the share of middle-class households in Indian villages has a sizeable and robust positive effect on educational outcomes. This effect is especially pronounced when focussing on female educational attainment. In addition, we find that the size of the middle class appears to have a stronger impact on educational outcomes in rural India.

3.4 Conclusion

We present a stylised household consumption model in which heterogeneous individuals can chose between three different categories of goods: subsistence consumption needs, education of children, and luxury goods. We show that the poor spend most of their income on subsistence consumption needs and the rich spend a positive (and potentially large) part of their budget on luxury goods such that the middle class has the highest expenditure share on education. Depending on class-specific differences, this provides a plausible pathway by which a rise in the relative size of the middle class raises the share of spending on education to the extent that educational outcomes depend positively on the share of the middle class. To test this hypothesis empirically, we use detailed survey data on Indian household incomes, educational attainment, and important control variables drawn from the household survey and village surveys.

In order to test the causal effect of the share of middle-class households on average educational attainment per village, we use detailed information on the shares of different castes according to the *Varna*-system. We use those shares in a – to our knowledge – novel instrumental variable specification. Our empirical analysis shows that larger shares of middle-class households in Indian villages indeed have a sizeable positive effect on average educational outcomes. Our results suggest that this effect is more pronounced in rural settings as compared to urban areas and for women as compared to men.

Altogether our results emphasize the importance of a sizeable middle class for education, and, via this pathway, potentially on other socioeconomic outcomes such as income growth and democratization. Therefore, it appears to be a warning sign if the share of the middle class shrinks and a larger part of the population belongs to the poor or to the rich.

Appendix

3.A The First-Order Conditions

The Lagrangian for the optimization problem is

$$\mathcal{L} = \omega_s \log(c_{s,t} - \bar{c}_s) + \omega_h \log(h_t) + \omega_l \log(c_{l,t} + \bar{c}_l) + \lambda_t(w_t - c_{s,t} - h_t - p_{l,t}c_{l,t}).$$

The associated necessary first-order conditions for an interior optimum are given by

$$\frac{\partial \mathcal{L}}{\partial c_{s,t}} = \frac{\omega_s}{c_{s,t} - \bar{c}_s} - \lambda_t \stackrel{!}{=} 0, \tag{3.4.1}$$

$$\frac{\partial \mathcal{L}}{\partial h_t} = \frac{\omega_e}{h_t} - \lambda_t \stackrel{!}{=} 0, \tag{3.4.2}$$

$$\frac{\partial \mathcal{L}}{\partial c_{l,t}} = \frac{\omega_l}{c_{l,t} - \bar{c}_l} - \lambda_t p_{l,t} \stackrel{!}{=} 0 \tag{3.4.3}$$

and the budget constraint $w_t = c_{s,t} + h_t + p_{l,t}c_{l,t}$. In this case, we have four equations to solve for the four unknowns $c_{s,t}$, h_t, $c_{l,t}$, and λ_t. Solving the corresponding system of equations for $c_{s,t}$, h_t, and $c_{l,t}$ yields

$$c_{s,t} = \frac{\omega_s(w_t + p_{l,t}\bar{c}_l) + (\omega_h + \omega_l)\bar{c}_s}{\omega_s + \omega_h + \omega_l}, \tag{3.4.4}$$

$$h_t = \frac{\omega_h(w_t + p_{l,t}\bar{c}_l - \bar{c}_s)}{\omega_s + \omega_h + \omega_l}, \tag{3.4.5}$$

$$c_{l,t} = \frac{\omega_l(w_t - \bar{c}_s) - p_{l,t}(\omega_h + \omega_s)\bar{c}_l}{p_{l,t}(\omega_s + \omega_h + \omega_l)}. \tag{3.4.6}$$

These results hold for an income level w_t for which the numerator of Equation (3.4.6) is positive which is the case as long as $w_t > p_{l,t}(\omega_h + \omega_s)\bar{c}_l/\omega_l + \bar{c}_s$. In the following, we denote the income level for which this expression is fulfilled with equality by $\hat{w}_t$. In case of a lower income level than $\hat{w}_t$, households do not consume luxury goods and face the following Lagrangian:

$$\mathcal{L} = \omega_s \log(c_{s,t} - \bar{c}_s) + \omega_h \log(h_t) + \lambda_t(w_t - c_{s,t} - h_t).$$

The necessary first-order conditions for this optimization problem are given by

$$\frac{\partial \mathcal{L}}{\partial c_{s,t}} = \frac{\omega_s}{c_{s,t} - \bar{c}_s} - \lambda_t \overset{!}{=} 0, \tag{3.4.7}$$

$$\frac{\partial \mathcal{L}}{\partial h_t} = \frac{\omega_e}{h_t} - \lambda_t \overset{!}{=} 0 \tag{3.4.8}$$

and the modified budget constraint $w_t = c_{s,t} + h_t$. In this case, we have three equations to solve for the three unknowns $c_{s,t}$, h_t, and λ_t. Solving the corresponding system of equations for $c_{s,t}$ and h_t yields

$$c_{s,t} = \frac{\omega_s w_t + \omega_h \bar{c}_s}{\omega_s + \omega_h}, \tag{3.4.9}$$

$$h_t = \frac{\omega_h (w_t - \bar{c}_s)}{\omega_s + \omega_h}. \tag{3.4.10}$$

We assume that incomes are sufficiently high so as to fulfil the basic subsistence consumption needs, i.e., it holds that $w_t > \bar{c}_s$. Altogether, we can therefore summarize our findings by means of the following system of demand functions

$$c_{s,t} = \begin{cases} \frac{\omega_s w_t + \omega_h \bar{c}_s}{\omega_s + \omega_h} & \text{for } \bar{c}_s < w_t < \hat{w}_t, \\ \frac{\omega_s (w_t + p_{l,t}\bar{c}_l) + (\omega_h + \omega_l)\bar{c}_s}{\omega_s + \omega_h + \omega_l} & \text{for } \hat{w}_t < w_t, \end{cases}$$

$$h_t = \begin{cases} \frac{\omega_h (w_t - \bar{c}_s)}{\omega_s + \omega_h} & \text{for } \bar{c}_s < w_t < \hat{w}_t, \\ \frac{\omega_h (w_t + p_{l,t}\bar{c}_l - \bar{c}_s)}{\omega_s + \omega_h + \omega_l} & \text{for } \hat{w}_t < w_t, \end{cases}$$

$$c_{l,t} = \begin{cases} 0 & \text{for } \bar{c}_s < w_t < \hat{w}_t, \\ \frac{\omega_l (w_t - \bar{c}_s) - p_{l,t}(\omega_h + \omega_s)\bar{c}_l}{p_{l,t}(\omega_s + \omega_h + \omega_l)} & \text{for } \hat{w}_t < w_t. \end{cases}$$

Since the Lagrangian is strictly concave because the utility function is strictly concave in all three arguments and the budget constraint is linear, the first-order conditions are not only necessary but also sufficient. Thus, they identify the global unique optimal choice.

3.B Data Appendix

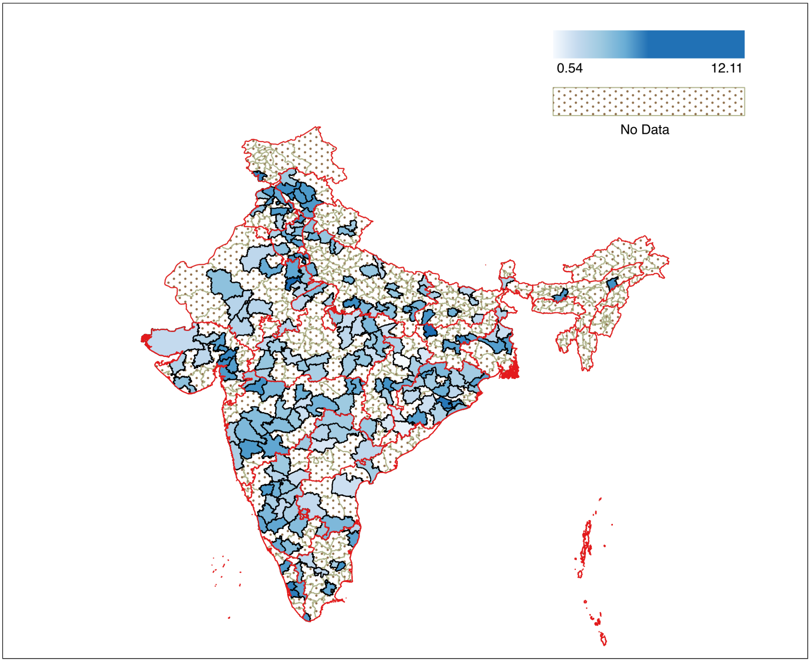


Figure 3.B.1. *District-Level Mean of Education in Years (Male Sample).*

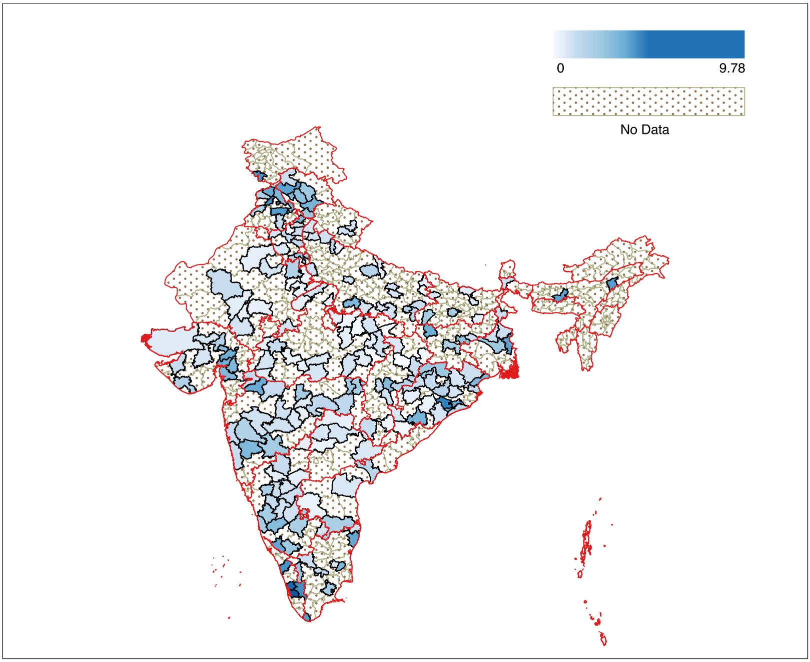


Figure 3.B.2. *District-Level Mean of Education in Years (Female Sample).*

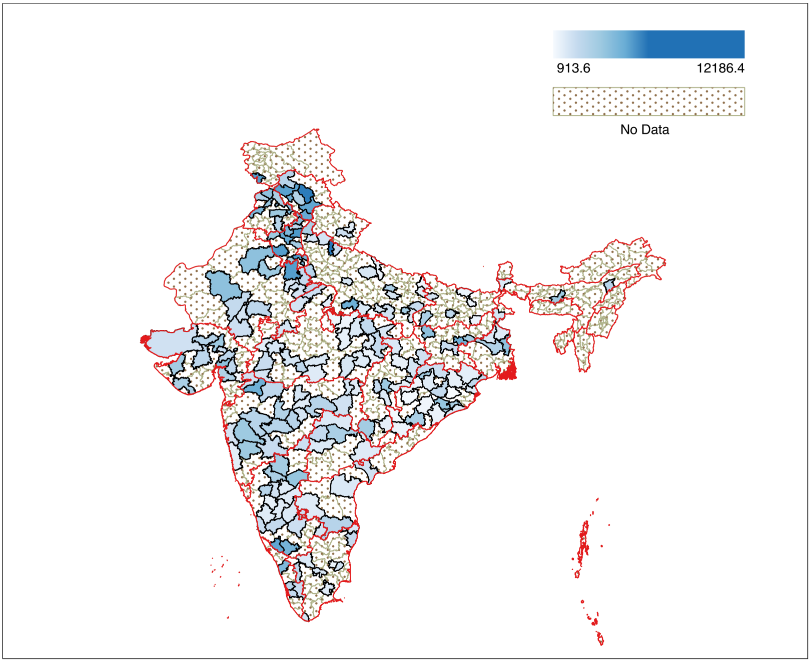


Figure 3.B.3. *District-Level Mean of Household Income per capita (in INR).*

Bibliography

Alesina, A. & Perotti, R. (1996). Income distribution, political instability, and investment. *European Economic Review*, *40*(6), 1203–1228.

Andreoni, J. (1989). Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence. *Journal of Political Economy*, *97*(6), 1447–1458.

Banerjee, A. V. & Duflo, E. (2008). What is Middle Class about the Middle Classes around the World? *Journal of Economic Perspectives*, *22*(2), 3–28.

Barro, R. J. (1999). Determinants of democracy. *Journal of Political Economy*, *Vol. 107*(No. 6), S158–S183.

Birdsall, N., Graham, C., & Pettinato, S. (2000). *Stuck In The Tunnel: Is Globalization Muddling The Middle Class?* Center on Social and Economic Dynamics. Washington, DC.

Bloom, D. E., Kuhn, M., & Prettner, K. (2017). Invest in Women and Prosper. *Finance & Development*, *Vol. 54*, 50–55.

Cancian, M., Danziger, S., & Gottschalk, P. (1993). Working Wives and Family Income Inequality Among Married Couples. In *Uneven tides* (pp. 195–222). Rising Inequality in America. Russell Sage Foundation.

Cass, D. (1965). Optimum Growth in an Aggregative Model of Capital Accumulation: A Turnpike Theorem. *The Review of Economic Studies*, *32*(3), 233–240.

Castelló-Climent, A., Chaudhary, L., & Mukhopadhyay, A. (2017). Higher Education and Prosperity: From Catholic Missionaries to Luminosity in India. *The Economic Journal*.

Chauvel, L. (2013). Welfare Regimes, Chorts, and the Middle Class. In J. C. Gornick & M. Jäntti (Eds.), *Income inequality - economic disparitites and the middle class in affluent countries* (1st, pp. 115–144). Stanford: Stanford University Press.

Collier, P. (2004). Greed and Grievance in Civil War. *Oxford Economic Papers*, *56*(4), 563–595.

Diamond, P. A. (1965). National Debt in a Neoclassical Growth Model. *The American Economic Review*, *55*(5), 1126–1150.

Diebolt, C. & Perrin, F. (2013). From Stagnation to Sustained Growth: The Role of Female Empowerment. *American Economic Review*, *103*(3), 545–549.

Doepke, M. & Zilibotti, F. (2008). Occupational Choice and the Spirit of Capitalism. *Quarterly Journal of Economics*, *123*(2), 747–793.

Fuller, C. & Narasimhan, H. (2008). Companionate marriage in India: the changing marriage system in a middleâfffdclass Brahman subcaste. *Journal of the Royal Anthropological Institute*, *14*(4), 736–754.

Galor, O. (2005). From Stagnation to Growth: Unified Growth Theory. In *Handbook of economic growth* (Vol. 1, Part A, pp. 171–293). Elsevier.

Galor, O. (2011). *Unified Growth Theory*. Princeton University Press.

Galor, O., Moav, O., & Vollrath, D. (2009). {I}nequality in {L}andownership, the {E}mergence of {H}uman-{C}apital {P}romoting {I}nstitutions, and the {G}reat {D}ivergence. *The Review of Economic Studies*, *Vol. 76*, 143–179.

Galor, O. & Weil, D. N. (2000). Population, Technology, and Growth: From Malthusian Stagnation to the Demographic Transition and Beyond. *American Economic Review*, *90*(4), 806–828.

Ghurye, G. (1932). *Caste and Race in India*. London: Kegan Paul, Trench, Trübner.

Glaeser, E. L., La Porta, R., Lopez-de-Silanes, F., & Shleifer, A. (2004). {D}o {I}nstitutions {C}ause {G}rowth? *Journal of Economic Growth*, *Vol. 9*, 271–303.

Hout, M. (1982). The Association between Husbands' and Wives' Occupations in Two-Earner Families. *American Journal of Sociology*, *88*(2), 397–409.

Juhn, C. & Murphy, K. M. (1997). Wage Inequality and Family Labor Supply. *Journal of Labor Economics*, *15*(1, Part 1), 72–97.

Koopmans, T. (1965). On the concept of optimal economic growth. In *The econometric approach to development planning*. Amsterdam: North Holland.

Lefgren, L. & McIntyre, F. (2006). The Relationship between Women's Education and Marriage Outcomes. *Journal of Labor Economics*, *24*(4), 787–830.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*(1), 3–42.

Milanovic, B. & Yitzhaki, S. (2002). DECOMPOSING WORLD INCOME DISTRIBUTION: DOES THE WORLD HAVE A MIDDLE CLASS? *Review of Income and Wealth Series*, *48*(2).

Ostby, G. & Urdal, H. (2011). *Education and Civil Conflict: A Review of the Quantitative, Empirical Literature*, UNESCO.

Pache-Huber, V. (2004). *Le mariage de l'amour et de la raison: stratégies matrimoniales de la classe moyenne en Inde*. Münster: LIT Verlag.

Pencavel, J. (1999). *Assortative Mating by Schooling and the Timing of Marriage*, Department of Economics, Stanford University.

Prettner, K. & Seiffert, S. (2018). *The Size of the Middle Class and Educational Outcomes : Theory and Evidence from the Indian Subcontinent*, University of Hohenheim, Institute of Economics. Stuttgart.

Prettner, K. & Strulik, H. (2017). Gender equity and the escape from poverty. *Oxford Economic Papers*, *69*(1), 55–74.

Ravallion, M. (2010). The Developing World's Bulging (but Vulnerable) Middle Class. *World Development*, *38*(4), 445–454.

Romer, P. M. (1990). Endogenous Technological Change. *Journal of Political Economy*, *98*(5), S71–S102.

Schwartz, C. R. & Mare, R. D. (2005). Trends in Educational Assortative Marriage From 1940 to 2003. *Demography*, *42*(4), 621–646.

Solow, R. (1956). A Contribution to the Theory of Economic Growth. *Quarterley Journal of Economics*, *70*(1), 65–94.

Stock, J. & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. W. K. Andrews (Ed.), *Identification and inference for econometric models* (pp. 80–108). New York: Cambridge University Press.

Strulik, H., Prettner, K., & Prskawetz, A. (2013). The past and future of knowledge-based growth. *Journal of Economic Growth*, *18*(4), 411–437.

Tadjoeddin, M. Z. & Murshed, S. M. (2007). Socio-Economic Determinants of Everyday Violence in Indonesia: An Empirical Investigation of Javanese Districts, 1994â"2003. *Journal of Peace Research*, *44*(6), 689–709.

Thorow, L. C. (1987). A Surge in Inequality. *Scientific American*, *256*(5), 30–37.

Vannemann, R. & Dubey, A. (2013). Horizontal and Vertical Inequalities in India. In J. C. Gornick & M. Jäntii (Eds.), *Income inequality - economic disparitites and the middle class in affluent countries* (1st, Chap. 16, pp. 439–458). Stanford: Stanford University Press.

Chapter 4

Go East: On the Impact of the Transsiberian Railway on the Economic Development of Eastern Russia[1]

4.1 Introduction

A sizeable amount of the annual World Bank budget is being allocated towards transport infrastructure improvements year by year. High trade costs are seen as key impediment to local economic development. It has been shown by a large range of theoretical contributions that reductions in trade costs result in higher levels of real income in trading as well as alleviate the impact of shocks (Donaldson, 2018). But can such investments in infrastructure be seen as the go-to method in order to foster local development? Is it not likely that also contexts, i.e. the specific local circumstances such as existing spatial equilibria or local natural endowments also matter? It is much likely that the impact of local measures is affected by the level of local attributes such as resource availability or the suitability for agriculture. While we have a good understanding of the theoretical mechanisms behind the nexus of trade costs and economic development, the empirical literature to this day remains quite sparse.

The historically most prominent and wide-spread strategy in order to reduce transport costs is the construction of railroad networks. It has been widely used in both the transfor-

[1] Single-authored. Published as Seiffert (2019)

mation of western countries towards industrialised economies as well as in the annexation and exploitation of countries in the era of colonialism. This opens the question if the spatial distribution of economic activity in such contexts is entirely determined by local natural advantages or if lower costs faced in accessing larger markets play decisive role. The main challenge in order to empirically assess the existence of said transport cost advantages is to separate the effects of state dependence which is the local availability of factors of production from those of local differences in trade costs and with it differences in the accessibility of remote markets.

In the underlying paper, I investigate the long-run causal effect of the construction of the Trans-Siberian Railway (henceforth: TSR) and the implied reduction in local transport costs on the level and spatial organisation of economic activity in Asian Russia. The assessment is based on data nocturnal lights emission, the historical and current network of railways in Eastern Russia as well as a wide array of control variables.

In order empirically disentangle the effects of local natural endowments from the effect of the TSR, I suggest to solve this endogeneity problem by employing data on historical tea trade in Tsarist Russia. The so-called Tea Road connected European Russia with first Nerchensk in Manchuria and later Kyaktha located on what is now the border between Russia and Mongolia. As there were hardly any local markets east of the Ural mountains at the time and the fact that the route was not suitable for the transport of large amounts of heavy goods as well as persons, the route was a mere transit route. its main purpose was to link the Chinese tea production sites with the tea markets in European Russia. This route was later extended to Vladivostok and used as a post route in addition to its original purpose. While the area which today constitutes the districts of Ural (in parts), Siberia and the Far East came under Russian control as early as the late 17th century, the population remained below 300.000 until the early 20th century. The construction of the TSR set in motion a large-scale settling process which lead to an increase in the population up to about 36 million in current times[2]. Accordingly, upon its completion the TSR facilitated both the colonisation of Eastern Russia as well as created a strong focal

[2] This implies an about 10 times stronger increase in the population of Eastern Russia as compared to European Russia over the same time period.

line for economic activity due to an early change in transport costs along its course. Hence, Asian Russia offers a research setting which - to my knowledge - is unique. Using this historical trade route as an instrument, I show that being located further away from the TSR has a causal and negative impact on contemporary economic activity as recorded by nocturnal lights emission. I further create a new data set linking information on railroads, historical (trade) routes, economic activity and local natural endowments at a high spatial granularity of 132,843 grid cells with the size of 0.1x0.1 decimal degrees which is equivalent to 11x11km.

I show how the construction of the TSR and the associated reduction in transport costs facilitated the colonisation of the Russian East. Until its construction the area it now transects was very scarcely populated and did not exhibit any noteworthy economic activity. Within the following 100 years the population in its vicinity highly increased and today there exists sizeable economic activity. By providing a novel instrumental variable estimation framework, I am able to demonstrate the causal effects of the reduction in transport costs.

In doing so, I contribute to the ongoing scientific discussion about the causal impact of large-scale infrastructure projects. Namely, to answering the question whether a large-scale transport infrastructure project and the implied reduction in trade costs have a causal impact on local economic growth. This is possible since with the underlying setup I am able disentangle the effects of such an investment from the ones of local endowments in a context of non-existing scale effects. Choosing the TSR is mainly driven by the fact that its specific location was chosen exogenously to local natural endowments. Further, it was built in an area lacking significant pre-existing populations which mattered at the time of its construction. As it appears, its location was solely determined by the course of the already existing tea and postal routes.

The remainder of the paper is structured as follows: I will start-off with summarising the related literature. Then, I will present a detailed description of the underlying data, the history of the TSR as well as the historical trade and post routes linking Europe and China. Third, I will present the empirical models and the results. Section 4.5 concludes.

4.2 Related Literature

With the underlying paper I combine findings from the literature on natural endowments, path dependence and transport infrastructure. I contribute to a deeper understanding of the question to which extend large scale transport infrastructure investments have a causal impact on subsequent development as well as the spatial organisation of economic activity. Scholars attributable to the former two branches investigate whether increasing returns or locational fundamentals determine the spatial equilibria of economies. Locational fundamentals being the sole determinant of where economic activity is concentrated would imply that (transitory) localised shocks, e.g. transport cost reductions, do only have short-run consequences which are set off in the long run.

In a recent article, Miguel & Roland (2011) assess the impact of the US bombings during the Vietnam War on subsequent economic activity and find no long-run effect. Their findings support the results of a closely related contribution by Davis & Weinstein (2002) who examine the effects of WW2 bombings in Japan. Another contribution coming to a similar conclusion is Davis & Weinstein (2008). On the other hand, there is a considerable range of contributions which find evidence for transitory localised shocks to persistently define spatial equilibria in the presence of increasing returns. In a recent contribution, Rauch & Michaels (2013) show that many French towns are still situated in Roman-Age town locations in spite of the fact that those locations often exhibit locational properties inferior to alternatives. Using the construction and the later demise of colonial railroads in Africa, Jedwab & Moradi (2016) show that the associated regional reduction in transport costs had a significant effect on the spatial equilibria during the operation of the railroads. This effect also persisted after their demise. Analysing the spatial equilibrium of the US, Bleakley & Lin (2012) find that modern US cities are often still located close to waterfalls which in the past either caused the foundation of portage cities or provided electricity. Despite the fact that those locational advantages are now obsolete, those cities persisted. Further studies in support of the existence multiple spatial equilibria are among others: Bosker, Brakman, Garretsen & Schramm (2007, 2008), Redding, Sturm & Wolf (2011), Bleakley & Lin (2012), Rauch & Michaels (2013) and Ahlfeldt, Redding, Sturm & Wolf

(2015).

While both local endowments as well as transitory shocks have been found to have significant effects on local development, there also exists a growing literature which puts these two forces into contrast with large scale infrastructure projects and their lasting effects. Fogel (1964) pioneered this field of research in being one of the first to apply the social savings methodology to the transport infrastructure field by assessing the impact of railroads on local economic dynamics in the US. Hurd (1983) applied a similar framework to India (Donaldson (2018)). The empirical literature regarding the issue was pioneered by Aschauer (1989). The author was one the first to estimate the relationship between aggregate productivity and stock and flow government spending. Among other things, his findings suggest decisive explanatory power of what he calls core infrastructure projects like highways or mass transit. Duflo & Pande (2007) show that the construction of dams in India increases the agricultural output of downstream districts[3]. In a recent contribution, Donaldson (2018) shows that the construction of the colonial railroad during the British Raj decreased trade costs and increased interregional as well as international trade in India.

In the identification of the causal effects of transport infrastructure on economic development, researchers face a serious problem in the light of the findings which have been laid out previously: the fact that oftentimes locations which are characterised by either strong localised natural advantages or the presence of increasing returns often exhibit better access to transport networks relative to their counterparts (Duranton et al. (2014)). This opens the issue of endogeneity in econometric models. I contribute to the understanding of the effects of local infrastructure investments by introducing a novel instrumental variable strategy and thereby solving the problem of endogeneity for the underlying empirical model. This allows me to clearly identify the causal impact of the construction of the TSR on local economic development in Russia.

[3] Further contributions assessing the effects of infrastructure projects in different context are: Jensen (2007), Michaels (2008), Dinkelman (2011) and Duranton, Morrow & Turner (2014).

4.3 Data and Historical Background

Due to the unavailability of data on economic activity in Eastern Russia, I employ data on nocturnal lights emission as provided by the US Airforce's Defence Meteorological Satellite Program (DMSP). Previous studies have established this data as a valid source of information about economic activity (see, among others: Michalopoulos & Papaioannou (2013b), Storeygard (2016) and Sala-i-Martin & Pinkovskiy (2010)). I combine this data with information on the contemporary rail network of Russia as well as with historical data on the tea trade and post routes of Tsarist Russia prior to 1830. My analysis is focused on the part of modern Russia which lies east of the Ural mountains[4]. The reason for that is the fact that this part of modern Russia was - while being largely controlled by Russia - home to less than 300,000 inhabitants until the early 20th century. The data is pre-processed in the following steps:

4.3.1 Lights Data and Local Economic Activity:

As there is hardly any reliable information on the location and population of settlements except for the major cities in the Russian East, I follow Jedwab & Moradi (2016) and construct a new data set based on a 0.1x0.1 decimal degree cell grid which covers the entire Russian East. I chose to focus on the area east of the Ural mountains[5] since this area remained largely untouched until the construction of the TSR. Second, as I am interested in the causal effect of the TSR, I further drop all grid cells outside a 500km buffer around the contemporary TSR network and the historical TSR main line, respectively. This leaves me with a data set consisting of about 133,000 grid cells in the contemporary TSR case and roughly 45,000 grid cells in the TSR main line scenario, respectively. As a measure of local economic activity, I extract the mean level nocturnal lights emission per grid cell[6]. As a measure for the local spatial organisation I use the standard deviation in illumination intensity between the 0.1 decimal degree cells which are coded from 0 (no lights) to 63

[4] More specifically, I focus on the area east of the 60.5 longitude line.
[5] This is approximated by dropping all grid cells west of 60.5 degrees longitude.
[6] Please refer to appendix 4.5 for a detailed description of the underlying data.

(sensor satiation) within a 11x11km grid cell.

4.3.2 Contemporary and Historical Russian Railroad Network:

The - to my knowledge - most comprehensive and accurate as well as freely obtainable shapefile of the Russian Railroad network is provided by diva-gis.org. This shapefile is used in order to compute the shortest distance from each grid cell centroid to the network using ESRI's NEAR tool in arcpy in the full network scenario. Figure 4.3.1 illustrates the underlying algorithm[7]. In order to identify the historical mainline of the TSR, I use the stops summarised in the TSR Wikipedia article[8]. Figures 4.3.2 and 4.3.3 illustrate the respective railroad networks.

4.3.3 Local (Natural) Advantages:

The empirical model laid out in section 4.4 encompasses several controls for initial local advantages which may act as potential confounding factors. As with the light intensity figures, these data have been extracted for the aforementioned grid cells. These are:

Caloric potential:

the mean local average caloric potential according to the Calorics Suitability Index[9] (CSI) introduced by Galor & Özak (2016) in order to control for local advantages in crop production. This data has been established as valid measure for long run locational advantages by a large number of studies attributable to the Unified Growth Theory spearheaded by Galor (2011)[10]. As with the night lights data, it is provided in form of raster file out of which I extracted the mean optimal caloric potential per grid cell [11].

---

7 For a detailed description, visit http://pro.arcgis.com/en/pro-app/tool-reference/analysis/near.htm.

8 The article can be accessed here: https://en.wikipedia.org/wiki/Trans-Siberian_Railway.

9 The CSI data can be obtained from https://ozak.github.io/Caloric-Suitability-Index/.

10 Prominent examples are: Michalopoulos & Papaioannou (2013b), Michalopoulos & Papaioannou (2013a) and Alesina, Giuliano & Nunn (2013).

11 "Optimal" meaning the caloric yield of cultivating the crop which is best suited for the agro-climtatic properties of the underlying area.

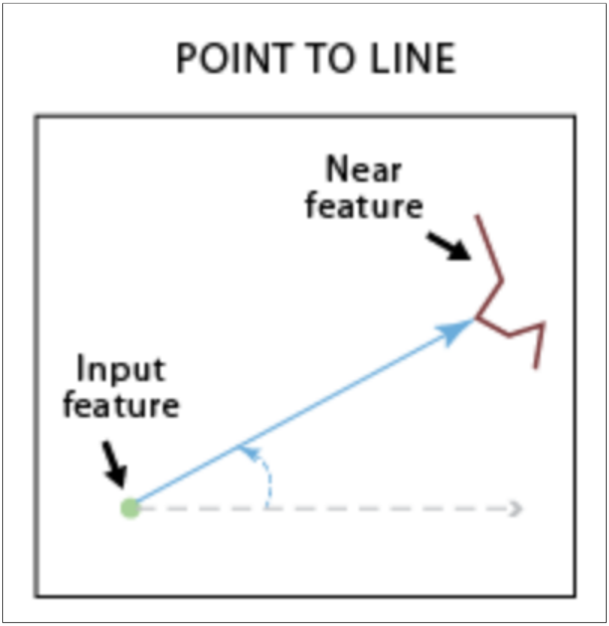


Figure 4.3.1. *Illustration of the NEAR Algorithm by ESRI.*

Flare Distance:

The minimum distance to the main areas of crude oil and natural gas extraction[12] in order to control for advantages in resource exploitation.

Precipitation:

While the average level of precipitation per year usually positively affects the local conditions for crop production on the northern hemisphere, the standard deviation might have adverse effects on both crop production as well settlement suitability due to flooding.

[12] There is no reliable data on the exact location of crude oil and gas extraction sites in Russia available. So I use shapefiles provided by DMSP which encompass the area of gas flares to approximate their position. The shapefiles were obtained from https://ngdc.noaa.gov/eog/interest/gas_flares.html.

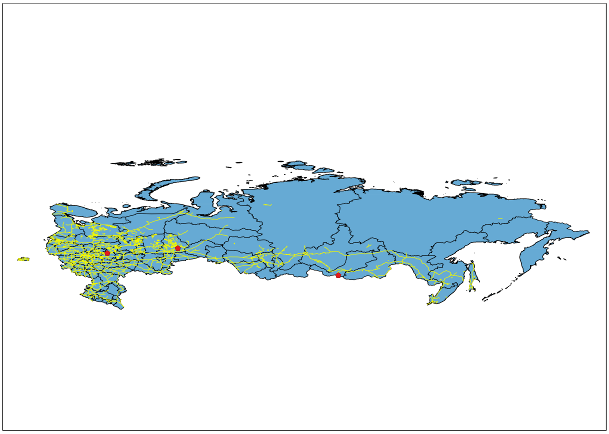


Figure 4.3.2. *The contemporary TSR network.*

The data is extracted from rasters of annual data averaged over 1900-2008[13]. Hence, both measures are included as controls.

Road density:

Since local transport infrastructure other than access to the TSR might distort the measurement of the causal effect on economic activity, I include the contemporary road density per grid cell as a control variable for local transport costs. This measure is derived from a shapefile encompassing the contemporary road network in Russia[14].

Population density:

Aiming to isolate the effect of TSR proximity on economic activity and agglomeration, I control for the population density per grid cell. This measure is derived from the Gridded

[13] The data is provided http://climate.geog.udel.edu/~climate/html_pages/download.html.

[14] The data is provided here: http://www.diva-gis.org/gdata. I create a density raster based on the shapefile using the Calculate-Density-tool in arcpy. Then I extract the mean road density per grid cell from this raster.

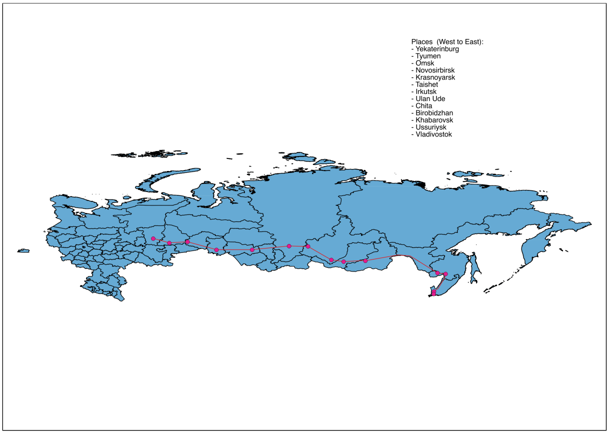


Figure 4.3.3. *The historical TSR mainline.*

Population of the World raster data[15].

The Russian Tea Road

Tea is one if not the main historical commonality between Russia, China and Central Asia. One of the historically most important trade ports in general and the most important one for the trade between Russia and China instead of being located at the sea shore - as one would intuitively assume taking the more prominent British tea trade as a reference - was located at the Sino-Russian border. The town of Kyakhta was defined as the exclusive border market by the Kyakhta treaty of 1727. Until the outbreak of the Opium War in 1840, Kyakhta and the seaport of Canton were the most prominent Chinese foreign markets (Lee, 2014). In this period, tea was distinguished into two different varieties: Overland Tea which was the one transported via the Russian Tea Route and Canton Tea which was transported by ship from Canton to Europe via the Indian Ocean. The Russian consumers considered the Overland quality to be far superior compared to the sea-borne Canton Tea. This had two reasons. On the one hand, Canton

[15] The data is provided here: http://sedac.ciesin.columbia.edu/data/collection/gpw-v4.

Tea was exposed to hot climate throughout the travel across the Indian Ocean which was said to cause the tea develop an aroma much different to the Overland quality. The transport route via the cold and dry deserts of the Russian Tea Route on the other hand was held to be enhancing the taste of the Overland alternative. Lee (2014) states:

> "The unpleasant taste from firing (*bei*) is removed (from tea) by transit through the cold dry climate of Mongolia and Siberia, and at the same time tea, which is nightly unloaded from the camel's back and placed upon the snow-covered steppes, is found to acquire, from light moisture it then absorbs, a delicacy of flavour obtainable in no other way, and it brings, in consequence, a much more lucrative price in the markets of Russia."

Accordingly, there was a strong preference for Overland Tea in the Russian markets where tea was usually consumed without adding milk. While the tea exported to Russia was produced in different provinces throughout China, its entirety was transported to Kyakhta by Chinese traders from where it then was transported to European Russia by camels and oxen cars operated by Russian merchants (Lee (2014)).
A second function the Russian Tea Route had was serving as a postal road which should ensure communication across the Russian sphere of influence in the east. It extended the Tea Route further east to Vladivostok. The Post Road travelled by Wenyon (1896) is depicted in figure 4.3.4[16]. Wenyon who travelled the Russian Post Route in 1893, makes a telling reference to his homeland England in the preface of his book:

> "The old post-roads of England have been superseded by the railway, and the same fate will soon befall the great post-road of Siberia."

The author describes the Siberian Route as a cordon of post-horse stations which were sixteen to twenty miles apart. They were installed by the Russian government for military

16 The basis for this simplified digitised version is the hand-drawn map by the author which can be found as figure 4.B.2 in appendix 4.5.

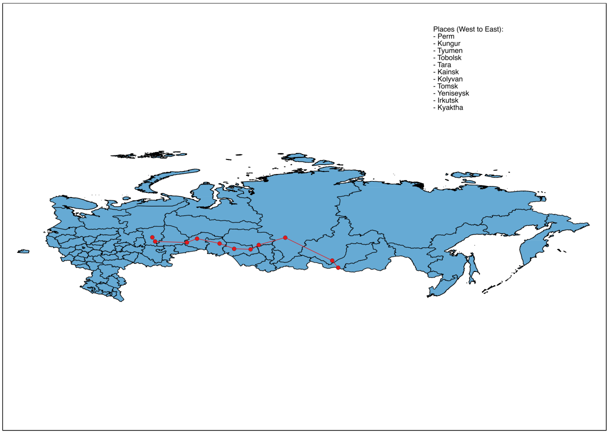


Figure 4.3.4. *The Tea Route in 19th Century Russia. (Own Depiction based on Avery (2003))*

purposes. Besides those stations he names only a few very scarcely populated places as sources of supply for his travel. He emphasises the notion of the route not being a properly built-up road but at best being a dirt track only defined by the post-horse stations. The author further emphasises his notion of the Route being by far not suitable for the transport of heavy goods or people.

The historical facts which have been laid out in this paragraph strongly support the appropriateness of using the Siberian Route in an instrumental variable approach in the context of my strategy to assess the causal long-run impact of the Russian Railroad on economic activity in Eastern Russia.

4.4 Econometric Specifications, Identification Strategies and Results

In this section, I will first lay out the basic empirical model used to give a first impression of the positive correlation between TSR proximity and economic activity using simple

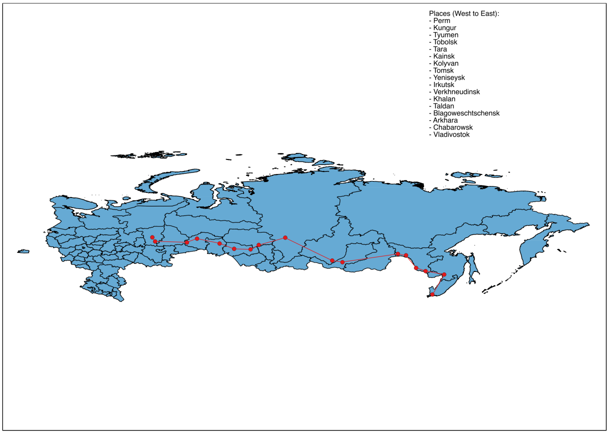


Figure 4.3.5. *The Post Route in 19th Century Russia. (Own Depiction based on Wenyon (1896))*

OLS. Second, I follow Jedwab & Moradi (2016) and use a spatial discontinuity framework to show that the effect found using OLS decrease in the distance to the TSR network. In a third and final step, I propose a novel instrumental variable approach in order to show the causal effect of TSR proximity on local economic development and agglomeration.

4.4.1 Baseline OLS Regression

The baseline regression model is defined in equations (4.4.1) through (4.4.4):

$$ln(Lights_i) = \beta ln(DistRail_i) + \chi X_i + \epsilon_i \tag{4.4.1}$$

$$ln(Lights_i) = \beta ln(DistMain_i) + \chi X_i + \epsilon_i \tag{4.4.2}$$

$$ln(Agglo_i) = \beta ln(DistRail_i) + \chi X_i + \epsilon_i \tag{4.4.3}$$

$$ln(Agglo_i) = \beta ln(DistMain_i) + \chi X_i + \epsilon_i \tag{4.4.4}$$

where $Lights_i$ is the mean nocturnal lights emission per location or grid cell and agglomeration which is measured by the standard deviation in illumination within each grid cell, respectively. $DistRail_i$ is the variable of interest. It is measured as the geodesic distance between the closest segment of the TSR and the centroid of the location or grid cell, respectively. X_i is a vector of control variables for both local natural advantages as well as current factors which might distort the identification of the local impact of the vicinity to a TSR segment. $DistMain_i$ is the analogue variable of interest in the scenario which focusses on the impact of the distance to the historical TSR mainline. ϵ_i is the error term. Tables 4.4.1 and 4.4.2 give an overview of the employed variables as well as their summary statistics. All variables are in logs with a small number added (0.001) to prevent grid cells with zero lights emission from dropping out of the sample[17].

Table 4.4.1. *Summary Statistics (TSR Scenario)*

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
DistRail	139,871	178.513	146.850	0.000	500.000
DistTea	139,871	1,089.965	881.771	0.004	3,432.420
DistPost	139,871	713.107	605.914	0.001	2,458.482
DistFlare	139,871	543.522	440.308	0.000	1,933.627
Lights	134,878	0.385	2.703	0.000	62.000
CalPot	137,930	1,378.965	1,921.020	0.000	10,827.360
PopDens	139,325	3.833	46.288	0.014	3,121.014
Precip	137,930	446.120	124.803	126.500	1,317.000
PrecipSD	137,930	0.397	1.126	0.000	27.290
DensRoa	139,695	1.962	2.076	0.000	10.282

Tables 4.4.3 through 4.4.6 report the results from the estimation of equations (4.4.1) through (4.4.4) using OLS with alternating the dependent variable from mean nocturnal lights emission to the nocturnal lights agglomeration. In both cases, the coefficient while decreasing in magnitude is smaller than zero as expected. The negative conditional correlation between the level of economic activity and the distance to the TSR exhibits

[17] This procedure has - among others - been used in Michalopoulos & Papaioannou (2013b) and is widely accepted in the night lights literature.

Table 4.4.2. *Summary Statistics (Mainline Scenario)*

	(1)	(2)	(3)	(4)	(5)
VARIABLES	N	Mean	SD	Min	Max
DistMain	46,140	236.053	143.524	0.001	499.991
DistTea	46,140	277.914	215.171	0.004	1,021.221
DistPost	46,140	207.084	136.814	0.001	672.323
DistFlare	46,140	475.086	279.277	0.000	1,174.234
Lights	45,283	0.744	3.513	0.000	61.618
CalPot	45,773	2,853.748	1,832.917	0.000	8,701.440
PopDens	45,806	8.755	74.513	0.022	3,121.014
Precip	45,773	453.260	103.088	238.000	1,029.000
PrecipSD	45,773	0.565	1.575	0.000	27.290
DensRoa	46,114	3.523	2.473	0.000	10.098

Table 4.4.3. *OLS Results (TSR Scenario)*

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Lights	Lights	Lights	Lights	Lights
DistRail	-0.543***	-0.540***	-0.524***	-0.477***	-0.471***
	(0.000)	(0.000)	(0.001)	(0.006)	(0.006)
CalPot		0.039***	0.003***	0.007***	0.004***
		(0.001)	(0.001)	(0.001)	(0.001)
PopDens			0.231***	0.249***	0.245***
			(0.005)	(0.005)	(0.005)
Precip				-0.086***	-0.095***
				(0.012)	(0.012)
PrecipSD					0.001
					(0.002)
RoadDens					0.026***
					(0.001)
Observations	134,878	133,167	132,843	132,843	132,843
R-squared	0.917	0.919	0.921	0.921	0.921

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports OLS estimates of equation (4.4.1). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the contemporary TSR network.

only slight changes in magnitude as a response to consecutively including additional controls. The change in magnitude of the coefficient of interest when comparing columns (1) through (6) of table 4.4.3 is below 15 percent. While varying in absolute values, all control variables exhibit the expected negative sign. Comparing the magnitude of the respective coefficients, it is obvious that the negative conditional correlation between the level of nocturnal lights emission and remoteness is the most pronounced. This strong persistence is not observable for the second variant using agglomeration as dependent variable. The coefficient on TSR remoteness decreases by roughly 85 percent comparing (1) through (6). This change is mainly triggered by the inclusion of the precipitation variable. While it does not break the correlation between agglomeration and TSR access, it strongly diminishes its absolute value. Further it appears that there exists a strong negative relationship between the level of precipitation and agglomeration per grid cell.

Table 4.4.4. *OLS Results (TSR Scenario)*

VARIABLES	(1) Agglo	(2) Agglo	(3) Agglo	(4) Agglo	(5) Agglo
DistRail	-0.525***	-0.523***	-0.504***	-0.154***	-0.141***
	(0.000)	(0.000)	(0.001)	(0.005)	(0.005)
CalPot		0.026***	-0.019***	0.004***	-0.002*
		(0.001)	(0.001)	(0.001)	(0.001)
PopDens			0.292***	0.422***	0.417***
			(0.006)	(0.006)	(0.006)
Precip				-0.651***	-0.662***
				(0.010)	(0.010)
PrecipSD					0.011***
					(0.002)
RoadDens					0.057***
					(0.001)
Observations	139,325	137,606	137,606	137,606	137,606
R-squared	0.892	0.895	0.899	0.903	0.903

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports OLS estimates of equation (4.4.3). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the contemporary TSR network.

As previously shown, the OLS results suggest a strong negative conditional correlation between the distance to the closest railroad segment and mean nocturnal lights emission as well as their spatial concentration. Tables 4.4.5 and 4.4.6 show the results which I receive from conducting the same OLS regression as above but changing the variable of interest to the distance to the TSR mainline. Comparing the coefficients on the variables of interest between the network scenario and the mainline scenario, we observe that changing the specification towards picking up the long run impact of the change in transport costs, they do not considerably change in magnitude if we focus on average night light emission per grid cell. This can be taken as evidence for a stable long run correlation between

Table 4.4.5. *OLS Results (Mainline Scenario)*

VARIABLES	(1) Lights	(2) Lights	(3) Lights	(4) Lights	(5) Lights
DistMain	-0.456***	-0.504***	-0.490***	-0.382***	-0.375***
	(0.001)	(0.001)	(0.001)	(0.017)	(0.017)
CalPot		0.106***	0.004***	0.007***	0.005***
		(0.001)	(0.002)	(0.002)	(0.001)
PopDens			0.811***	0.846***	0.836***
			(0.010)	(0.011)	(0.012)
Precip				-0.221***	-0.227***
				(0.034)	(0.034)
PrecipSD					0.012***
					(0.004)
RoadDens					0.026***
					(0.007)
Observations	45,283	44,944	44,636	44,636	44,636
R-squared	0.794	0.803	0.830	0.830	0.831

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports OLS estimates of equation (4.4.2). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

Table 4.4.6. *OLS Results (Mainline Scenario)*

VARIABLES	(1) Agglo	(2) Agglo	(3) Agglo	(4) Agglo	(5) Agglo
DistMain	-0.432***	-0.451***	-0.437***	-0.060***	-0.072***
	(0.001)	(0.001)	(0.001)	(0.017)	(0.017)
CalPot		0.038***	-0.069***	-0.058***	-0.054***
		(0.002)	(0.002)	(0.002)	(0.002)
PopDens			0.841***	0.960***	0.981***
			(0.013)	(0.014)	(0.015)
Precip				-0.769***	-0.742***
				(0.034)	(0.035)
PrecipSD					0.004
					(0.004)
RoadDens					-0.056***
					(0.010)
Observations	45,806	45,465	45,465	45,465	45,465
R-squared	0.758	0.763	0.795	0.798	0.799

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports OLS estimates of equation (4.4.4). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

low transport costs and aggregate economic activity per unit of observation. Looking at the two last columns of table 4.4.6, we observe that while diminishing in magnitude, the conditional correlation between vicinity to the TSR mainline and agglomeration of economic activity is negative in the mainline scenario but much less pronounced.

4.4.2 Further Assessment of the Spatial Equilibrium

After having established a negative conditional correlation between TSR remoteness and economic activity as well as its spatial organisation, I follow Jedwab & Moradi (2016) and estimate equations (4.4.5) through (4.4.8) in order to assess in how far the installation of the TSR established a specific spatial equilibrium. More specifically, I will lay out the diminishing effect of TSR as one moves further and further away from it.

$$ln(Lights_i) = \beta RailDu_i + \chi X_i + \epsilon_i \tag{4.4.5}$$

$$ln(Agglo_i) = \beta RailDu_i + \chi X_i + \epsilon_i \tag{4.4.6}$$

$$ln(Lights_i) = \beta MainDu_i + \chi X_i + \epsilon_i \tag{4.4.7}$$

$$ln(Agglo_i) = \beta MainDu_i + \chi X_i + \epsilon_i \tag{4.4.8}$$

while the dependent variables and the included controls are the same as in the baseline OLS specifications, $RailDu_i$ and $MainDu_i$ are cell dummies which capture the vicinity to the TSR and the TSR mainline, respectively. Those dummies are equal to one if the respective cell lies within 0-10, 10-20, 20-30 or 30-40km distance to the TSR or the TSR mainline, respectively.

Table 4.4.7 presents the results obtained from estimating equations (4.4.5) through (4.4.8). Columns (1) and (3) indicate a strong positive and significant effect on the level economic activity of a cell being located relatively close to the TSR or TSR mainline, respectively.

Table 4.4.7. *OLS Results (Vicinity Dummies)*

VARIABLES	(1) Lights	(2) Agglo	(3) Lights	(4) Agglo
0-10km Dummy (TSR)	2.5055***	6.9259***		
	(0.0688)	(0.7506)		
10-20km Dummy (TSR)	0.3875***	-0.3789		
	(0.0378)	(0.2426)		
20-30km Dummy (TSR)	0.1394***	-1.1176***		
	(0.0301)	(0.1176)		
30-40km Dummy (TSR)	0.0728***	-1.0330***		
	(0.0247)	(0.0868)		
CalPot	0.0000***	0.0001	0.0001***	0.0002*
	(0.0000)	(0.0001)	(0.0000)	(0.0001)
PopDens	0.0295***	0.4894***	0.0300***	0.5021***
	(0.0012)	(0.0402)	(0.0013)	(0.0448)
Precip	0.0002***	-0.0015***	-0.0006***	-0.0032***
	(0.0000)	(0.0001)	(0.0001)	(0.0008)
PrecipSD	-0.0020	0.0936	0.0175**	0.1148
	(0.0050)	(0.0615)	(0.0070)	(0.0939)
DistFlare	-0.0001***	0.0003***	0.0001*	0.0011***
	(0.0000)	(0.0000)	(0.0000)	(0.0004)
0-10km Dummy (Main)			2.3608***	13.6747***
			(0.2043)	(3.6104)
10-20km Dummy (Main)			1.5268***	6.9187***
			(0.0624)	(1.1122)
20-30km Dummy (Main)			-0.0579	-1.4132
			(0.0558)	(0.8698)
30-40km Dummy (Main)			-0.2619***	-2.0304***
			(0.0419)	(0.5446)
Observations	132,843	137,606	44,636	45,465
R-squared	0.3734	0.5350	0.5291	0.5462

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports OLS estimates of equations (4.4.5) through (4.4.8). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR network or the historical TSR mainline, respectively.

This effect diminishes strongly moving towards cells located further away from the TSR. Same holds true for the effect on spatial agglomeration of economic activity within cells as illustrated in columns (2) and (4).

The results presented in this section further substantiate my findings from section 4.4.1. Still, these results cannot be seen as proof of a causal impact of the closeness to the railroad on economic activity and its spatial organisation within a grid cell as there remains a possibility for the existence of endogeneity. Hence, I propose a novel instrumental variable specification in order to verify the existence of such a relationship. As laid out in section 4.3, the Transsiberian Routes existed long before the construction of the TSR mainline. And even though, there clearly would have been feasible alternative routes for the TSR since prior to it starting its services the area which is now fairly densely populated was basically empty of people there has been no attempt to relocate it. While the TSR (mainline) was clearly constructed in order to colonise the Russian East, the Transsiberian Routes' locations were not based on existing populations or local natural advantages. This makes them ideal candidates for an IV approach which will be presented in section 4.4.3.

4.4.3 Instrumental Variable Approach

Equations (4.4.9) through (4.4.12) are equivalent to equations (4.4.1) through (4.4.4) with regards to the included control variables. The specification is estimated using a Two Step Least Squares model where the log of the distance to the nearest Transsiberian Route segment is used as an instrument for the distance to closest TSR (mainline) segment.

$$ln(Lights_i) = \beta(ln(DistRail_i) = ln(DistTea_i)) + \chi X_i + \epsilon_i \tag{4.4.9}$$

$$ln(Lights_i) = \beta(ln(DistRail_i) = ln(DistPost_i)) + \chi X_i + \epsilon_i \tag{4.4.10}$$

$$ln(Lights_i) = \beta(ln(DistMain_i) = ln(DistTea_i)) + \chi X_i + \epsilon_i \tag{4.4.11}$$

$$ln(Lights_i) = \beta ln(DistMain_i) = ln(DistPost_i)) + \chi X_i + \epsilon_i \tag{4.4.12}$$

Table 4.4.8 shows the results for estimating the impact on the mean nocturnal lights emission. The first stage results listed in Panel A show a positive and significant correlation between the logs of the respective distances which while diminishing in magnitude remains positive. This relationship is significant at the 1%-level across all columns. The

Kleinbergen-Papp and the Anderson-Rubin-Wald F statistics both indicate that the proposed IV model is not suffering from weak instrument issues[18]. Comparing columns (1) through (7) in panel B, we observe while almost tripling in absolute value, the coefficient on the instrumented variable remains negative and significant.

Table 4.4.8. *IV Results (TSR Scenario)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: First Stage Results			
DistTea	0.448***	0.445***	0.160***	0.047***	0.052***	0.052***	0.031***
	(0.003)	(0.003)	(0.004)	(0.005)	(0.003)	(0.003)	(0.003)
				Panel B: Second Stage Results			
VARIABLES	Lights	Lights	Lights	Lights	Lights	Lights	Lights
DistRail	-1.067***	-1.076***	-1.512***	-2.490***	-2.076***	-2.069***	-3.027***
	(0.013)	(0.014)	(0.048)	(0.205)	(0.167)	(0.166)	(0.344)
DistFlare		0.017***	0.016***	-0.013**	-0.031***	-0.031***	-0.008
		(0.004)	(0.004)	(0.006)	(0.006)	(0.006)	(0.005)
CalPot			-0.056***	-0.072***	-0.041***	-0.040***	-0.051***
			(0.005)	(0.008)	(0.006)	(0.006)	(0.009)
PopDens				-0.499***	-0.328***	-0.325***	-0.616***
				(0.077)	(0.062)	(0.062)	(0.118)
Precip					-1.167***	-1.171***	-1.717***
					(0.057)	(0.057)	(0.146)
PrecipSD						0.008***	0.012***
						(0.002)	(0.003)
RoadDens							-0.206***
							(0.030)
Observations	134,878	134,878	133,167	132,843	132,843	132,843	132,843
Kleibergen-Papp LM Stat.	13228	12665	1783	183.2	228	227.9	84.56
Kleibergen-Papp F Stat.	20848	20013	1929	186.7	233.9	233.9	85.83
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (4.4.9). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the contemporary TSR network.

Panel B of table 4.4.9 reports the second stage results for mean nocturnal lights emission as dependent variable and using the Post Route as instrument. The coefficients obtained by IV estimation are all smaller than zero and significant at the 1%-level. Comparing columns (1) through (7), one observes that the coefficient's absolute value decreases by

[18] The critical values of the test provided by Stock and Yogo (2005) are 16.38, 8.96, 6.66, and 5.53 for a 10%, 15%, 20%, and 25% bias of the obtained estimator, respectively. Accordingly, the null hypothesis of the underlying estimators being biased due to weak instrumentation are rejected in all cases.

Table 4.4.9. *IV Results (TSR Scenario)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistPost	0.627***	0.628***	0.436***	0.309***	0.305***	0.305***	0.276***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
			Panel B: Second Stage Results				
VARIABLES	Lights	Lights	Lights	Lights	Lights	Lights	Lights
DistRail	-0.778***	-0.778***	-0.537***	-0.219***	-0.334***	-0.333***	-0.293***
	(0.009)	(0.009)	(0.016)	(0.024)	(0.025)	(0.025)	(0.028)
DistFlare		0.003	0.005	0.030***	0.013***	0.012***	0.006
		(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
CalPot			0.040***	0.015***	0.019***	0.019***	0.018***
			(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
PopDens				0.351***	0.317***	0.318***	0.323***
				(0.010)	(0.011)	(0.011)	(0.011)
Precip					-0.673***	-0.679***	-0.628***
					(0.020)	(0.020)	(0.022)
PrecipSD						0.008***	0.007***
						(0.002)	(0.002)
RoadDens							0.030***
							(0.003)
Observations	134,878	134,878	133,167	132,843	132,843	132,843	132,843
Kleibergen-Papp LM Stat.	25945	25897	9577	5814	5253	5252	4216
Kleibergen-Papp F Stat.	41664	41889	12938	7059	6314	6313	5173
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (4.4.10). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

roughly 70 percent. The second stage coefficients are much smaller than in the Tea Route case. This difference stems from a different underlying relationship between the instrument and the instrumented variable due to the fact that the Post Route extends much further east as observable in figure 4.3.5.

While reacting more strongly to the inclusion of additional controls, the results obtained by instrumental variable estimation suggest a sizeable, significantly negative impact of remoteness to the TSR. All control variables exhibit the expected signs. Further their magnitudes do not change noticeably compared to the OLS results. The picture somewhat changes when we turn to the results from estimating equations (4.4.11) and (4.4.12). Here I focus on the impact of remoteness relative to the historical TSR mainline. Comparing the first stage results from tables 4.4.10 and 4.4.11 to the ones from the respective TSR scenarios presented earlier, we observe a stronger and more robust relationship between the instrument and the instrumented variable. This is mainly due to the fact that the historical mainline is closer to the historical routes which is also mirrored in the fact that the differences in the second stage results are by far not as pronounced as in the TSR scenario. Contrasting the insights from both scenarios, we can summarise that there exists a causal negative impact of TSR (mainline) remoteness on local economic activity in Eastern Russia.

Table 4.4.10. *IV Results (Mainline Scenario)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistTea	0.456***	0.464***	0.426***	0.381***	0.396***	0.396***	0.382***
	(0.005)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
			Panel B: Second Stage Results				
VARIABLES	Lights	Lights	Lights	Lights	Lights	Lights	Lights
DistMain	-1.039***	-1.043***	-0.697***	-0.400***	-0.398***	-0.397***	-0.384***
	(0.029)	(0.029)	(0.034)	(0.036)	(0.035)	(0.035)	(0.037)
DistFlare		0.019***	0.096***	0.018***	0.019***	0.018***	0.017***
		(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
CalPot			0.097***	0.007***	0.007***	0.007***	0.005***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PopDens				0.837***	0.838***	0.838***	0.832***
				(0.015)	(0.015)	(0.015)	(0.015)
Precip					0.076	0.062	0.048
					(0.064)	(0.064)	(0.066)
PrecipSD						0.009**	0.009**
						(0.004)	(0.004)
RoadDens							0.022***
							(0.008)
Observations	45,283	45,283	44,944	44,636	44,636	44,636	44,636
Kleibergen-Papp LM Stat.	5882	6062	4954	4683	5110	5106	4794
Kleibergen-Papp F Stat.	6928	7075	5538	5106	5529	5527	4997
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (4.4.11). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

$$ln(Agglo_i) = \beta(ln(DistRail_i) = ln(DistTea_i)) + \chi X_i + \epsilon_i \tag{4.4.13}$$

$$ln(Agglo_i) = \beta(ln(DistRail_i) = ln(DistPost_i)) + \chi X_i + \epsilon_i \tag{4.4.14}$$

$$ln(Agglo_i) = \beta(ln(DistMain_i) = ln(DistTea_i)) + \chi X_i + \epsilon_i \tag{4.4.15}$$

$$ln(Agglo_i) = \beta(ln(DistMain_i) = ln(DistPost_i)) + \chi X_i + \epsilon_i \tag{4.4.16}$$

Turning to the IV results when using spatial agglomeration as dependent variable as I

Table 4.4.11. *IV Results (Mainline Scenario)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistPost	0.507***	0.503***	0.473***	0.441***	0.424***	0.424***	0.420***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
			Panel B: Second Stage Results				
VARIABLES	Lights	Lights	Lights	Lights	Lights	Lights	Lights
DistMain	-0.599***	-0.591***	-0.410***	-0.232***	-0.227***	-0.227***	-0.223***
	(0.028)	(0.029)	(0.031)	(0.030)	(0.032)	(0.032)	(0.032)
DistFlare		0.054***	0.127***	0.029***	0.028***	0.027***	0.024***
		(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)
CalPot			0.112***	0.009***	0.010***	0.009***	0.006***
			(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
PopDens				0.890***	0.892***	0.891***	0.874***
				(0.014)	(0.014)	(0.014)	(0.014)
Precip					-0.083	-0.097	-0.105
					(0.064)	(0.064)	(0.064)
PrecipSD						0.010**	0.010**
						(0.004)	(0.004)
RoadDens							0.044***
							(0.008)
Observations	45,283	45,283	44,944	44,636	44,636	44,636	44,636
Kleibergen-Papp LM Stat.	5837	5812	5598	5570	5461	5461	5537
Kleibergen-Papp F Stat.	6058	6019	5599	5410	5167	5166	5145
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (4.4.12). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

estimate equations (4.4.13) through (4.4.16), a similar picture arises. As expected, there is no significant difference between the first stage coefficients reported in panel A of table 4.4.13 and the ones in table 4.4.9. Comparing the coefficients of TSR remoteness from columns (1) through (7), we observe that they are all negative and significant at the 1%-level. Other than the results from estimating the model using the level nocturnal lights emission as dependent variable, the impact of TSR remoteness on lights agglomeration loses only about 30% of its magnitude. The differences between the two instruments in the first stage relationships in the TSR and the mainline scenario differ by similar magnitude as compared to the light emission case. Same holds for the mainline scenario. An interesting insight emerges once comparing the second stage results of the agglomeration specification in both scenarios. Looking at columns 7 in tables 4.4.14 and 4.4.15, we ob-

Table 4.4.12. *IV Results (TSR Scenario)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
				Panel A: First Stage Results			
DistTea	0.445***	0.441***	0.162***	0.046***	0.050***	0.050***	0.033***
	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
				Panel B: Second Stage Results			
VARIABLES	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo
DistRail	-0.949***	-0.996***	-1.464***	-1.824***	-1.633***	-1.630***	-2.187***
	(0.014)	(0.015)	(0.054)	(0.209)	(0.183)	(0.183)	(0.314)
DistFlare		0.069***	0.067***	0.061***	0.057***	0.056***	0.068***
		(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)
CalPot			-0.061***	-0.066***	-0.050***	-0.050***	-0.053***
			(0.005)	(0.008)	(0.006)	(0.006)	(0.007)
PopDens				-0.186**	-0.104	-0.103	-0.274**
				(0.079)	(0.068)	(0.068)	(0.110)
Precip					-0.649***	-0.651***	-0.994***
					(0.063)	(0.063)	(0.134)
PrecipSD						0.003	0.005*
						(0.002)	(0.003)
RoadDens							-0.129***
							(0.025)
Observations	139,325	139,325	137,606	137,606	137,606	137,606	137,606
Kleibergen-Papp LM Stat.	13223	12673	1843	181.9	217.9	217.7	96.38
Kleibergen-Papp F Stat.	20653	20034	1991	185.3	223.3	223.1	97.89
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The table reports IV estimates of equation (4.4.13). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR.

serve that the coefficient on remoteness shows no statistical significance. While it appears that there is a pronounced causal impact of remoteness on economic activity when focusing on the long run relationship mirrored by the mainline case, agglomeration appears not to be causally impacted by mainline remoteness. Regarding the coefficients obtained for the control variables the picture is highly similar compared to IV results using mean lights emission as dependent variable.

Taking all the above presented results from instrumental variable estimation into account, we can clearly state that there is sizeable positive and highly significant causal effect of being closer to the contemporary TSR network on both the level as well as the concentration of economic activity. While the impact of mainline remoteness on the level of economic activity remains negative and highly significant when focussing on the mainline

Table 4.4.13. *IV Results (TSR Scenario)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistPost	0.622***	0.627***	0.439***	0.311***	0.307***	0.307***	0.280***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
			Panel B: Second Stage Results				
VARIABLES	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo
DistRail	-0.816***	-0.800***	-0.723***	-0.407***	-0.460***	-0.459***	-0.446***
	(0.010)	(0.010)	(0.018)	(0.025)	(0.027)	(0.027)	(0.030)
DistFlare		0.064***	0.062***	0.077***	0.074***	0.073***	0.072***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CalPot			0.010***	-0.015***	-0.012***	-0.012***	-0.013***
			(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
PopDens				0.347***	0.333***	0.333***	0.335***
				(0.011)	(0.011)	(0.011)	(0.012)
Precip					-0.322***	-0.325***	-0.307***
					(0.023)	(0.023)	(0.026)
PrecipSD						0.004*	0.003*
						(0.002)	(0.002)
RoadDens							0.010***
							(0.003)
Observations	139,325	139,325	137,606	137,606	137,606	137,606	137,606
Kleibergen-Papp LM Stat.	25612	25814	9788	5922	5365	5363	4408
Kleibergen-Papp F Stat.	41067	41868	13259	7225	6495	6492	5425
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (4.4.14). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR.

scenario, this changes when estimate the impact on spatial agglomeration within the respective grid cells. My results suggest that TSR mainline remoteness has no significant impact on agglomeration. A potential explanation for this is that it is possible that in the long run centrifugal forces such as congestion economies break the relationship. These results are in line with the theoretical predictions presented in section 4.1. Further, they support previous empirical research on the topic.

4.4.4 Validity of the Suggested IV Approach

In order to justify the utilisation of the distance to the historical Transsiberian Route as a valid instrument, the instrument needs to satisfy two conditions according to Cameron

Table 4.4.14. *IV Results (Mainline Scenario)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistTea	0.458***	0.457***	0.420***	0.380***	0.395***	0.395***	0.381***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)
			Panel B: Second Stage Results				
VARIABLES	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo
DistMain	-0.468***	-0.463***	-0.311***	0.017	0.022	0.022	-0.014
	(0.030)	(0.031)	(0.036)	(0.036)	(0.035)	(0.035)	(0.037)
DistFlare		0.072***	0.089***	0.048***	0.050***	0.050***	0.051***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
CalPot			0.049***	-0.056***	-0.056***	-0.056***	-0.052***
			(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
PopDens				0.973***	0.975***	0.975***	0.990***
				(0.017)	(0.017)	(0.017)	(0.017)
Precip					0.150**	0.159**	0.198***
					(0.069)	(0.070)	(0.071)
PrecipSD						-0.006	-0.006
						(0.004)	(0.004)
RoadDens							-0.062***
							(0.011)
Observations	45,806	45,806	45,465	45,465	45,465	45,465	45,465
Kleibergen-Papp LM Stat.	5878	5920	4822	4707	5136	5132	4809
Kleibergen-Papp F Stat.	6962	6951	5403	5124	5548	5546	5015
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (4.4.15). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

& Trivedi (2005): First, it must show a sufficient correlation with the variable of interest X which is to be instrumented. Second, the instrumental variable Z must be exogenous. This means it must not be affected by other variables in the system or in other words: the impact of the instrument on the dependant variable must only be exerted via the instruments impact on the variable which is to be instrumented. The first condition in our case, namely a sufficient (conditional) correlation between the proximity to the TSR (mainline) and the proximity to the Transsiberian Routes is fulfilled as shown by the first stage results as well as the Kleibergen-Paap LM statistics reported in tables 4.4.8 through 4.4.15. The null of no correlation between the endogenous regressors and the excluded instruments is rejected in all cases. Further the test statistics suggest no presence of bias due to weak instruments. This is indicated by the Kleibergen-Paap F statistics. As

Table 4.4.15. *IV Results (Mainline Scenario)*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: First Stage Results				
DistPost	0.510***	0.500***	0.471***	0.443***	0.426***	0.426***	0.421***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
			Panel B: Second Stage Results				
VARIABLES	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo	Agglo
DistMain	-0.262***	-0.216***	-0.164***	0.032	0.024	0.024	0.019
	(0.029)	(0.030)	(0.032)	(0.030)	(0.032)	(0.032)	(0.032)
DistFlare		0.085***	0.099***	0.049***	0.050***	0.050***	0.052***
		(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
CalPot			0.056***	-0.055***	-0.056***	-0.056***	-0.052***
			(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
PopDens				0.978***	0.975***	0.975***	0.998***
				(0.016)	(0.016)	(0.016)	(0.016)
Precip					0.148**	0.157**	0.167**
					(0.068)	(0.069)	(0.069)
PrecipSD						-0.006	-0.006
						(0.004)	(0.004)
RoadDens							-0.058***
							(0.011)
Observations	45,806	45,806	45,465	45,465	45,465	45,465	45,465
Kleibergen-Papp LM Stat.	5931	5787	5567	5630	5510	5510	5587
Kleibergen-Papp F Stat.	6117	5946	5529	5448	5196	5195	5174
Hansen J Stat.	0	0	0	0	0	0	0

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: The table reports IV estimates of equation (4.4.16). All variables are in logs with 0.001 added. The specification includes all grid cell within a 500 km buffer around the TSR mainline.

with the Kleibergen-Paap LM statistics they all are decisively larger than the critical values suggested by Stock & Yogo (2005)[19]. The final potential flaw which is testable, is the failure to fulfill the overidentifying restrictions. As every endogenous regressor is instrumented by exactly one instrument, the equation is exactly identified. Therefore the overidentification restriction is fulfilled. This is also mirrored by the Hansen J statistics reported in tables 4.4.8 through 4.4.15.

The exogeneity condition can not be tested directly. Still, there are key aspects about the underlying setup in this contribution which strongly support the notion of exogeneity.

[19] The critical values of the test provided by Stock and Yogo (2005) are 16.38, 8.96, 6.66, and 5.53 for a 10%, 15%, 20%, and 25% bias of the obtained estimator, respectively. Accordingly, the null hypothesis of the underlying estimators being biased due to weak instrumentation are rejected in all cases.

Table 4.4.16. *Summary Statistics (Mean) for Treated and Control Cells (Tea Road)*

Group of Cells	0-10 km	10-20 km	10-40 km	t-test (means) 0-10km vs. 10-40km
	(1)	(2)	(3)	(4)
CalPot	3663.119	3637.006	3610.04	0.120
	(30.075)	(29.188)	(16.947)	
DistFla	411.470	411.144	409.958	0.850
	(6.902)	(6.928)	(4.017)	
Precip	454.985	456.071	455.353	0.862
	(1.741)	(1.788)	(1.080)	
PrecipSD	0.554	0.639	0.605	0.349
	(0.044)	(0.049)	(0.028)	
Number of Cells	1,001	1,001	2991	

Note: The table reports the means and standard deviations (in parentheses) for the respective cell groups. Column 4 reports Pr(|T| > |t|) for $H_0! = 0$.

The first aspect is the considerable time span between the implementation of the Transsiberian Route and the construction of the TSR mainline which amounts to about 180 years as well as the historical contexts of the two events. When the Route was implemented, the Russians subsequently wiped out large parts of the small indigenous population which led to a situation in which Eastern Russia was basically unpopulated until the settlement promoted by the TSR took up. Accordingly, it is highly unlikely that the location of the Route was determined by existing local populations. A second potential concern regarding the exogeneity of the Route's location is the existence of local natural advantages which determined the Route's course. In other words: it could be that the route developed from connecting advantageous places. While theoretically possible, this argument is undermined by two factors. First, as both the Tea Road as well as the Transsiberian Route were mere transport routes for goods and information it appears not logical that local advantages might have played a roll in determining it's location. This is further supported by the results presented in tables 4.4.16 and 4.4.17. If the assumption that the course of the routes was not determined by differences in observables, we should not find any once we move further away from the Tea Road or the Post Route, repsectively. This is illustrated by the absence of significant differences in the means of observable controls included in the IV specification as reported in tables 4.4.16 and 4.4.17.

Column 4 reports the results from performing t tests on the equality of the means of

Table 4.4.17. *Summary Statistics (Mean) for Treated and Control Cells (Transsiberian Route)*

Group of Cells	0-10 km	10-20 km	10-40 km	t-test (means) 0-10km vs. 10-40km
	(1)	(2)	(3)	(4)
CalPot	3764.791	3724.435	3666.457	0.092
	(52.873)	(50.541)	(28.730)	
DistFla	580.640	576.656	579.912	0.931
	(7.230)	(7.235)	(4.224)	
Precip	470.179	468.786	468.866	0.649
	(2.510)	(2.504)	(1.441)	
PrecipSD	0.438	0.475	0.464	0.419
	(0.025)	(0.026)	(0.016)	
Number of Cells	1,596	1,591	4,849	

Note: The table reports the means and standard deviations (in parentheses) for the respective cell groups. Column 4 reports Pr(|T| > |t|) for $H_0! = 0$.

the respective control variables. I compare the means of the cells which are closer than 10 kilometres to the Tea Road (treated cells) to the ones which are between 10 and 40 kilometres away. As we can see, the null of no differences in the means is rejected in all cases. This can be seen as strong support for the assumption that local advantages did not play decisive role in the Tea Road's location. Table 4.4.17 reports the means and test statistics for the Transsiberian Route. While differences in the means are slightly different in magnitude as compared to the Tea Road scenario, they still do not exhibit statistically significant differences. Accordingly, this supports the notion of appropriateness of both of the suggested instrumental variable approaches.

Summarising the above explained aspects, all testable conditions of sufficient correlation between the proximity to the TSR (mainline) and the respective instrument variables is fulfilled as shown by the first stage results and the Kleibergen-Paap LM statistics. Potential issues stemming from the failure to fulfill the weak identification as well as the overidentification restrictions are absent as suggested by the Kleibergen-Paap F statistic and the Hansen J statistic. There is ample historical as well as descriptive support for the second, non-testable condition of exogeneity to be fulfilled. Taking all those insights together, there is strong support for the validity of the underlying IV approach.

4.5 Conclusion

This paper has aimed to contribute to the ongoing scientific debate about the impact of infrastructure projects on local economic development and the local spatial organisation of economic activity. Further, it aimed to empirically disentangle those effects from the ones of existing local natural advantages as well as localised returns to scale by including most controls for both aspects. In doing so, I provide novel cross-section of gridded nightlight emission data, local agro-climatic, resource exploitation in combination with historical data on Tsarist trade and post routes. Due to the (historical) context in which the Transsiberian Railroad was built, I am able to empirically demonstrate a causal negative effect of remoteness to transport infrastructure on local economic activity in Eastern Russia. This effect - while varying in magnitude - is persistent to focussing on either the contemporary TSR network or the historical mainline. In addition to that, I show that this negative causal effect also impacts the local spatial organisation of economic activity when focussing on the contemporary TSR network. This effect vanishes when centering the analysis around the historical mainline which could potentially be explained by centrifugal forces which emerge over time. As a more thorough examination of this question would go beyond the scope of this paper, I relegate it to future research.

Appendix

4.A Data Appendix Nightlights

One of the main challenges in the underlying project is the lack of reliable sub-national GDP figures for Russia. In their seminal contribution, Henderson, Storeygard & Weil (2012) suggest to use the amount of light that can be observed from outer space as proxy for economic activity. While they show that nightlight emission are a viable proxy for economic activity at the national level, the authors further stress that nightlights data is of even greater value in a sub-national setting since it is available at a great geographic fineness of about a 1 square-kilometre resolution. Using this data together with geo-spatial data on for example administrative divisions this data can then be aggregated and be used to construct city or regional-level indices on economic activity. In the following, I will summarise Henderson et al. (2012)'s technical remarks on the data.

The nightlights data is recorded and provided to the public by the United States Air Force Defense Meteorological Satellite Program (DMSP). The program's satellites complete 14 orbits per day since the 1970s. The data is being digitally processed, archived and published since 1992. Originally, the data was being recorded in order to detect clouds and was meant to be used to improve operational accuracy of air strikes and the like. In the process also nocturnal light emissions of human settlements are being recorded as well. The operational pattern of the program ensures that every satellite records every given point on the earth's surface at somewhen between 20:30 and 22:00 o'clock local time. After being transmitted to the program's headquarters, the data is processed by members of the National Oceanic and Atmospheric Administration's (NOAA) National Geophysical Data Center (NGDC). This process entails the removal of unwanted lights emissions like for example forest fires, solar flares or extreme cloud cover with the aim to filter out all natural light emissions or obstructions of of the same. In the end, the cleaned data is aggregated to one composite raster file in order to produce a satellite-year data set which is then made publicly available.

As aforementioned every final product is a raster file which consists of a grid with a

30 arc-second cell size. Such a grid cell approximately covers an area of 0.86 square-kilometres at the equator. The grid extends between 65 degrees south and 75 degrees north latitude. Every grid cell / pixel reports the intensity of nocturnal light on an integer scale from 0 (no lights recorded) to 63 (sensor satiation). While the exclusion of the extreme north and south latitudes clearly means leaving a sizeable portion of the earth's surface uncovered, it is stressed by Henderson et al. (2012) that this area is only inhabited by roughly 10.000 people which is equivalent to 0.0002 percent of the world's population. Accordingly, for the underlying project this means that parts of Russia are left out of the analysis. Still, it is highly unlikely that this significantly distorts the findings. The author's further stress that the recorded night lights reflect all indoor and outdoor use of man-made light. Accordingly, both the use of light in production as well as consumption is recorded and cannot be abstracted. Still, there is a stable relationship between night lights and economic activity which is more than what has been available at this high spatial granularity for Russia. This raw data is then used in order to aggregate the light emission data to less granular grid as described in section 4.3.

4.B Appendix: The Russian Routes Across Siberia

The instrumental variable approaches suggested in this contribution is based on two historical sources. The source used in order to retrace the course and the historical facts regarding the Russian Tea Route is Avery (2003). In her contribution, the author provides a detailed picture about both the known stations of the Tea Route as well its political foundations. Figure 4.B.1 presents the original map which I used in order create a simplified and digitised map. While still being in use, the Tea Route has been extended both its sphere of influence as well as its function after the first operations have been taken up around 1730. As the Russian Tsardom aimed at extending its influence further east, more and more post stations have been installed east of Ulan-Ude. When Wenyon travelled what he then called the Russian Post Route in around 1894, it extended to the most eastern point of the Russian empire, Vladivostok. Figure 4.B.2 presents the original scan of the hand-drawn map published in Wenyon (1896). I used this original map in order to create a simplified and digitised version. This version has been used in order to compute the grid centroid distances to the Post Route which I then used in order to instrument the actual distances to the respective rail networks.

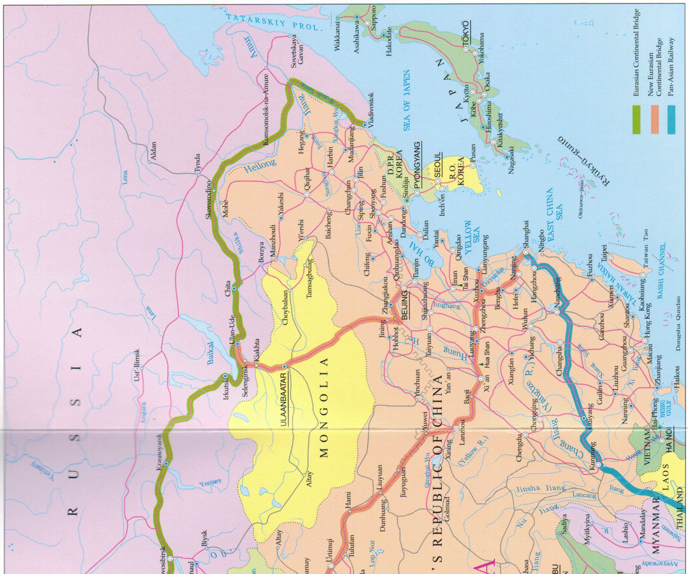


Figure 4.B.1. *Tea Road (Avery (2003))*

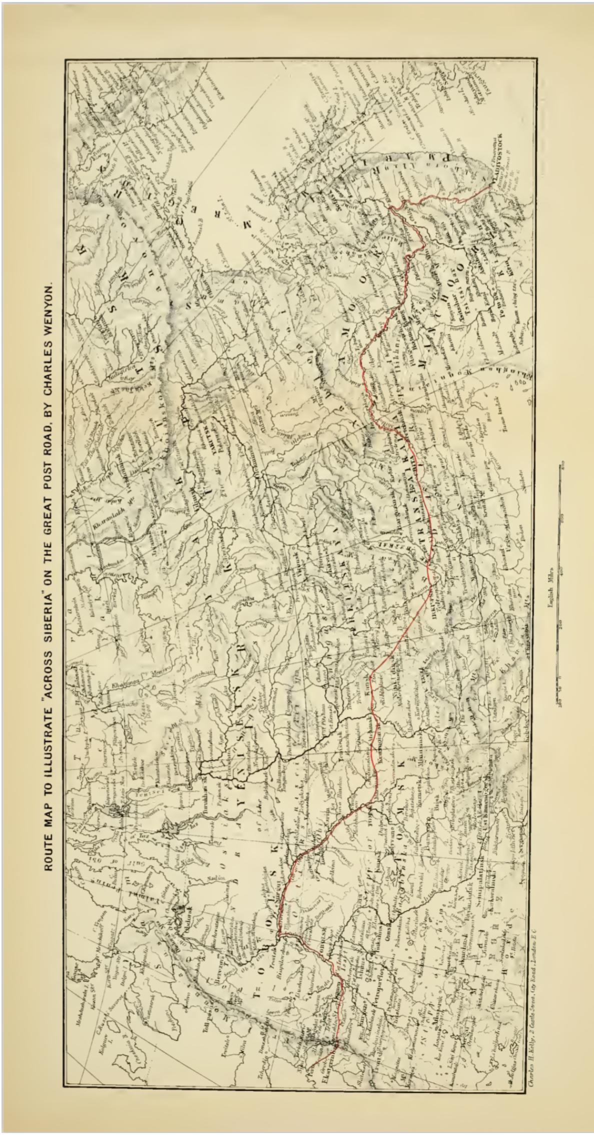


Figure 4.B.2. *Post Road* (Wenyon (1986))

Bibliography

Ahlfeldt, G. M., Redding, S. J., Sturm, D. M., & Wolf, N. (2015). The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, *83*(6), 2127–2189.

Alesina, A., Giuliano, P., & Nunn, N. (2013). On the Origins of Gender Roles: Women and the Plough. *The Quarterly Journal of Economics*, *128*(2), 469–530.

Aschauer, D. A. (1989). Is Public Expenditure Productive? *Journal of Monetary Economics*, *23*(2), 177–200.

Avery, M. (2003). *The Tea Road : China and Russia meet across the Steppe* (1st ed.). Beijing: China Intercontinental Press.

Bleakley, H. & Lin, J. (2012). Portage and Path Dependence. *The Quarterly Journal of Economics*, *127*(2), 587–644.

Bosker, M., Brakman, S., Garretsen, H., & Schramm, M. (2007). Looking for multiple equilibria when geography matters: German city growth and the WWII shock. *Journal of Urban Economics*, *61*(1), 152–169.

Bosker, M., Brakman, S., Garretsen, H., & Schramm, M. (2008). A century of shocks: The evolution of the German city size distribution 1925â“1999. *Regional Science and Urban Economics*, *38*(4), 330–347.

Cameron, A. C. & Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*.

Davis, D. R. & Weinstein, D. E. (2002). Bones, Bombs, and Break Points: The Geography of Economic Activity. *American Economic Review*, *92*(5), 1269–1289.

Davis, D. R. & Weinstein, D. E. (2008). A Search for Multiple Equilibria in Urban Industrial Structure. *Journal of Regional Science*, *48*(1), 29–65.

Dinkelman, T. (2011). The Effects of Rural Electrification on Employment: New Evidence from South Africa. American Economic Association.

Donaldson, D. (2018). Railroads of the Raj: Estimating the Impact of Transportation Infrastructure. *American Economic Review*, *108*(4-5), 899–934.

Duflo, E. & Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, *122*(2), 601–646.

Duranton, G., Morrow, P. M., & Turner, M. A. (2014). Roads and Trade: Evidence from the US. *The Review of Economic Studies*, *81*(2), 681–724.

Fogel, R. W. (1964). *Railroads and American Economic Growth: Essays in Econometric History.* Baltimore: Johns Hopkins Press.

Galor, O. (2011). *Unified Growth Theory*. Princeton University Press.

Galor, O. & Özak, Ö. (2016). The Agricultural Origins of Time Preference. *American Economic Review*, *106*(10), 3064–3103.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring Economic Growth from Outer Space. *American Economic Review*, *102*(2), 994–1028.

Hurd, J. (1983). Railways. In D. Kumar (Ed.), *Cambridge economic histroy of india* (2nd ed., Chap. 8). Cambridge: Cambridge University Press.

Jedwab, R. & Moradi, A. (2016). The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa. *Review of Economics and Statistics*, *98*(2), 268–284. arXiv: arXiv:1011.1669v3

Jensen, R. (2007). The Digital Provide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector. *The Quarterly Journal of Economics*, *122*(3), 879–924.

Lee, C. (2014). From Kiachta to Vladivostok: Russian Merchants and the Tea Trade. Slavica Publishers.

Michaels, G. (2008). The Effect of Trade on the Demand for Skill: Evidence from the Interstate Highway System. *Review of Economics and Statistics*, *90*(4), 683–701.

Michalopoulos, S. & Papaioannou, E. (2013a). National Institutions and Subnational Development in Africa. *The Quarterly Journal of Economics*, *129*(1), 151–213.

Michalopoulos, S. & Papaioannou, E. (2013b). Pre-Colonial Ethnic Institutions and Contemporary African Development. *Econometrica*, *81*(1), 113–152.

Miguel, E. & Roland, G. (2011). The long-run impact of bombing Vietnam. *Journal of Development Economics*, *96*(1), 1–15.

Rauch, F. & Michaels, G. (2013). Resetting the Urban Network: 117-2012. *Economics Series Working Papers*.

Redding, S. J., Sturm, D. M., & Wolf, N. (2011). History and Industry Location: Evidence from German Airports. *Review of Economics and Statistics*, *93*(3), 814–831.

Sala-i-Martin, X. & Pinkovskiy, M. (2010). African Poverty is Falling...Much Faster than You Think! *NBER Working Papers*, (15775).

Seiffert, S. (2019). *Go East: On the Impact of the Transsiberian Railway on Economic Development in Eastern Russia*, University of Hohenheim, Institute of Economics. Stuttgart.

Stock, J. & Yogo, M. (2005). Testing for Weak Instruments in Linear IV Regression. In D. W. K. Andrews (Ed.), *Identification and inference for econometric models* (pp. 80–108). New York: Cambridge University Press.

Storeygard, A. (2016). Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa. *The Review of Economic Studies*, *83*(3), 1263–1295.

Wenyon, C. (1896). *Across Siberia: On the Great Post Road*. London: Charles H. Kelly.

Chapter 5

Conclusion

This thesis aimed at offering a comprehensive analysis of a multitude of factors impacting the well-being of individuals in the spirit of Samuelson & Nordhaus (1998). The motivation was to identify core factors impacting individual well-being both directly or indirectly via various transmission channels. Chapter 2 was devoted to the assessment of the correlates and causes of US gun violence. It comprised a stylised model of economic crime in which assailants decide between using or not using a firearm while committing a crime. This model implicates that the utilisation of firearms is increasing in the availability of illegal guns and decreasing in both the level of social capital as well as in police intensity. The second part of chapter 2 focussed on empirically testing the hypotheses derived from the theoretical model. Using a detailed panel data set on the county level which was created by combining data published by the FBI, we found empirical support for our main hypotheses. In order to be able to make a statement about the magnitude and direction of the causal effect of the availability of illegal guns, we used the plausibly exogenous variation in the number of stolen guns in neighbouring states. The empirical evidence presented in chapter 2 suggests that the higher prevalence of illegal guns causally increases the rate at which offenders use firearms in order to achieve their criminal goals. From our insights, we could derive a number of policy recommendations. First, a possible tool in order to reduce gun-related crime could be to impose higher penalties for armed crime. Secondly, another area in which lawmakers could coerce criminals is by introducing more strict legislation regarding the possession of illegal firearms. Finally, gun violence could be lowered by strengthening social cohesion. Such increased cohesion could lower criminal inclinations in the respective society. This suggestion was derived from the fact that we show that associational density as measured by the prevalence of civic, social and religious societies has a robust negative effect on gun offenses. Developing new strategies in capacity-building with regards to social cohesion might prove a useful tool in order to combat gun crime. As laid out in chapter 1, gun-related violence is a factor severely impeding physical well-being. By contributing to the development of a better understanding of the correlates and causes of gun-crime in the US, we tried and help to improve the situation in this context.

Chapter 3 focussed on the assessment of the impact of a society's class structure with regards to income on educational outcomes. Namely, we presented a simple model of household consumption in which households can decide between subsistence consumption, investing in the next generation's education, and luxury goods. Assuming that the society is stratified into three income classes, we demonstrated that while the poor use most of their income in order to secure their basic consumption, the rich spend a considerable fraction of their budget on luxury goods. From that we derived that the middle income stratum devotes the largest share to education as compared to the other two groups. Based on those considerations, it is corollary that the relative size of the middle class in a society positively impacts their average educational outcomes. Chapter 3 further provided an empirical test of the hypothesis derived from the theoretical model. Using detailed survey data from the Indian Household Development Survey regarding household incomes, educational attainment as well as wide range of important controls, we empirically showed that higher shares of middle class households in Indian districts are associated with higher average educational outcomes. Further, since the underlying relationship is potentially affected by reverse causality, we proposed a novel instrumental variable approach in order to identify the causal effect. The specific situation in the predominantly Hinduistic society of the Indian subcontinent enabled us to do so. More specifically, we made use of the fact that the ancient caste system unique to Hinduistic cultures divides them into different strata. In the implementation of the IV approach presented in chapter 3, the district share of the middle class was instrumented by the share of people belonging to the middle two castes. This allows us to identify a causal positive effect of a larger share of the middle class on educational outcomes per district. Accordingly, strengthening the middle income stratum is recommendable if policymakers aim at fostering average educational outcomes. Thereby, individual well-being is not only fostered directly by better education but also indirectly since higher educational attainment has been shown to usually translate into higher incomes by a multitude of scientific contributions.

In the 4th chapter, I addressed the question if and how large-scale transport-infrastructure projects influence local economic development and the spatial organisation of economic activity. Namely, I investigated in how far the construction of the Transsiberian Railway and the associated reduction in transport costs had a positive impact on economic activity in Eastern Russia. While there is ample theoretical work which suggests a positive effect of such projects, empirical contributions remain rather scarce. I contributed to the existing literature in several ways. First, utilising satellite data on nocturnal lights emissions I was able to construct a novel data set. This data set entails detailed information on economic activity in Eastern Russia with a high spatial resolution. To my knowledge, there exists no reliable sub-national accounts data for said region. For this reason, chapter 4 is possibly the first economic research project which focusses on such a large part of Eastern Russia. Secondly, the specific history of the planning and construction of the Railroad enabled me to suggest a novel instrumental variable approach. This approach is based on the fact that the Transsiberian Railway follows the course of historical trade and postal routes. Those routes are plausibly exogenous with regards to local natural endowments as well as the local availability of production factors. By instrumenting the distance of specific locations in Eastern Russia with their distance to those historical routes, I was able to empirically isolate the causal impact of the transport cost reductions brought along by the Transsiberian Railway from the two previously mentioned confounding factors. In doing so, I was able to empirically prove a causal and negative effect of remoteness from the railway on economic activity as well as local spatial agglomeration of the same. Since lower transport costs usually imply lower barriers to accessing distant markets, being closer to the railway translates into higher economic activity and with it higher incomes which in turn translates into improved well-being of the local population. Hence based on the insights presented in chapter 4, investments in transport-infrastructure projects commend themselves if one aims at fostering regional development and economic well-being.

The chapters included in this thesis contributed new insights into three very heterogenous topics. This was achieved by using novel data and constructing new data sets often employing data wrangling techniques which extend beyond the usual computer-science skillset of economic scholars. Hoping to have made some people wiser, I also want to encourage scholars to dig more deeply into the touched areas and relegate such endeavours to future research.