Universitat Autònoma de Barcelona

Essays on Organised Crime and Internal Conflicts

Lavinia Piemontese

Universitat Autònoma de Barcelona

Doctoral Thesis

Essays on Organised Crime and Internal Conflicts

Author: Lavinia Piemontese

Advisor: Hannes Mueller

Tutor: Jordi Brandts

A dissertation submitted to the *Departament d'Economia i d'Història Econòmica* in fulfillment of the requirements for the degree of Doctor of Philosophy by the International Doctorate in Economic Analysis (IDEA).

May 23, 2019

A tutti i dottorandi. Passati, presenti e futuri.

Acknowledgements

I am thankful to the Fondazione Luigi Einaudi onlus for the scholarship given to me during the Academic Year 2014-2015. I am also indebted to the World Bank Macroeconomic and Fiscal Management Global Practice Group and Kevin Carey for the financial support I have received during the second semester of year 2015. Finally, I am grateful to the Ministerio de Ciencia y Innovacion and to Prof. Jordi Brandts for the FPI scholarship ECO2014-59302-P, linked to the project "Experimental economics, two-sided markets and subjective well-being". These financial aids allowed me to pursue the research presented in the thesis.

I am deeply indebted to my advisor Hannes Mueller, an extremely professional, bright, and fun person who never saves on the interest, the passion, the time, and the rigor he puts into anything he does. I believe that the quality of my research would have been worse without his supervision.

I would also like to thank Shanker Satyanath for hosting me at New York University, and for many useful conversations and valuable insights. I also thank Pablo Querubin.

I would like to praise the faculty members involved in IDEA, a group of fine and dedicated researchers and good teachers. I have benefited from the deep help and the scrupulous feedback they gave me during both the course work part and the research part of this experience. A special thank you also to Angela, Ángels and Mercè for their kindness and their excellent administrative work they did for me.

All of these years would have been different without all my colleagues with whom I have shared good and tough experiences. To those colleagues that now I can call friends. To the *IDEA CEOs* for their love and constant support.

I also want to thank all of my friends near and far. To my beloved *LMEC girls*, who make me feel I have a home in every place they live. To my dearest *cucciole* from Bologna, who know my essence and how to make me happy, despite the many years of distance.

Finally I thank every single member of my family, of which I admire the brightness. I dedicate this thesis also to them.

Contents

Preface

In this dissertation, I analyse the economic consequences of two categories of violence: organised crime and internal conflicts. It is well-established in the economic literature that both organised crime and civil wars imply significant economic losses. For example, Abadie & Gardeazabal (2003) investigate the economic effect of the Basque terror campaign and estimate a reduction of the GDP per capita of approximately 10 percent. Similarly, Pinotti (2015) studies the economic effect of the Italian mafia expansion into two regions not historically affected by this phenomenon, and measures a reduction of the GDP per capita of approximately 16 percent. Here, I estimate the cost of organised crime by looking at the economic consequences of resource misallocation linked to the Italian mafia presence. I study the cost of internal conflicts by taking a close look at the contemporaneous effects of civil wars and at the humanitarian crisis triggered by these violent events.

In Chapter 1, I study how the response of firms to mafia infiltrations can generate economic costs. I explain theoretically that extortion racketing imposed on firms located in Northern Italy is linked to resource misallocation, which is measured using the within-industry covariance between size and productivity. Using a model of monopolistic competition, I prove that extortion racketing generates misallocation, because it introduces distortions in the alignment of the rankings of firms' productivity and size. I then test this hypothesis using a panel dataset that provides information on the economic performance of 12 sectors, located in the 46 Italian provinces, between years 1998 and 2012. Through a triple-Difference-in-Differences strategy I provide evidence that the industries that are appealing for mafia groups, located in territories that experience mafia presence, suffer resource misallocation after mafia arrival.

In Chapter 2, I take the findings in the previous chapter to develop a method to estimate the economic cost of mafia infiltrations in the northern Italian economy. Specifically, I quantify the share of output that mafia groups extort from firms and the output loss that these transfers imply. The novelty of this methodology relies on the use of both panel data analysis and structural econometrics. The results of this analysis suggest that the transfer ranges between 1 and 8 percent of firm-level output for the taxed firms. I simulate the counterfactual northern Italian economy without infiltrations and estimate the cost due to mafia expansion. The results suggest that the northern Italian economy, between 2000 and 2012, suffered an aggregate loss of approximately 2.5 billion Euros. Only one fourth of this cost is the aggregate transfer to mafia groups. The remaining three fourths correspond to the contraction of production that impacted firms suffer because of resource misallocation.

In Chapter 3, I study the economic impact of internal conflicts. This chapter is extracted by the World Bank Policy Report *Recovery from conflict: lessons of success* (2017), joint work with Hannes Mueller and Augustin Tapsoba. In this report we study long-term impacts of violent conflict, to provide insights into the costs of conflict and policies to prevent conflict relapse. In this chapter I present two sections of the original report. In the first one, we analyse the contemporaneous effect of civil wars by looking at the economic growth of country or region that experiences violence. Using panel data regressions we show that the economic growth of the impacted territories experiences dramatic decreases. In the second section, we analyse the humanitarian crisis triggered by civil war. We look at the close relationship between violence and refugees. We show that in the average civil war year around 500,000 persons leave their country.

Chapter 1

Mafia and misallocation in Northern Italy: exploring extortion racket practice

1.1 Introduction

Extortion racket is a profitable practice run by criminal organisations in almost every country in the world. For example, already in the late nineteenth century, Sicilian miners were coerced by the emerging Sicilian mafia groups (Buonanno et al., 2015), or even large firms in Italy, Unites States and South America can be subject to extortion, by having a criminal organisation threaten credibly severe damages if the levy is not paid (Gambetta & Reuter (2000), CITpax (2012)). Extortion is a crucial activity because, by successfully forcing entrepreneurs to pay the tribute, criminal groups guarantee themselves control over the infiltrated territories and a stable flow of income (Konrad & Skaperdas, 1998). Clearly, what is a gain for the perpetrator is a cost for the victim. Moreover, extortion activity might also have repercussions at the aggregate level. For example, the aggregate output of the infiltrated markets can suffer significant losses (Ranasinghe, 2017).

This chapter presents a model of extortion to account of its possible aggregate affects. This model builds on the large literature on misallocation among heterogeneous producers (see, for example, Restuccia & Rogerson (2008), Hsieh & Klenow (2009), and Bartelsman et al. (2013)). In particular, I focus on the mechanism though which extortion leads to resource misallocation. To see why this might be the case, consider a standard model of heterogeneous producers, in which firms differ in their productivity level and choose employment to produce in a monopolistic competition environment. Extortion is a proportional tax on firm-level output, this distortion is idiosyncratic because notably some producers are not coerced, while others are. A central prediction of the model is that in an environment without extortion, the distribution of firms' productivity and the distribution of their sizes, measured respectively with revenue per worker and firm-level employment, are perfectly correlated. Instead, if criminal organisations infiltrate the market and coerces a randomly chosen group of firms, the strength of the link between the two distributions is weakened, meaning that the market is affected by resource misallocation.

I then explore the relationship between extortion and resource misallocation empirically. I focus on the Italian mafia, one of the most famous criminal organisations in the world, and how the response of firms located the northern Italian economy to *pizzo* has economic consequences that are detectable at the aggregate effect.[1]. Specifically, I employ a panel dataset that reports aggregate information on 12 sectors of the economy, located in the 46 northern Italian provinces, from year 1998 to year 2012, and I provide empirical evidence that mafia presence correlates negatively with allocative efficiency, measured with the within-sector covariance between size and productivity introduced by Olley & Pakes (1996) (OP covariance henceforth).

As argued in Bartelsman et al. (2013), the OP covariance is an instructive measure of allocative efficiency, as distributions of firms' productivity and firms' size tend to be correlated. Moreover, in markets with a low degree of distortions, more productive firms are larger than the less productive ones. However, the strength of this relationship is weakened if firms face idiosyncratic distortions that impact their scale of businesses. Thus, in a market infiltrated by the mafia the extent of resource misallocation is higher that the one of the counterfactual scenario without mafia, and this should be reflected in a lower OP covariance.

The main purpose of this chapter is to theorise and test the idea that the link between firm-level productivity and size depends also on mafia infiltrations. To do so, I employ the three-dimensional variation of the panel data at my disposal (sector-province-year) to: define a subgroup of industries as mafia-appealing; distinguish between infiltrated and not infiltrated provinces; and determine the timing of mafia arrival. I define mafia-infiltrated markets as specific sectors of the economy that are mafia-appealing, located in mafia infiltrated provinces, after mafia arrival. I then use this definition in a triple Difference in Differences strategy and show that the OP covariance characterising the sectors defined as mafia-appealing, located in mafia-infiltrated provinces, decreases significantly, only after mafia arrival.

This analysis contributes to the literature on the economic consequences of weak institutions. Taking a macro approach, Acemoglu et al. (2001) and Hall & Jones (1999) show how differences in institutions and government policies can explain why per capita

[1]The world *pizzo*, derived from the Sicilian language, indicates the extortion perpetrated by the mafia

income differs considerably among countries. An emerging strand of this literature joins the macro and the micro aspects, using firm-level data to analyse cross-country differences in income and aggregate productivity. In this perspective, Ranasinghe (2017) explores the role of property rights and their link to acts of extortion. Ranasinghe & Restuccia (2018) quantify the effects of institutional differences in the degree of financial development and the rule of law on aggregate outcomes and economic development. Finally, Besley & Mueller (2018) study the consequences of predation and estimate the welfare loss due to the misallocation of labour across firms and within firms, when labour is moved from production to protection. Compatibly with this literature, I study how extortion perpetrated by mafia groups affect the allocative efficiency of the infiltrated markets.

This study relates also to a number of theoretical and empirical contributions on misallocation. The relationship between firm-level idiosyncratic distortions and aggregate productivity has been theorised by Restuccia & Rogerson (2008) and developed by Hsieh & Klenow (2009), who provide an empirical analysis of resource misallocation to explain cross-country difference in productivity, measured with Total Factor Productivity (TFP). Bartelsman et al. (2013) take a step further and propose the within-industry covariance between firms' productivity and size as the most instructive index of resource misallocation. Along these lines, I study how acts of extortion imposed by mafia groups to heterogeneous firms generate resource misallocation. I analyse this relationship *directly*, in the sense that I analyse the role of a specific distortion that is idiosyncratic across heterogeneous establishments.[2]

Naturally, this analysis adds to the literature on the diffusion of mafia groups in new areas, topic that has been given little attention so far. Buonanno & Pazzona (2014) study possible channels that favoured the spreading of Italian mafia to the northern Italian provinces.[3] Moreover, Piemontese (2013) and Scognamiglio (2018) show that the construction sector is a crucial environment for mafia migration. In this chapter I study how the response of firms to mafia infiltrations in the northern economy can generate economic costs. I quantify these costs in the next chapter.

The remainder of the chapter proceeds as follows. Section 1.2 lays out the model of extortion and misallocation and presents a theoretical description of the OP covariance. Section 1.3 introduces the data and the definition of mafia-infiltrated markets. The empirical strategy and the results are presented in Section 1.4. Section 1.5 concludes the chapter and sets the basis of the next chapter.

[2]For a comprehensive review of the literature on misallocation and a deeper understanding of the difference between the *direct* and the *indirect* approach to the topic, see Restuccia & Rogerson (2013).

[3]In particular, the authors study the interaction between *Confino law* and the influx of southern workers into northern regions. *Confino* is a peculiar Italian policy measure that imposed the compulsory displacement of people strongly suspected of being part of mafia-like organisations.

1.2 Theoretical framework

The main point of this chapter is to argue theoretically and empirically that mafia groups affect economic activity by introducing distortions in the functioning of the impacted markets. In particular, the extortion of *pizzo* from a randomly chosen group of firms alters the allocative efficiency of the infiltrated markets.

In this section I first explain what the OP covariance is and why it can be used as an instructive index of allocative efficiency. I then describe theoretically the mechanism through which mafia infiltrations can lead to resource misallocation. I do so using a multisectorial static model of monopolistic competition. This is a partial equilibrium model, in which firms choose labor taking the wage as given and the mafia is simply an idiosyncratic distortion to firms' scale of output.

1.2.1 The OP decomposition

Before moving to the model, it is worth reviewing the OP decomposition of industry-level productivity and describing how it is measured in this analysis. The OP decomposition comes from a measure of aggregate efficiency introduced by Olley & Pakes (1996). In this seminal contribution, aggregate productivity is defined as the weighted sum of firms' productivity, where the weight is firms' size. This index can be decomposed into two components: the unweighted average of firms' productivity and the covariance component, which measures the extent to which most productive firms are larger than the less productive ones.

Consider a market populated by N firms. The OP covariance is derived by the following decomposition of the aggregate productivity:

$$\Omega \equiv \sum\nolimits_{i \in N} \theta_i \omega_i = \bar{\omega} + \sum\nolimits_{i \in N} ((\theta_i - \bar{\theta})(\omega_i - \bar{\omega})) \tag{1.1}$$

where Ω is the aggregate productivity of the market, ω_i and θ_i are firm-level productivity and size, respectively, and a "bar" over a variable represents the unweighted average of the firm-level measure. Thus, the first term of the right hand side $\bar{\omega}$ is the unweighted average productivity of the N firms operating in this market. The second term is the so called OP covariance between productivity and size of the N firms.

Proposition 1 helps to see how the OP covariance contributes to increasing aggregate productivity and why it can be used as an informative measure of resource misallocation.

The proof is provided in Appendix A.1.

Proposition 1 *Consider the vector ω containing N firm-level productivities ranked as follows $\omega_1 > \omega_2 > \omega_3 > ... > \omega_N$. Consider the vector θ containing N of firm-level size ranked as follows $\theta_I > \theta_{II} > \theta_{III} > ... > \theta_N$.*
If aggregate productivity is defined as the sum of firm-level productivity weighted by firm-level size, the way of maximising it is to have the ranking of firm-level productivity and firm-level size perfectly aligned. In other words, aggregate productivity is maximised when ω_1 is matched to θ_I, ω_2 is matched to θ_{II}, and so forth.

Proposition 1 tells that aggregate productivity is maximised when the rankings of firm-level productivity and firm-level size are perfectly aligned. Looking at the RHS of Equation 1.1, aggregate productivity is maximised when the OP covariance is maximised (because the first term of the RHS is constant). Hence, the highest value of the OP covariance is obtained when the most productive firm is also the largest one, the second productive one is the second largest one, and so forth. A market where these rankings are not aligned is a market affected by resource misallocation. I rely on Proposition 1 below, where I show how mafia extortive behaviour can lead to resource misallocations.

In the empirical analysis proposed in this chapter, I measure resource misallocation with the OP covariance using (log-) labour productivity, i.e. revenue per worker, and share of employment over total industry employment to measure productivity and size, respectively. These arguments have the advantage that their computation requires information widely available in firm-level datasets and less subject to measurement error than variables such as capital or total factor productivity.

1.2.2 The model

Model setting

This model builds on Restuccia & Rogerson (2008) and Hsieh & Klenow (2009). Firms are heterogeneous in their level of productivity and face idiosyncratic output distortions. Production units have access to a decreasing return to scale technology and operate in a monopolistic competition framework. Finally, as in Bartelsman et al. (2013), the model includes overhead labour as a part of the technology, to guarantees dispersion of labour productivity even in the absence of mafia distortions.[4]

Consider an economy defined by $p \times t$ labour markets, where each market is defined as a specific geographical area p at a given point in time t. Assume that the final aggregate

[4]Overhead labour can be seen as employment that is not used for production, e.g. personnel for protection, reception, or supportive services.

output Y_{pt} is produced by a representative firm that operates in a perfectly competitive market: the final good is sold at price P, which is taken as given. To produce Y_{pt}, this single representative firm combines s intermediate outputs Y_s produced from all S sectors of the economy in a Cobb-Douglas technology:

$$Y_{pt} = \Pi_{s=1}^{S} Y_s^{\theta_s} \tag{1.2}$$

with $\sum_{s=1}^{S} \theta_s = 1$. Given that the final good price P is taken as given, cost minimisation implies that $P_s Y_s = \theta_s PY$ for every sector s, where P_s is the sector specific price of the intermediate output Y_s.

In each sector s the production of the intermediate output Y_s is carried out by a single representative firm that combines N_s differentiated inputs using a CES production function. Every input Y_{ptsi} is supplied by firm i at price P_{ptsi} in monopolistic competition. Industry s production is given by:

$$\begin{aligned} Y_{pts} &= \left(\sum\nolimits_{i \in N_s} Y_{ptsi}^{\frac{\sigma-1}{\sigma}}\right)^{\frac{\sigma}{\sigma-1}} \\ &= \left(\sum\nolimits_{i \in N_s} Y_{ptsi}^{\gamma}\right)^{\frac{1}{\gamma}} \end{aligned} \tag{1.3}$$

where σ >1 is the elasticity of substitution between any two varieties. For computational simplicity σ is rewritten as $\frac{1}{1-\gamma}$, with γ <1.

Mafia infiltrations

From now on, consider a given market defined by industry s located in area p at time t. In this market, as mentioned above, there are N_s firms that produce N_s differentiated products in a monopolistic competition regime.[5] The production function of firm i exhibits decreasing returns to scale, labour is the unique input, and includes overhead labour as friction:

$$Y_i = \Gamma_{pt} A_i (L_i - f_s)^{\alpha} \tag{1.4}$$

with α smaller than one because of decreasing returns to scale.

I make the following assumptions: (i) firm i's productivity has a firm-specific component A_i, i.e. firm i's total factor productivity (TFP) and a second time-varying and province-specific exogenous component Γ_{pt}, which captures all the province-time specific factors that affect aggregate outcomes that are common to every industry s located in

[5] To simplify the notation, given that the focus is on a given market pts, I use the subscript i instead of $ptsi$.

that area; (ii) the firm-specific productivity component A_i is drawn from a log-Normal distribution with average μ_{ps} and standard deviation σ_{ps}, and these moments of TFP are sector and province-specific but time-invariant, suggesting that different provinces specialise in different sectors;[6] and (iii) overhead labour f_s is exogenously determined and sector-specific.

The mafia enters in the model as an exogenous disturbance on firms' level of output. Given that some firms are coerced and others are not, mafia infiltrations can be seen as idiosyncratic distortions that are orthogonal to firms' individual productivity.[7] This distortion is the result of the interaction between two terms. First, the idiosyncratic component, called "mafia exposure" parameter τ_i, which is Bernoulli distributed with average λ. This dummy variable equals one if firm i at time t is exposed to mafia infiltrations and zero otherwise. I assume that if τ_i equals one, firm i is forced to pay *pizzo*. If instead τ_i is zero, firm i will not interact with mafia at all. Second, the "mafia intensity" component, which measures the share of output that mafia groups extort, is given by δ. This parameter is assumed to be the same for all firms with τ_i equal to one; i.e. each infiltrated firm pays the same amount of *pizzo* δ.

If the mafia infiltrates the market, firm's i profit is:

$$\Pi_i = (1 - \tau_i \delta) Y_i P_i - w_{st} L_i \tag{1.5}$$

where w_{st} is the cost of labour. I assume that w_{st} is exogenous and that it changes over sector and time but not across provinces.[8]

Note that this maximisation problem is static, i.e. there is no link between current profit-maximising decisions and choices made in other time periods. *Pizzo* can be seen as a *one-off* tax, i.e. a payment that is not demanded regularly but on sporadic occasions.

The fact that firms operate in monopolistic competition implies that each firm benefits from some degree of market power (given by the parameter γ): firm i supplies its differentiated good at price P_i, which is endogenous to Y_i. As a consequence, profit maximisation yields a price P_i which is a constant markup over the cost of labour:

[6] The log-Normal distribution assumption is consistent with evidence provided by Angelini & Generale (2008) and Donati & Sarno (2015).

[7] This assumption is compatible with the analysis of mafia extortion provided by Balletta & Lavezzi (2014). In fact, the authors model mafia behaviour as a principal-agent model where the criminal organisation does not observe firm-level productivity. Later on, I present two extensions of this model: one where I assume that mafia impact is positively correlated with firm-level productivity and another, demonstrating the reverse scenario where the correlation between mafia impact and firm-specific productivity correlate negatively.

[8] This assumption is compatible with the fact that, in Italy, wages are mainly established through collective bargaining. Thus, in every industry, wages are homogeneous at least across the northern regions.

$$P_i^* = \left[\frac{1}{\gamma}\right]^{\frac{\alpha(1-\gamma)}{1-\alpha\gamma}} \left[\frac{w_{st}}{\alpha}\frac{1}{(1-\tau_i\delta)}\right]^{\frac{\alpha(1-\gamma)}{1-\alpha\gamma}} \left[\frac{1}{\Gamma_{pt}A_i}\right]^{\frac{(1-\gamma)}{1-\alpha\gamma}} \tag{1.6}$$

$$(L_i^* - f_s) = \left[\frac{\alpha\gamma}{w_{st}}(1-\tau_i\delta)\right]^{\frac{1}{1-\alpha\gamma}} [\Gamma_{pt}A_i]^{\frac{\gamma}{1-\alpha\gamma}} \tag{1.7}$$

Finally, plugging Equation 2.4 into the production function (Equation 2.1) yields the following expression of optimal output:

$$Y_i^* = \left[\frac{\alpha\gamma}{w_{st}}(1-\tau_i\delta)\right]^{\frac{\alpha}{1-\alpha\gamma}} [\Gamma_{pt}A_i]^{\frac{1}{1-\alpha\gamma}} \tag{1.8}$$

Optimal P_i, Y_i and L_i can be combined to compute firm i's equilibrium level of labour productivity as following:

$$LPR_i^* = \frac{P_i^* Y_i^*}{L_i^*} = \left[\frac{w_{st}}{\alpha\gamma}\frac{1}{(1-\tau_i\delta)}\right] - \left[\frac{w_{st}}{\alpha\gamma}\frac{f_s}{(1-\tau_i\delta)L_i^*}\right] \tag{1.9}$$

Assume for a moment that mafia groups do not infiltrate the market at hand, i.e. no firms pay *pizzo* thus $\lambda = 0$. In this case, the equilibrium level of firm i's productivity would be:

$$LPR_i^* = \left[\frac{w_{st}}{\alpha\gamma}\right]\left[1 - \frac{f_s}{L_i^*}\right] \tag{1.10}$$

First of all, notice that the inclusion of overhead labour guarantees that labour productivity varies across firms even in the absence of distortions. Moreover, Equation 1.10 shows that, in absence of the mafia, labour productivity LPR_i^* is increasing in firm size, measured as labour demand L_i^*, and that the firm with the highest value of LPR_i^* is the one with the highest value of L_i^*. This means that in the absence of mafia the distribution of firms' productivity and firms' size are perfectly correlated. Proposition 1 states that, if this is the case, the resource allocation characterising this market is optimised.

If, instead, mafia groups infiltrate this market by extorting from a randomly chosen group of firms the share δ of their output, the equilibrium level of labour productivity of firm i is given by Equation 2.7, from which we can deduce that the statement that the most productive firm is also the largest one is not necessarily true anymore, thus, the correlation between firms' productivity and firms' size is weakened. Following Proposition 1, if this is the case, the OP covariance characterising this market is lower to the one characterising the corresponding scenario without mafia. This means that mafia extortive behaviour, by mis-aligning the ranking between firm-level productivity and firm-level size, can lead to resource misallocation. In the following section I provide suggestive evidence of this fact.

1.3 Data

1.3.1 Definition of mafia-infiltrated markets

The first step of this analysis is the definition of mafia-infiltrated markets, which are specific sectors of the economy located in mafia infiltrated provinces and observed after a time period. This definition requires three distinctions, the first one is at the sector level. I recognise sectorial mafia presence using data about firms that have been seized by the judicial system because they were found to be directly managed or linked to mafia groups. The dataset, provided by the National Agency for the Administration and Management of Real Estate and Firms Confiscated from Criminal Organisations (ANBSC), reports information on the location and the sector to which the confiscated firm belongs, as well as the year of the final verdict of confiscation.[9]

Figure 1.1 ranks seven sectors of the economy according to the absolute and relative number (amount per 1000 firms) of firms seized from the mafia. Mafia industries are those with a high number of confiscations, while non-mafia industries have an absolute and relative number of confiscations that is close to zero. Following this criteria, mafia-appealing sectors are: accommodation and food service activities, construction, wholesale and retail trade, services, community social and personal services.[10]

The second and third dimensions needed to define infiltrated markets are province and

[9] I can distinguish among seven 1-digit ISIC rev. 4 sectors. The five 2-digit ISIC rev. 4 manufacturing sub-sectors are grouped in one. I use this information only to distinguish among mafia-appealing sectors and the remaining ones. I don't exploit the time variation because the year of seizure could provide only a "vintage" measure of mafia presence, as years can pass between the initial mafia infiltration and the year of the final verdict of seizure.

[10] Recall that, in my dataset, manufacturing is divided in six sub-sectors; thus the non-mafia appealing sectors are seven: the six manufacturing sub-sectors and transportation and storage sector.

time. To define mafia-infiltrated territories and periods, I follow Piemontese (2013), which explores the role of public investment in infrastructure on mafia expansion in Northern Italy, and shows that there was a significant increase in extortion cases in the provinces that received funding for the modernisation of the A4 motorway and the construction of the high-speed rail between Milan and Bologna, after their approval between 2000 and 2002. According this paper, mafia infiltrated provinces are those that received public funding for the renewal of the Turin-Trieste highway and/or the Milan-Bologna railway. These provinces are visible in Figure 1.2, which maps the northern territories and the infrastructure studied by Piemontese (2013).

For each infiltrated province, the mafia period starts with the approval of the funding, hence in year 2000 for some provinces and in year 2002 for the remaining provinces. Given the little available work on mafia diffusion in northern Italy, there is not a unique theory on when exactly mafia groups migrated towards the northern regions. Other studies assume that the mafia started moving in the late 70s, both because of legal practices forcing suspected *mafiosi* (people linked to the mafia) to relocate to towns that were historically unaffected by the mafia and due to migration flows from southern Italy (Buonanno & Pazzona (2014) and Scognamiglio (2018)).[11] The assumption of these studies of an earlier arrival does not necessarily contradict the assumption used in this context, of a later arrival. Even if *mafiosi* might have reached the north before the 2000s, it is plausible to assume a time lag between the mere arrival and the effective infiltration of the legal economy, which boomed in the early 2000s.

To sum up, the definition of a mafia-infiltrated market employed in this study relies on the simultaneous occurrence of industrial, geographical, and temporal dimensions: the sectors of the economy defined as mafia appealing (Figure 1.1), located in the mafia provinces (Figure 1.2), and observed after year 2000 or 2002.

1.3.2 Data on economic performance and mafia presence

The three-dimensional panel dataset employed in this analysis uses data from the *Survey on Small and Medium Enterprises* provided by the Italian National Statistical Institute (ISTAT). This dataset collects information on small and medium enterprises (SME) registered in the Italian Statistical Firms Register (ASIA). The data is stratified by sector, year, and geographical region and is representative of the universe of the SMEs operating in Italy. I use a subsample that includes information on firms operating in the northern regions over the time period from 1998 to 2012. The resulting dataset, containing between 20,000 and 29,000 observations per year, provides information on location, sector,

[11] Both Buonanno & Pazzona (2014) and Scognamiglio (2018) focus on northern and central regions, thus their results do not specifically refer to the regions where I base my analysis.

employment, sales and expenditure in intermediate inputs.[12] Data on sales and inputs is deflated with an industry-level deflator.[13] I aggregate this firm-level data at sector-province-year level.[14] The resulting dataset is an unbalanced panel dataset that provides information on the 46 northern Italian provinces, over 12 sectors, from 1998 to 2012.[15] The variables included in the dataset are: OP covariance (in terms of value added per worker and share of employment over total industry employment), mean and variance of labour, mean and variance of value added, mean and variance of value added per worker (all the variables are both in levels and logs).

For expositional simplicity, I introduce here the measure of mafia intensity that I use mainly in the following chapter. I use information on reports of extortion provided by the Yearly Book of Criminal Statistics published by the Italian Statistical Institute (ISTAT). Figure 1.3 shows the average extortion cases over time for two groups of observations: mafia and non-mafia provinces. Average extortion in mafia-infiltrated provinces is larger than in the other group; moreover, there is a clear change in the trend of the average extortion cases for mafia-infiltrated provinces starting approximately in the year 2004, i.e. after the approval of the public funding for infrastructure as mentioned above.

I use this information to construct the variable λ_{pt}, which measures the share of impacted firms for every province-year. In other words, λ_{pt} measures the extensive margin of mafia diffusion. The infiltrated markets are specific industries, located in specific provinces, observed after a given point in time. Therefore λ_{pt} has to be equal to zero in each market that is not defined as mafia-infiltrated. As Figure 1.3 shows, acts of extortion are also different than zero in non-infiltrated markets, i.e. in non-mafia provinces and in mafia provinces before mafia arrival. In order to handle this fact and have λ_{pt} equal zero in non-mafia markets, I adopt a "Difference-in-differences" approach and use the increase of reports of extortion registered in the infiltrated provinces after mafia arrival to measure mafia intensity. In particular, I adjust the number of extortion cases observed in every mafia-infiltrated province-year as follows:

$$\tilde{e}_{pt} = \text{mafia province year}_{pt} \times \left[(e_{pt} - \bar{e}_{-pt}) - \left(\bar{e}_{p,PRE} - \bar{e}_{-p,PRE} \right) \right]$$

where $\tilde{e}_{pt}$ is the adjusted number of extortion cases, mafia province year$_{pt}$ is a dummy

[12]Sectors are defined according to ISIC activity Rev.4.

[13]Industry-level deflators are gathered from the EU KLEMS Growth and Productivity Accounts: 2012 Release.

[14]Firm-level data has been aggregated to maintain confidentiality.

[15]Five of these are 2-digits ISIC rev. 4 manufacturing sub-sectors. The remaining ones are 1-digit ISIC rev. 4 sectors.

that equals one if the province-year is infiltrated, e_{pt} is the number of extortion cases observed in the raw data, $\bar{e}_{p,PRE}$ and $\bar{e}_{-p,PRE}$ are the averages of reports of extortion from 1998 until year of the approval of the infrastructure in mafia provinces and non-mafia provinces, respectively, and $\bar{e}_{-pt}$ is the yearly average for every mafia-year of the number of extortion cases observed in non-mafia provinces.[16]

Then, I compute the share of impacted firms as follows:

$$\lambda_{pt} = max\left[\frac{\tilde{e}_{pt}}{N_{pt}}, 0\right]$$

where N_{pt} is the number of firms operating in province p at time t.

In other words, I include in the dataset a variable that measures the share of firms impacted in every mafia province-year that equals the number of reports of extortion observed in the specific province-year, adjusted by the difference of the average extortion cases between mafia and non-mafia provinces from 1998 until the year of approval of the renewal works, and adjusted by the average extortion cases in the non-mafia provinces after mafia arrival. If this value is lower than zero, I replace it with zero and I exclude that specific province-year from the analysis.[17]

1.4 Empirical analysis

In order to test whether the predictions of the model hold in the data, I propose a triple Difference-in-differences strategy, which employs the three-dimensional variation of the data, i.e. sector-province-year. This method allows me to compare the average allocative efficiency of the infiltrated provinces to the one characterising the remaining northern territories, in sectors that are mafia appealing and in the other sectors of the economy, over time. To do so, I estimate the following regression:

$$\text{OPcovariance}_{spt} = \alpha + \sum_t \beta_t(\text{mafia}_{ps} \times \text{year}_t) + \theta_{st} + \theta_p + u_{spt} \tag{1.11}$$

[16]Notice that some mafia-infiltrated provinces start being infiltrated in 2000. For these provinces, $\bar{e}_{p,PRE}$ and $\bar{e}_{-p,PRE}$ are computed between the years 1998 and 2000. For the remaining mafia-infiltrated provinces, $\bar{e}_{p,PRE}$ and $\bar{e}_{-p,PRE}$ are computed between 1998 and 2002.

[17]Given that I don't have information on the sectorial intensity of extortion, I assume that mafia groups impact every sector evenly. Thus, in order to obtain the sectorial share of firms that is impacted (λ_{spt}), I divide λ_{pt} by five, i.e. the number of sectors defined as mafia appealing.

where the left hand side is the OP covariance in terms of (log) labour productivity and share of employment over total industry employment; mafia_{ps} denotes a dummy that varies at sector-province level and equals one if sector s is mafia appealing and it is located in an infiltrated province p; this dummy is interacted with each available time dummy year_t; θ_p and θ_{st} are province and time fixed effects respectively.

Figure 1.4 plots the estimates of the β_t's included in Equation 1.11, i.e. β_{1998}, β_{1999}, ..., β_{2012}, together with their 95% confidence bands. These coefficients follow a decreasing and statistically significant (or marginally significant) trend, starting after year 2000.[18]

These results point to the existence of a negative correlation between mafia presence and allocative efficiency.Note that the change in the slope of this trend happens at the same time period when we observe the sharp increase in extortion cases in the mafia-infiltrated provinces presented in Figure 1.3. This is an interesting result, since the decrease in the OP covariance is computed controlling for any province-specific and sector-year-specific determinants of allocative efficiency and without using data on extortions.

In Appendix A.2 I propose a robustness check that allows not only for mafia presence, but also for mafia intensity, which is measured with the data on extortion presented above. The results I obtain corroborate the idea that mafia presence correlates negatively with the allocative efficiency of the affected markets.

1.5 Concluding remarks

Despite the considerable attention and the many efforts that have been made in studying the development and the functioning of organised crime, the nature and the magnitude of the economic cost of this phenomenon is still a topic of active research. This chapter approaches this topic by studying the economic consequences of mafia diffusion in northern Italy. In particular, I focus on the relationship between extortion, a typical mafia activity, and allocative efficiency.

I theorise the mechanism through which mafia extortive behaviour can impact negatively the allocative efficiency characterising the infiltrated markets. To do so, I use a multisectorial model of monopolistic competition and I prove that acts of extortions generate misallocation, by introducing distortions in the alignment of the rankings of firm-level productivity and firm-level size.

I then test the prediction of the model in the data. I use panel dataset that varies at sector-province-year level and define as mafia-infiltrated markets only some sector of

[18]The magnitude of the estimates is robust to the inclusion of province-year fixed effects. However, most of the betas lose statistical significance, probably because of the sizeable reduction in variation implied by the high number of fixed effects.

the economy, located in specific provinces, after a given point in time. I then employ this definition in a triple-Difference in Differences and I show that, after controlling for any other province-specific and sector-time-specific determinant of allocative efficiency, the OP covariance between size and productivity characterising the infiltrated sectors and territories changes after mafia arrival.

To sum up, this chapter grounds the idea that a way to study economic consequences of mafia infiltrations is analysing the allocative efficiency characterising the infiltrated markets. The next step is quantifying this cost. Based on the theoretical description and the suggestive evidence presented in this chapter, in the next chapter I propose a novel methodology to assess the amount of the extortion and then compute the total loss suffered by the northern Italian economy due to resource misallocation.

Chapter 2

Mafia and misallocation in Northern Italy: a new approach of estimating the cost of crime

2.1 Introduction

Organised crime is a frightening phenomenon present in almost every country in the world. Criminal organisations engage in a wide range of economic activities, both legal and illegal, and generate huge revenues. For example, the UN Office of Drugs and Crime (UNODC, 2012) has reported that the annual turnover of transnational organised criminal activities is comparable to 1.5% of the global GDP. Measuring the economic consequences of organised crime is thus a relevant as well as a challenging task, because of the hidden nature of this phenomenon.

This chapter develops a method to estimate the cost of organised crime. I use panel data in a structural model that builds on the large literature on misallocation among heterogeneous producers (see, for example, Restuccia & Rogerson (2008), Hsieh & Klenow (2009), and Bartelsman et al. (2013)). I focus on the Italian mafia, one of the most famous criminal organisations in the world, and I study its impact on the northern Italian economy. The foundation of this analysis is the model described in the previous chapter, where I explain how firms' response to extortion can generate resource misallocation. Extortion racketing is a central mafia activity that has been classified as one of its largest sources of profit.[1] In this chapter, I use this theoretical explanation to propose a novel

[1]The spread of the mafia in northern Italy in the last decade has received considerable attention in the Italian news. A relatively recent investigation promoted by the Italian Ministry of the Interior (Transcrime, 2013) highlights a strong presence of mafia groups in the northern regions, especially in Piedmont, Lombardy and Emilia Romagna. Transcrime (2013) has estimated average revenues from mafia illegal activity. According to this report drug trafficking is the activity that generates the highest

method to estimate the economic cost due to resource misallocation.

In addition to the output losses from extortion, there is a cost in terms of reduction in production, which can be seen as a welfare loss. The key contribution of this analysis is that the method I develop provides a way to estimate the amount of *pizzo* imposed in northern Italy, and a way to understand and measure the welfare loss that occurs by transferring resources to mafia groups that would otherwise have been used productively.[2]

Some of the crucial elements of this analysis have been already described in the previous chapter of this dissertation: (i) the OP covariance between productivity and size is the proxy of allocative efficiency; (ii) the dataset that provides aggregate information on the northern Italian economy at three dimensions, i.e. sector, province, year; (iii) the distinction between mafia and non-mafia infiltrated markets according to which only specific sectors of the economy, located in a subset of provinces, observed after a given point in time are defined as mafia-infiltrated.

The way I combine tools from panel data analysis with structural estimation techniques makes my methodology innovative. I take advantage of the three-dimensional variation of the data to estimate all the determinants of allocative efficiency different than mafia. I do so performing a system of Fixed Effects regressions on the subsample of non-mafia markets. This means that, in a way, I use non-mafia markets as the best counterfactual for the infiltrated markets if the mafia were not coercing firms. I use these parameters and actual data on the extensive margin of extortion in a structural model that estimates the intensive margin of extortion, i.e. the amount of *pizzo*. Specifically, I estimate the magnitude of *pizzo* applying the Method of Simulated Moments (MSM henceforth) on the subsample of mafia-infiltrated markets and matching the OP covariance simulated by the model to the one observed in the data. Since I control for every possible determinant of allocative efficiency, the estimates of *pizzo* represent the extra distortion to allocative efficiency that occurs only in mafia-infiltrated markets. Thus, I am able to shed light on the relationship between extortion and misallocation and I exploit the structure of my model to measure the economic cost due to this phenomenon.

Beyond the estimates of the sectorial amount of *pizzo*, which ranges between 1 and 8 percent of the output level of impacted firms, I compute the economic loss due to mafia presence. These estimates imply a total cost of almost 2.5 billion euros. Almost three-fourths of this cost can be seen as a welfare loss due to resource misallocation, as this share is the forgone output of the firms that are forced to pay *pizzo*. The remaining share of the cost is the aggregate transfer to mafia groups coerced to firms.

revenues (on average 7.7 billion euros), followed by extortion (4.7 billion euros), sexual exploitation (4.6 billion euros) and counterfeiting (4.5 billion euros).

[2]Recall that the world *pizzo*, derived from the Sicilian language, indicates the extortion perpetrated by the mafia.

There is a large related literature on the economic cost of crime, which has been thoroughly reviewed by Soares (2010). There are some contributions that focus on organised crime. For example, Besley et al. (2015) explicitly address the welfare cost of Somali piracy using data on shipping contracts in the dry bulk market. Pinotti (2015) studies the economic consequences of mafia expansion in two southern regions not historically plagued by the mafia, and compute the cost of mafia expansion in terms of GDP per capita. Likewise, my analysis provides an explicit estimate of the cost of mafia diffusion in an area that has not yet been explored, northern Italy. Naturally, this analysis adds also to the literature on the economic impact of the Italian mafia. Barone & Narciso (2015), Acconcia et al. (2014), and Galletta (2017) look at the impact of mafia presence on public transfer to firms and on public spending. Pinotti (2013) and Daniele & Geys (2015) look at the implications of mafia presence on politicians' characteristics established in the infiltrated areas.

As the previous chapter, this study contributes to the literature on the economic consequences of weak institutions. Recalling the contributions that join the macro and the micro aspects, using firm-level data to analyse cross-country differences in income and aggregate productivity, Ranasinghe (2017) explores the role of property rights and their link to acts of extortion. Ranasinghe & Restuccia (2018) quantify the effects of institutional differences in the degree of financial development and the rule of law on aggregate outcomes and economic development. Finally, Besley & Mueller (2018) study the consequences of predation and estimate the welfare loss due to the misallocation of labour across firms and within firms, when labour is moved from production to protection. Compatibly with this literature, I study how extortion perpetrated by mafia groups affect the aggregate productivity and the allocative efficiency of the infiltrated markets. The main contribution of this chapter is to widen the range of methods that can be used to explore this field, by proposing a new methodology to measure explicitly the economic cost do to weak institutions, even without having complete information on the distortion analysed (e.g. the amount of extortions used in Ranasinghe (2017)).

The remainder of the paper proceeds as follows. Section 2.2 revises the model presented in the previous chapter. Section 2.3 covers the description of the estimation method developed in this analysis. Section 2.4 illustrates the results of the model estimation, the model simulations and the counterfactual analysis which measures the cost suffered by the infiltrated northern territories. Section 2.5 considers two alternative specifications of the model. Section 2.6 concludes the chapter.

2.2 Model

In this section I revise the model developed in the previous chapter. I do so to recall how the mafia enter in the model as an idiosyncratic distortion to firms' scale of output, the identifying assumption on which I rely when I develop the estimation method, and the results of the model that are the basis of the structural estimation I develop in Section 2.3.

Consider a given market defined by industry s located in area p at time t, as in the previous chapter. In this market there are N_s firms that produce N_s differentiated products in a monopolistic competition regime.[3] The production function of firm i exhibits decreasing returns to scale, labour is the unique input, and includes overhead labour as friction:

$$Y_i = \Gamma_{pt} A_i (L_i - f_s)^{\alpha} \tag{2.1}$$

with α smaller than one because of decreasing returns to scale.

I make the following assumptions: (i) firm i's productivity has a firm-specific component A_i, i.e. firm i's total factor productivity (TFP) and a second time-varying and province-specific exogenous component Γ_{pt}, which captures all the province-time specific factors that affect aggregate outcomes that are common to every industry s located in that area; (ii) the firm-specific productivity component A_i is drawn from a log-Normal distribution with average μ_{ps} and standard deviation σ_{ps}, and these moments of TFP are sector and province-specific but time-invariant, suggesting that different provinces specialise in different sectors;[4] and (iii) overhead labour f_s is exogenously determined and sector-specific.

The mafia enters in the model as an exogenous disturbance on firms' level of output. Given that some firms are coerced and others are not, mafia infiltrations can be seen as idiosyncratic distortions that are orthogonal to firms' individual productivity.[5] This distortion is the result of the interaction between two terms. First, the idiosyncratic component, called "mafia exposure" parameter τ_i, which is Bernoulli distributed with

[3]To simplify the notation, given that the focus is on a given market pts, I use the subscript i instead of $ptsi$.

[4]The log-Normal distribution assumption is consistent with evidence provided by Angelini & Generale (2008) and Donati & Sarno (2015).

[5]This assumption is compatible with the analysis of mafia extortion provided by Balletta & Lavezzi (2014). In fact, the authors model mafia behaviour as a principal-agent model where the criminal organisation does not observe firm-level productivity. Later on, I present two extensions of this model: one where I assume that mafia impact is positively correlated with firm-level productivity and another, demonstrating the reverse scenario where the correlation between mafia impact and firm-specific productivity correlate negatively.

average λ. This dummy variable equals one if firm i at time t is exposed to mafia infiltrations and zero otherwise. I assume that if τ_i equals one, firm i is forced to pay *pizzo*. If instead τ_i is zero, firm i will not interact with mafia at all. Second, the "mafia intensity" component, which measures the share of output that mafia groups extort, is given by δ. This parameter is assumed to be the same for all firms with τ_i equal to one; i.e. each infiltrated firm pays the same amount of *pizzo* δ. Note that I use actual data on reports of extortions to compute λ (the procedure I use to compute λ is presente in Section 1.3 of the previous chapter), while I estimate the "mafia intensity" component using the procedure developed in Section 2.3.

If the mafia infiltrates the market, firm's i profit is:

$$\Pi_i = (1 - \tau_i \delta) Y_i P_i - w_{st} L_i \tag{2.2}$$

where w_{st} is the cost of labour. I assume that w_{st} is exogenous and that it changes over sector and time but not across provinces.[6]

Note that this maximisation problem is static, i.e. there is no link between current profit-maximising decisions and choices made in other time periods. *Pizzo* can be seen as a *one-off* tax, i.e. a payment that is not demanded regularly but on sporadic occasions.

The fact that firms operate in monopolistic competition implies that each firm benefits from some degree of market power (given by the parameter γ): firm i supplies its differentiated good at price P_i, which is endogenous to Y_i. As a consequence, profit maximisation yields a price P_i which is a constant markup over the cost of labour:

$$P_i^* = \left[\frac{1}{\gamma}\right]^{\frac{\alpha(1-\gamma)}{1-\alpha\gamma}} \left[\frac{w_{st}}{\alpha}\frac{1}{(1-\tau_i\delta)}\right]^{\frac{\alpha(1-\gamma)}{1-\alpha\gamma}} \left[\frac{1}{\Gamma_{pt}A_i}\right]^{\frac{(1-\gamma)}{1-\alpha\gamma}} \tag{2.3}$$

$$(L_i^* - f_s) = \left[\frac{\alpha\gamma}{w_{st}}(1-\tau_i\delta)\right]^{\frac{1}{1-\alpha\gamma}} [\Gamma_{pt}A_i]^{\frac{\gamma}{1-\alpha\gamma}} \tag{2.4}$$

Finally, plugging Equation 2.4 into the production function (Equation 2.1) yields the following expression of optimal output:

[6]This assumption is compatible with the fact that, in Italy, wages are mainly established through collective bargaining. Thus, in every industry, wages are homogeneous at least across the northern regions.

$$Y_i^* = \left[\frac{\alpha\gamma}{w_{st}}(1-\tau_i\delta)\right]^{\frac{\alpha}{1-\alpha\gamma}} [\Gamma_{pt}A_i]^{\frac{1}{1-\alpha\gamma}} \tag{2.5}$$

Optimal P_i and Y_i can be combined to compute firm i's equilibrium level of value added as following:

$$VA_i^* = \left[\frac{\alpha\gamma}{w_{st}}\right]^{\frac{\alpha}{1-\alpha\gamma}} [A_i\Gamma_{pt}]^{\frac{\gamma}{1-\alpha\gamma}} \tag{2.6}$$

Moreover, optimal P_i, Y_i and L_i can be combined to compute firm i's equilibrium level of labour productivity as following:

$$LPR_i^* = \frac{P_i^* Y_i^*}{L_i^*} = \left[\frac{w_{st}}{\alpha\gamma}\frac{1}{(1-\tau_i\delta)}\right] - \left[\frac{w_{st}}{\alpha\gamma}\frac{f_s}{(1-\tau_i\delta)L_i^*}\right] \tag{2.7}$$

In the following section I show how I use the results of this model to estimate δ for every sector-province-year defined as mafia-infiltrated. In particular, I use Equation 2.6 to derive expressions for the average and the variance of value added in an environment without mafia, and I use Equation 2.4 and Equation 2.7 to simulate the OP covariance.

2.3 Estimation

In this section, I describe how to bring the model presented above to the data. The estimation technique is a structural analysis that consists of two main stages. In the first step, I use data on non-infiltrated markets to estimate all the parameters of the model that are unrelated to the mafia. I use these estimates in the second part of the analysis to generate firm-level data. By using these parameters, I control for the determinants of allocative efficiency different from the mafia, which are not idiosyncratic. The other crucial ingredient of the second stage of this analysis is the data on the share of firms that pay *pizzo* described in the previous chapter. I use this information to generate the idiosyncratic distortion that is present only the mafia-infiltrated market, i.e. the mafia extortion racketing that involves a subgroup of firms. I then implement the MSM to quantify the amount of *pizzo*, i.e. the share of output that coerced firms has to transfer to mafia groups.

2.3.1 First stage

In this stage, I focus on non-mafia markets. Recall from the description presented in the previous chapter (Section 1.3) that the three-dimensional variation of the data allows the recognition of a group of N_{NM} markets composed by: (i) sectors that are not appealing for mafia groups, in every northern Italian province, observed during years 1998-2012; (ii) mafia appealing sectors, located in mafia immune provinces, observed during years 1998-2012; (iii) mafia-appealing sectors, in mafia-infiltrated provinces, before mafia arrival, i.e before 2000 for some mafia infiltrated provinces and before 2002 for the remaining provinces.

I use data on these N_{NM} sector-province-year in order to estimate two sets of parameters presented in the theoretical model. The first group contains the parameters related to firm-level productivity, i.e. μ_{ps}, σ_{ps} and Γ_{pt}. The second set includes the exogenous macroeconomic variables, i.e wages w_{st} and overhead labour f_s. According to the assumptions stated above, firm i's productivity has an idiosyncratic component A_i and second element Γ_{pt}, which is province-year specific. Moreover, the mean and the standard deviation of A_i, namely μ_{ps} and σ_{ps}, of firms located in province p belonging to sector s are time invariant. Finally, overhead labour f_s is sector-specific and wages w_{st} are sector-specific and vary over time.

The three-dimensional structure of the data and the assumptions I make allow me to use non-mafia markets as the best counterfactual of infiltrated markets without the mafia. In fact, I utilise information on mafia appealing sectors located in mafia impacted provinces before mafia arrival to estimate two moments of the TFP μ_{ps} and σ_{ps}. I compute the province-time specific productivity component Γ_{pt} with data on sectors that are not mafia-appealing. Finally, sectorial wage w_{st} and overhead labour f_s are estimated with data on non-mafia provinces and mafia-infiltrated provinces before mafia arrival.

I compute these parameters through a reduced form analysis. Specifically, I estimate a system of Fixed Effects regressions whose equations are three moment conditions derived by the model presented in Section 2.2.

Specifically, I build the three moment conditions using the equilibrium level of firm i's value added (Equation 2.6). Since this stage of the analysis is performed on non-mafia markets, firm i's equilibrium value added can be written as follows:

$$VA_i^* = \left[\frac{\alpha\gamma}{w_{st}}\right]^{\frac{\alpha}{1-\alpha\gamma}} [A_i\Gamma_{pt}]^{\frac{\gamma}{1-\alpha\gamma}} \tag{2.8}$$

Consequently, firm i's log-value added is:

$$\log(VA_i^*) = \frac{\alpha}{(1-\alpha\gamma)}\left[\log(\alpha\gamma) - \log(w_{st})\right] + \frac{\gamma}{(1-\alpha\gamma)}\left[\log(A_i) - \log(\Gamma_{pt})\right] \tag{2.9}$$

An expression for value added as a function of optimal labour can be derived using Equations 2.7 and 2.4:

$$VA_i^* = \frac{1}{\alpha\gamma} w_{st}(L_i^* - f_s) \tag{2.10}$$

I use Equation 2.8, Equation 2.9, and Equation 2.10 to compute the following moment conditions for every non-mafia market: (i) average of log-value added (Equation 2.11); (ii) variance of the log-value added (Equation 2.12); (iii) average of value added as a function of optimal labour (Equation 2.13).

$$\overline{\log(VA)}_{spt} = \frac{\alpha}{(1-\alpha\gamma)}\left[\log(\alpha\gamma) - \log(w_{st})\right] + \frac{\gamma}{(1-\alpha\gamma)}\log(\Gamma_{pt}) + \frac{\gamma}{(1-\alpha\gamma)}\tilde{\mu}_{A_{ps}} + \epsilon^1_{spt} \tag{2.11}$$

where $\tilde{\mu}_{A_{ps}}$ is the mean of $\log(A_i)$ that is province-sector specific.

$$\text{VAR}(\log(VA))_{spt} = \left[\frac{\gamma}{(1-\alpha\gamma)}\right]^2 \tilde{\sigma}^2_{A_{ps}} + \epsilon^2_{spt} \tag{2.12}$$

where $\tilde{\sigma}^2_{A_{ps}}$ is the variance of $\log(A_i)$ that is province-sector specific.

$$\overline{VA}_{spt} = \frac{1}{\alpha\gamma} w_{st}(\overline{L}_{spt} - f_s) + \epsilon^3_{spt} \tag{2.13}$$

I estimate the average and the standard deviation of TFP, μ_{ps} and σ_{ps}, for each sector-province, the productivity component Γ_{pt} for each province-year, the wage w_{st} for each sector-year and the overhead labour f_s for each sector, by solving the system of $3 \times N_{NF}$

equations denoted by equations 2.11, 2.12 and 2.13. In fact, for each of the N_{NM} non-mafia sector-province-year, I observe $\overline{\log(VA)}$, VAR($\log(VA)$), $\overline{VA}$, i.e. the dependent variables of the system, and $\overline{L}$ that will be used as a regressor. Moreover, I assume specific values of α and γ.[7]

This system of equations can be solved performing Fixed Effect regressions. In fact, μ_{ps}, Γ_{pt} and w_{st} can be estimated from the sector-province, province-time and sector-time fixed effects in Equation 2.11, while Equation 2.12 serves to estimate σ_{ps}. Estimates of the sectorial time-varying wage $\widehat{w}_{st}$ are then plugged into Equation 2.13, whose estimation yields $\widehat{f}_s$.[8]

As a final remark, it is worth mentioning that the estimates obtained at this stage are robust to two alternative specifications of the system. In the first one I use as moment conditions average labour, variance of labour, and average value added as a function of labour for every non-mafia province-sector-years. In the second alternative, I estimate a system of $5 \times N_{NF}$ equations, where I add to the baseline system average and variance of labour for each non-mafia sector-province-year as extra conditions. The results I obtain in both alternatives are strongly correlated to the estimates that the baseline system yields.

2.3.2 Second stage

The second stage of the analysis focuses on the mafia-infiltrated markets, defined as mafia appealing sectors, located in mafia-infiltrated provinces, observed after the arrival of mafia. For each of these N_M markets, I estimate the amount of *pizzo* that mafia groups extort from a randomly chosen group of firm. I do so performing the MSM where the main moment condition is the OP covariance between labour productivity and labour. The set of parameters that I compute in this stage is the vector Δ, which includes the N_M values of δ_{spt} that mafia groups extort in each infiltrated market.

Given that mafia distortion cannot be disentangled using aggregate data, I cannot estimate vector Δ using standard econometrics techniques. Thus, I compute it using MSM, which minimises a distance criterion between key moments from actual data and corresponding moments computed using simulated data.

For every sector-province-year, I generate a vector of firm-specific productivity A_i from a log-Normal distribution with average $\widehat{\mu}_{ps}$ and standard deviation $\widehat{\sigma}_{ps}$.[9] Firm-specific

[7] I follow Bartelsman et al. (2013) and Bloom (2009) and assume a 20 percent markup that yields γ equal to 0.83. Moreover I introduce additional curvature to the firm-level profit function and assume α equal to 0.85 as in Atkeson & Kehoe (2005).

[8] Alternatively, this system of $3 \times N_{NF}$ can be solved simultaneously through MLE. In fact, linearity guarantees that the estimates obtained using OLS are equivalent to the ones obtained performing MLE.

[9] The length of this vector is the number of firms that operate in the specific sector-province-year

productivity is then rescaled with the province-year component $\widehat{\Gamma}_{pt}$.

Then, I let mafia groups enter in the model. Recall that mafia impact is modelled as an idiosyncratic distortion constituted by the interaction of the "mafia exposure" component τ_i with the "mafia intensity" component δ. The "mafia exposure" component is assumed to follow a Bernoulli distribution with average λ. Hence, in each mafia-infiltrated market, λ gives the share of firms that are impacted by mafia groups. The computation of λ characterising each impacted market is described in detail in Section 1.3 of the previous chapter of this disseration. Mafia infiltrations are orthogonal to firms' productivity, thus I randomly select the share λ of firms that have to pay the tax δ, which I estimate structurally through MSM. In other words, I feed the model with actual data on a component of the distortion, i.e. the average λ of τ_i, in order to measure the other component δ, which is unobserved.

I implement the MSM as follows: I create a grid Θ of values that δ can take and choose a set of observed data moments Φ^O that the model has to match.[10] For each possible vector Δ formed by each possible combination of values of δ in Θ, the model is solved and the simulated moments of interest $\Phi(\delta)^S$ are computed. The estimate of the vector containing the amount of *pizzo* paid in each mafia-infiltrated market $\widehat{\Delta}$ is derived according to the following criterion:

$$\widehat{\Delta} = \arg\min_{\Delta}[\Phi^O - \Phi(\Delta)^S]' W [\Phi^O - \Phi(\Delta)^S] \tag{2.14}$$

where W is a weighting matrix.[11]

The most important actual data moment that I use in this analysis is the OP covariance between productivity and size. In fact, this analysis relies on the assumption, tested in the previous chapter, that mafia extorting behaviour generates factor misallocation in mafia-infiltrated markets and that can be detected by looking at the OP covariance that characterise the affected sector-province-year. Any other factor that contributes to allocative efficiency has been estimated in the first stage using data on mafia-free environments. I add to the set of actual data moments the average and the variance of labour characterising each mafia-infiltrated market to estimate an overidentified model.

[10] The grid search method is useful to avoid convergence problems with possible non-convexities of the objective function (see Equation 2.14).

[11] For a comprehensive description of the functioning of MSM and of its statistical properties see McFadden (1989).

2.4 Results

This section presents and discusses the results obtained from the estimation described in Section 2.3. As illustrated above, the mafia infiltrates markets by forcing a randomly selected group of firms to pay *pizzo*, i.e. imposing a tax δ on firm i's output. I use the MSM to estimate the vector Δ containing the amount of *pizzo* imposed in each infiltrated market. The structural approach allows me to go beyond the mere estimate of Δ and to perform two relevant exercises. First, a model simulation aimed at understanding the source of the overall economic costs imposed by the mafia. Second, a counterfactual analysis that quantifies this cost.

I simulate the impact of extortion in each mafia sector on aggregate value added and employment, and average TFP, changing both the extensive and the intensive margins of this distortion. I find that mafia infiltrations can reduce the number of firms that operate in the impacted markets, because some impacted firms can incur negative profits, and thus they have to leave the market. Moreover, by extorting *pizzo*, mafia groups make the coerced firms reduce employment, and, in turn, production. This economic cost, which is increasing with the imposed mafia-tax δ, is thus composed by the transfer of money that extorted firms have to make, the forgone value added of the impacted firms that exit the market because they incur negative profits, and the forgone value added of the impacted firms that have to reduce employment. The last two components, which account for approximately three-fourths of the total cost, can be seen as the welfare loss due to mafia infiltrations. This is a key result of this analysis, because it sheds light on the magnitude economic costs that go beyond the mere transfer of resources that would have been otherwise used productively. I measure this cost with a counterfactual analysis, in which I compare the aggregate value added of the infiltrated economy to the one that would have been without the mafia, i.e. by setting the share of impacted firms λ equal to zero.

2.4.1 Model estimation

Figure 2.1 plots the histogram of the estimates of *pizzo* that I obtain for each mafia-infiltrated market, i.e. $\hat{\delta}_{spt}$.[12] The distribution is right-skewed, with most of the estimates ranging between 0.01 and 0.05. This means that in most of the infiltrated markets, coerced firms pay an extortion that ranges between 1 and 5 percent of their output.

Another way of studying the results is by looking at Table 3.1, which reports the average of $\hat{\delta}$ computed for each mafia appealing sector s, i.e. $\bar{\hat{\delta}}_s$. The least impacted sector

[12]A brief comment on the fit of the model is provided in the appendix of this chapter.

is wholesale and retail trade with average *pizzo* of 1%. Construction and services have an average *pizzo* of 3%, while community social and personal services has a slightly higher average tax of 5%. The most impacted sector is accommodation and food services, where the average *pizzo* is 8%. All in all, these estimates suggests that on average, depending on the sector where they operate, impacted firms are extorted between 1 and 8 percent of their output.[13]

Table 3.2 reports the results of a crucial validation exercise, where I estimate the amount of *pizzo* in the seven non-mafia sectors located in the mafia-infiltrated provinces after mafia arrival. In other words, assuming that mafia groups infiltrate also sectors that are not defined as mafia appealing, I estimate the amount of extortion imposed in this environment. First of all, I compute the parameters necessary to generate firm-level data, using information on the new control group. Wages w_{st} and overhead labour f_s are measured using data on non-infiltrated mafia provinces. Mean and standard deviation of TFP, $\mu_{A_{ps}}$ and $\sigma_{A_{ps}}$, are computed with data on mafia sectors-provinces before mafia arrival. In order to estimate the productivity component common to all firms that operate in the same province-time, Γ_{pt}, I need information on non-mafia sectors. Since, in this exercise, every sector is infiltrated by mafia, I compute Γ_{pt} performing seven rounds of the first stage of the estimation (one per sector), each one including non-appealing sector within the group of mafia-appealing sectors. Using these parameters, I can perform the second stage of the estimation, where I match the OP covariance simulated by the model to the corresponding one observed in the data (together with the other moment conditions), and I back out one δ for each affected market.

The results presented in Table 3.2 validate the hypothesis that extortion racketing distorts allocative efficiency only in mafia-infiltrated markets. In fact, for each mafia non-appealing sector, average estimates of δ are very close to zero. This means that the allocative efficiency characterising each market is entirely explained by factors that are province-sector, province-time, sector-time and sector specific, and not by acts of extortion imposed by the mafia.

2.4.2 Model simulation and counterfactual analysis

In order to understand where the economic loss due to mafia infiltration derives, I perform two simulation exercises of the model described in Section 2.2. In the first test, I

[13]It is worth to mentioning that these numbers are compatible with those presented by Balletta & Lavezzi (2014). The authors use a unique dataset on extortion in Sicily in which *pizzo* rate is observed. The reported magnitudes of *pizzo* (Figure 9 pg. 18) are on average even higher than the estimates obtained in the present analysis.

focus on the intensive margin of the distortion, studying how aggregate outcomes change by keeping the number of infiltrated firms fixed and changing the amount of *pizzo* they are forced to pay. The second simulation considers the reverse exercise, i.e. the intensive margin of the distortion. I look at aggregate outcomes, fixing the amount of *pizzo* and changing the share of impacted firms, and I explain the composition of cost that extortion brings to the infiltrated economy. Finally, I perform a counterfactual analysis that quantifies this cost.

The intensive margin of the mafia

In this exercise, I simulate five economies, one for each mafia sector. I simulate a market with 1000 firms, choosing parameters obtained in Stage 1 (Section 2.3) as follows: (i) for every sector, the mean and the standard deviation of TFP are the median values of the estimates of $\widehat{\mu}_{ps}$ and $\widehat{\sigma}_{ps}$ repsectively; (ii) the province-year component of firms' productivity is the median of $\widehat{\Gamma}_{pt}$ for year 2006; (iii) the sectorial wage $\hat{w}_{st}$ is the estimate obtained for year 2006; and (iv) the overhead labour $\hat{f}_s$ is the one obtained for the sector at hand.[14] I hold the share of impacted firms λ_{pt} constant at its median level observed in year 2006 and I let the amount of *pizzo* δ change between zero and one. Then, for each value of δ, I compute the total number of active firms (i.e. those that do not incur negative profits), their average TFP, and the aggregate value added and employment. Each simulation is repeated 25000 times.

Recall that firm i extracts its value of TFP A_i from a log-Normal distribution and its value of the "mafia exposure" parameter τ_i (i.e. the idiosyncratic component of the distortion) from a Bernoulli distribution. This means that firm i knows its level of productivity and whether it is forced to pay the tax δ. Given that the only input that firm i utilises is labour, whose price (wage) is taken as given, its draws of A_i and τ_i are enough to determine whether it produces or not. In other words, if firm i is extorted the share δ of its output two alternatives are possible: (i) it hires less workers and produces less; or (ii) if its draw of A_i is too low, its profit is negative and it stays inactive. This fact has three implications. First of all, aggregate employment and aggregate value added decrease with δ. In fact, recall from Equations 2.15 and 2.16 below that both optimal labour and equilibrium value added depend negatively on the magnitude of the extortions.

$$(L_i^* - f_s) = \left[\frac{\alpha\gamma}{w_{st}}(1-\tau_i\delta)\right]^{\frac{1}{1-\alpha\gamma}} [\Gamma_{pt}A_i]^{\frac{\gamma}{1-\alpha\gamma}} \tag{2.15}$$

[14] Comparable results are obtained with different parameters combinations and baseline years.

$$VA_i^* = \left[\frac{\alpha\gamma}{w_{st}}(1-\tau_i\delta)\right]^{\frac{\alpha\gamma}{1-\alpha\gamma}} [\Gamma_{pt}A_i]^{\frac{\gamma}{1-\alpha\gamma}} \tag{2.16}$$

Second, given that the impacted firms that incur negative profits exit the market, there will be a lower number of operating firms than in the corresponding scenario without mafia infiltrations. Third, it is possible that the average TFP of the operating firms increases with *pizzo*, because the productive units that exit are those with lower draws of TFP.

Figure 2.2 presents the outcome of the construction industry. These results confirm the premise discussed above. The first panel of Figure 2.2 shows that as δ increases, the number of firms that produce in the market decreases. The second and the third panels show that both total employment and aggregate value added decrease as δ increases. Interestingly, we can notice that average TFP is inverse-U shaped, i.e. it increases up to a given threshold of δ and then it starts decreasing. This is explained by the fact that when the rate at which firms are taxed is too high, both firms with high and low draws of TFP are unable to produce, and mafia simply removes a portion of firms from the market.[15]

The extensive margin of the mafia

The simulation presented above highlights that the welfare loss brought by the mafia depends on the forgone production of the infiltrated firms that either reduce employment, or do not produce at all. To have a taste of the magnitude of these components, I perform a second simulation. I generate firm level data with the same parameters of the previous application, but here, for each sector, I fix the mafia tax δ at its average level (values presented in Table 3.1) and I make the share of impacted firms λ vary between zero and one. I focus on infiltrated firms and, for each value of λ, I compute: (i) the total amount of *pizzo* paid by the firms that do not exit the market —this is not a component of the welfare loss but simply a waste of resources; (ii) the total forgone value added of the firms that reduced employment; and (iii) the forgone value added of the firms that leave the market. I also compute the average TFP of the firms that produce in each scenario, i.e. for every value of λ.

Figure 2.3 plots the results for the construction industry.[16] The cost, which is the sum of (i), (ii) and (iii), is clearly increasing with the share λ of impacted firms.

Quite remarkably, the welfare loss, i.e. the forgone value added due to either employment reduction or exit, accounts for almost three fourths of the total cost. This is a key

[15]The same exercise is repeated for the remaining mafia appealing sectors. Results, available upon request, are analogous to the ones obtained for the construction sector.

[16]Results obtained from the simulation of the other mafia-appealing sectors are comparable to the ones presented. Figures are available upon request.

result of the paper; in fact, the model developed here provides a way to estimate the economic cost of the misallocation generated by mafia extortion. In fact, by extorting *pizzo* from randomly selected firms, the mafia makes some firms hire less workers than what they would have hired had they not been coerced. Second, the mafia makes other firms exit the market because the transfer imposed make them unable to hire labor. As a consequence of mafia extortion some firms produce below their optimal level and other firms do not produce at all. The cost due to misallocation is almost three times higher than the mere transfer to mafia groups.

Finally, the second panel of Figure 2.3 shows that TFP is steadily increasing. This result suggests that imposing a relatively small *pizzo*, makes only the least productive firms exit the market, implying an increase of the average TFP of the producing firms.

Counterfactual analysis

The previous subsection sheds light on the sources of the economic cost due to extortion. To quantify this cost, I implement a counterfactual analysis where each infiltrated market is compared to its corresponding scenario without mafia infiltrations, i.e. where λ is set equal to zero. For each market, I compute the three components of the cost explained above, i.e. the total transfer to mafia and the forgone value added due to both reduction in employment and firms' exit. Results are aggregated by sector in Table 3.3. The most heavily impacted sector is Accommodation and food services, with a total cost of more than 800 million euros, while the least impacted is Services, with a total cost of approximately 340 million euros. The total cost suffered by northern Italy between years 2000 and 2012 is approximately 2.5 billion euros. Recall that, as explained above, only one-fourth of this cost is the transfer obtained by mafia groups. The remaining three-fourths account for the welfare loss due to the reduction in production carried out by the infiltrated firms.

2.5 Alternative specifications

This section considers two extensions of the model presented in Section **??**. First, mafia impact is assumed to be positively correlated with firm-level productivity. In other words, the probability that a firm is asked to pay *pizzo* increases with its level of TFP. The second specification examines the reverse scenario, in which mafia targeting is negatively correlated with firm-level productivity.

2.5.1 Positively correlated distortions

In this scenario, the only deviation from the baseline model concerns the assumption on the relationship between the mafia targeting and firm-level TFP. Indeed, in this case, the distribution of mafia-related distortions is assumed to be positively correlated with firm i's draw of A_i. Specifically, firms can be ranked according to their individual probability of being impacted by mafia and this probability correlates positively with firm-level TFP. In each mafia-infiltrated market, firm i's maximisation problem can be rewritten as:

$$\Pi_i = (1 - \tilde{\tau}_i \delta) Y_i P_i - w_{st} L_i \tag{2.17}$$

The idiosyncratic component of the distortion, i.e. the "mafia exposure" parameter $\tilde{\tau}_i$, is now an interaction between two terms: (i) τ_i that is Bernoulli distributed with average λ (the observed share of firms that are forced to pay *pizzo*); and (ii) ζ_i that also follow a Bernoulli distribution, where the probability that ζ_i equals one increases with A_i, i.e. $\frac{\partial \mathrm{P}(\zeta_i=1)}{\partial A_i} > 0$. This interaction guarantees that the share of firms impacted by the mafia is exactly the share observed in the data. I estimate the model following the steps described in Section **??**. Table 3.4 shows the obtained results. The estimates of *pizzo* are comparable to the ones obtained in the baseline model.

2.5.2 Negatively correlated distortions

The second alternative specification considers the reverse assumption: mafia targeting and firm-level TFP are negatively correlated. Firms that operate in mafia-infiltrated markets maximise profit given by Equation 2.17. However, in this case, the "mafia exposure" parameter $\tilde{\tau}_i$ is the interaction between τ_i and γ_i, where the probability of being impacted by the mafia decreases with the firm-specific draw of A_i, i.e. $\frac{\partial \mathrm{P}(\gamma_i=1)}{\partial A_i} < 0$. Table 3.5 shows the obtained results, which are comparable to the ones estimated in the baseline model.

2.5.3 Model simulation and counterfactual analysis of the alternative specifications

In order to understand the composition of the cost due to mafia infiltrations in these alternative specifications, I perform the model simulation described above ("The extensive margin of the mafia"), in which I fix the value of *pizzo* δ and I let vary the share of impacted firms λ between zero and one. I focus on infiltrated firms and, for each value of

λ, I compute: (i) the total amount of *pizzo* paid by the firms that do not exit the market; (ii) the total forgone value added of the firms that reduced employment; and (iii) the forgone value added of the firms that leave the market.

Results are presented in Figure 2.4. The first panel of the figure reports the outcome of the simulation of the first scenario, where there is positive correlation between firms' productivity and probability of being coerced by the mafia. In this framework, the forgone value added due to employment reduction accounts for more than one half of the total cost, while the forgone value added due to firms' exit is not even one-fourth of the total cost. Reasonably, if mafia groups prefer to extort more productive firms, the impacted firms are more likely to reduce employment, and, in turn, produce less, than to incur negative profits, and, thus, exit the market. Instead, if mafia distortion is negatively correlated with firms' productivity, impacted firms are more likely to exit the market. This fact is highlighted in second panel of Figure 2.4, in which it can be noticed that the forgone value added due to firms' exit is higher than the previous case, accounting for more than one-third of the total cost.

In order to measure the magnitude of the total cost due to mafia infiltrations, I perform the same counterfactual analysis described in the previous section, where each infiltrated market is compared to its corresponding scenario without mafia infiltrations, i.e. where λ is set equal to zero. For each market, I compute the three components of the cost, i.e. the total transfer to mafia and the forgone value added due to both reduction in employment and firms' exit. Table 3.6 reports the results. By impacting firms in the upper tail of the TFP distribution mafia groups generate a total cost of approximately 4 billion euros. Not surprisingly, this cost is higher than the one computed in the baseline model, which is around 2.5 billion euros, and the one computed assuming negatively correlation between mafia targeting and productivity, which is approximately 1.8 billion euros.

2.6 Concluding remarks

Despite the considerable attention and the many efforts that have been made in studying the development and the functioning of organised crime, the nature and the magnitude of the economic cost of this phenomenon is still a topic of active research. This chapter tries to answer this question by proposing a new estimation approach that integrates tools from panel data analysis and structural econometrics. I study the economic consequences of mafia diffusion in northern Italy. In particular, I focus on the relationship between extortion, a typical mafia activity, and resource misallocation.

I build on the results presented in the previous chapter where I theorise the mechanism through which mafia infiltrations imply misallocation by introducing distortions in the alignment of the rankings of firm-level productivity and firm-level size. In this chapter I develop a methodology to measure the amount of *pizzo* and, consequently, the economic cost due to resource misallocation. This method uses panel data that vary across sector-province-year in a structural model of resource misallocation. The fact that only specific sectors of the economy, located in specific provinces, observed after a given point in time, are defined as mafia-infiltrated, enables to estimate all the factors that explain allocative efficiency beyond the mafia using information on non-mafia market. These parameters and actual data on mafia diffusion are then used to estimate the *extra* determinant of misallocation that depends on mafia extortion racketing.

This method delivers estimates of the average sectorial *pizzo* that ranges between 1 and 8 percent of the output level of the taxed firms. The counterfactual analysis performed using these estimates, measures the total cost suffered by the northern Italian economy during years 2000-2012. This cost is approximately 2.5 billion euros. A crucial result of this analysis, which goes beyond the quantification of *pizzo* and of the cost of the mafia, is that the resources that mafia groups coerce from the impacted firms account for only one-fourth of the total cost. The monetary cost of misallocation implied by extortion is three times higher the amount of extortion transferred to mafia groups.
To sum up, this analysis proposes new insights into the economic consequences of the Italian mafia, that are relevant for the policy. Moreover, it builds on the existing contributions on the economic consequences of weak institutions, by proposing a method to measure something that is not observed. Finally, it opens a methodological debate on how to use structural models to give sense to panel data.

Chapter 3

The economic cost of internal conflicts

3.1 Introduction

The danger of internal conflict and its potential importance for the process of economic and social development hardly needs emphasis. The set of countries which are failing most chronically to meet the Millennium Development Goals (MDGs) consists disproportionately of conflict-affected states. This has been recognised by the international community and the analysis of, and initiatives for, fragile and conflict-affected states (FCS) have been growing rapidly. In addition to addressing drivers of conflict and fragility these initiatives, so is the hope, should improve economic development.

Civil wars affect economic conditions in two stages: during conflict and during the post-war period. In this chapter (extract of Mueller et al. (2017)) we focus our analysis by exploring the first stage, i.e. the relationship between violence and economic activity during conflict. It turns out that the contemporaneous cost of conflict is essential to comprehend the observed lack of long term development in FCS. This is because many FCS do not experience stable peace (as shown in Section 3 of Mueller et al. (2017)). Economic recovery happens on average but it does not last long enough. This is why understanding the contemporaneous effects of conflict is essential to understand long term development.

However, the direct economic cost of conflict incidence is not the only channel through which civil wars impact the development of the affected territories. In this chapter we explore one channel in which the detrimental effects of conflict spread to the recovery period: the humanitarian crisis trigged by wars.[1] These effects will force growth in the

[1] Mueller et al. (2017) in Section 6 study the contraction of investments due to instability as another possible channel.

recovery period downwards. This is important as it indicates that the contemporaneous effects we turn towards now might be an underestimate.

This chapter sheds lights on the detrimental effects of internal conflicts by looking at the contemporaneous effect of conflict. We show, using standard panel fixed effect regressions, that economic growth during conflict incidence is significantly lower. This is true both at the country and ethnic group levels. Local disruptions of the economy in conflict seem to play an important role in this loss.

We then build on these findings by studying the humanitarian crisis triggered by civil war. In particular analyse the close relationship between violence and refugees. In the average civil war year around 500,000 persons leave their country. This is an important finding as it gives an idea of the scale of the humanitarian crisis entailed by mass violence.

3.2 Conflict and Economic Growth

The macroeconomic impact of civil wars has been studied extensively. Collier (1999) analyses the effects of violence on GDP growth using cross-county data on internal wars which occurred between 1960 and 1992. He finds that civil conflict is correlated with a contemporaneous reduction of GDP per capita growth of 2.2 percent.[2] On a regional level, Abadie & Gardeazabal (2003) investigate the economic effect of the Basque terror campaign and estimate a reduction of the GDP per capita of approximately 10 percent.

There are several channels through which violence retards development. During conflict there is a direct cost caused by the destruction of resources that would have been employed in production. This directly impacts contemporaneous economic performance in the affected territory and can affect recovery as well. In addition, the economic activity of countries involved in civil conflicts is damaged indirectly by an increase in production costs and insecurity in transport. Fear spreads the economic costs of conflict. We will show, for example, that the number of people running away from violence is far larger than the number of fatalities. Apart from this very direct impact of fear there are several indirect effects including the deviation of resources towards armament or the faltering of investment due to a lack of perspective. All these ideas have been explored in the academic literature.[3]

In what follows we will take an agnostic view regarding the channel through which the contemporaneous effect of violence on growth arises. Instead, we will simply look at correlates at the country and ethnic group level. It is nonetheless an important question

[2]This effect is comparable to estimates we compute and discuss in this section.

[3]See Blattman & Miguel (2010) for a review of the literature. Mueller (2013) provides a recent review of the literature on the economic costs of violence.

whether the effect of violence is coming from the local level or whether the main channel is the political and state crisis that coincides with civil war. We will turn towards this first.

3.2.1 The Per Capita Model of Violence

Assumptions regarding the main channels through which violence affects the economy have a direct impact on how the issue needs to be investigated empirically - even at the macro level. If local effects drive the economic costs of conflict then we expect countries with high per capita violence to suffer most economically. If macro effects are most important, the economy of the whole country will be affected by local violence. The economic impact of a fatality would then be independent of population size. In other words, if violence was only a sign of a national fragility which also affected the economy then we would expect that the same intensity of violence in India and Nicaragua have the same growth effect. At the same time, local and macro effects call for different policy responses. If effects are predominantly, local then the policy response should also have a strong local element.

There are good reasons to believe that both channels are active. Ksoll et al. (2009), for example, provide an example of local effects in their study of election violence in Kenya. Their paper provides direct evidence for the increase in labour costs that occur due to local violence risk. In a recent work Amodio & Di Maio (2017) also provide evidence for similar increases in costs due to break-down in transportation links caused by violence locally. At the same time the disruption of transport will affect international trade. Blomberg & Hess (2006) argue that the presence of violence is equivalent to a 30 percent tariff on trade. Martin et al. (2008) estimate a reduction of trade with conflict of between 20 and 25 percent. In addtion, there are direct state level effects as well. Collier et al. (2003), for example, show that during civil war countries increase their military expenditure from 2.8 percent of 5 percent of GDP. Mueller (2016) argues that if this is the main channel in which violence affects the economy then it could be regarded as pure public bad at the country level. All inhabitants of the country will suffer from the distortion in budgets regardless of their direct exposure to violence or the population size of the country. However, it can be shown that a per capita model of violence leads to a better description of the impact of conflict on the economy.

Figure 3.1 demonstrates that aggregate growth changes with the per capita intensity of violence. It displays growth averages for different levels of violence intensity controlling for country and year fixed effects. Violence intensity in the graph is displayed as the number of battle-related deaths from UCDP/PRIO per 1000 population.[4] Each observation in the Figure contains 10 percent of all country/year observations which experienced some violence. The observation at the far right contains the years with the most intense violence. These countries experienced extreme levels of violence with more than 2 fatalities per 1000 population (0.002 fatalities per capita). In this group growth was 5.2 percentage points lower compared to average growth during peacetime in the same country. Growth in the three next most affected groups was about 3 percentage points lower, despite the fact that violence per capita in these groups was only around 0.2 fatalities per 1000 population. Below this intensity, the effect of violence on growth cannot be distinguished from zero with this data. A populous country like India, for example, can grow seemingly uninhibited by constant violence on its territory. However, local economic effects could be just as strong.

The main insight from Figure 3.1 is that per capita violence provides a good approximation of the impact of violence on growth which comes mainly from the local level. New data from the UCDP GED and G-Econ projects for the African continent allows us to provide even more direct evidence for this. Mueller (2016) shows, using this data, that the growth damage of violence can be traced to 100kmx100km cells. Each year with violence in such a cell reduces growth in the cell by more than 2 percentage points. High intensity violence of 100+ fatalities reduces growth by about 5.8 percentage points locally.[5] At this level, violence seems to be a pure public bad, i.e. the population size within the cell does not affect how much damage a fatality does to the economy. An easy way to understand the magnitudes involved here is to go back to the country level and to use the data on locally affected population from the micro data.

In Figure 3.2 we show the result. The x-axis displays the share of the population in a country affected by 100+ fatalities in their cell. Again, we generate ten groups defined by this measure and see how growth in these groups changes with increasing violence intensity. The group of the 10 percent most affected countries had over 65 percent of their population affected by violence in the same cell. Growth in these countries was in free fall - their economies grew by almost 9 percentage points less than average.

Figure 3.2 reveals that there is an almost linear relationship between the country growth rate and the share of population affected by violence locally. As the locally affected population increases, so does the damage to growth associated with conflict. This suggests that understanding the local impact of violence is therefore key to understanding its overall impact. An absolute count of fatalities is then a misleading measure of violence

[4]For a data description see the appendix. The most intense civil war in this regard took place in the Lebanon which experienced more than 20 fatalities per 1000 population.

[5]These results are derived with cell and year fixed effects so that omitted variable bias is less likely to drive the result.

intensity. Throughout this chapter we will use this insight and approximate conflict intensity not through fatality counts but through fatality per capita counts.

3.2.2 Conflict Incidence and Economic Performance

As a first step into the analysis we run standard panel regressions of growth on conflict incidence. For this purpose we assemble three indicators of country-level economic performance during the post War World II period. The first and the second indices are GDP per capita growth computed using data provided respectively by Penn World Tables and World Bank Open Data. The third proxy for economic activity is given by per capita growth of night light, computed using satellite data from the National Oceanic and Atmospheric Administration (NOAA).[6] Night light data has the benefit of being available on a yearly basis independently of the quality of local statistical offices and data gathering. While it comes with its own problems it can shed light on local economic activity where gathering of statistical data is incomplete.[7] This makes it a great fit for measuring growth in a context of civil conflict. Conflict incidence is measured through the number of battle-related deaths from UCDP/PRIO dataset.

We run the following regression for country i at time t:

$$g_{it} = \beta \times \text{incidence}_{it} + \mu_i + \eta_t + \epsilon_{it} \tag{3.1}$$

where g_{it} is economic performance per capita growth of country i in year t, incidence_{it} is conflict incidence, μ_i and η_t are respectively country and year fixed effects.

A cross-country analysis as in equation (3.1) bears considerable potential for both reverse causality and omitted variable bias. Thus, a priori, a convincing causal link is hard to establish. However, here we expect the resulting bias to be small for two reasons. First, the cross-country literature has not found that negative, contemporaneous shocks to growth systematically lead to violence.[8] Second, we have run a large number of robustness checks by adding time trends or lagged growth to our specification and controlling for rainfall shocks directly.[9] The upshot from this is not only that results remain significant

[6] Satellite data is available for a shorter time period, 1992-2013.

[7] For a discussion see Henderson et al. (2012). In order to calculate light per capita we use population data that is provided by World Bank dataset.

[8] The standard reference here is Miguel et al. (2004). Ciccone (2011) shows that high rainfall levels three years earlier seem to be best predictors of conflict in the reduced form. Miguel & Satyanath (2011) argue that lagged negative growth shocks are a predictor of conflict onset. In any case, there is no evidence from this literature that contemporaneous growth declines cause conflict. Bazzi & Blattman (2014) corroborate the view that the relationship between income shocks and conflict is not straightforward. They do not find evidence of an effect of price shocks on conflict onset and only weak evidence on incidence.

[9] Results from this are presented in the Appendix of this chapter.

but also that the estimated coefficients barely change. This does not mean that a causal link from falling growth to conflict can be ruled out. But it is unlikely to drive the macro relationship we see in the data.

In order to further explore the relationship between violence and country-level output we run two specifications of the model described above. In the first model conflict in country i at time t is defined by any violence, i.e. if at least one battle related deaths occurs. In the second specification, conflict is defined by a higher threshold, by 0.008 deaths per 1000 population.[10]

We expect to get different results from the two specifications. From the analysis of Figure 3.1 we know that economic damage of civil war increases with the severity of conflict. The estimated impact from the second model should therefore be more acute.

Table 3.7 panel A and B reports the results. Each column contains one of our measures for economic growth. In all specifications, conflict incidence correlates negatively with country-level economic performance. The estimated coefficients of conflict incidence are statistically significant and negative. We also find that the coefficients reported in panel B are approximately twice as high (in absolute value) as the ones shown in panel A.

In Table 3.7 panel B we find that the estimated coefficients of the incidence variable reported in columns (1) and (2) are fairly similar. According to these results, civil conflict in country i at time t correlates negatively with GDP per capita growth which ranges from 2 to 3 percentage points. In column (3) we use night light as an economic indicator. The estimated effect of conflict on night light is statistically significant and is equal to -0.075. What does this number mean in terms of GDP per capita growth? One way to think about this are the estimates from Henderson et al. (2012) who argue that the relationship between GDP and light can be expressed fairly well in a constant elasticity model in which an increase of night light by 1 percent implies an increase of GDP of about 0.25 percent. If we apply this model to our result from Table 3.7 panel B, the conflict reduces GDP growth by almost 1.9 percentage points. This is strikingly similar to the 2 percent found in column (1).

Another way to understand the size of the decline in economic performance with conflict is to use the estimate of $\hat{\beta}$ in equation (3.1) to simulate how output in each country would have developed as a function of the number of years in which the country was affected by civil conflict. For the simulation we focus on the Post War Word II period, i.e. 1946-2014 and count the number of years each country has been affected by conflict. Call this number T_i. In Table 3.7 panel B column (1) we estimated that for each

[10]We take the threshold from Mueller (2016) who shows that a threshold like this leads to a similar number of coded civil wars as the threshold of 1000 battle-related deaths often used in the conflict literature. In the context here, this is a conservative approach as it is not the threshold which yields the biggest difference between conflict and non-conflict countries.

year a country is in conflict the GDP per capita will decrease by around 2 percentage points. From this we calculate the total loss for country i as

$$\text{OutputLoss}_i = (1 + \hat{\beta})^{T_i} - 1$$

where $\hat{\beta}$ is the estimate of the panel regression (3.1). This simulation predicts output, had the country not experienced conflict but otherwise kept its growth path.

The findings are displayed in Figure 3.3. Around 70 percent of countries hit by conflict experienced only one or two years of high intensity violence which, according to our estimates, would have directly lowered their output by 2 to 4 percentage points. The remaining countries experienced conflict for a longer time period. Accordingly, the total output loss in these groups ranges from around 10 to over 50 percentage points. These are quite striking magnitudes.

In order to interpret these results it is important to keep in mind that our estimates are coming from a country fixed effects regression. In this way we are comparing growth in a country during conflict to other years in the same country without conflict. The hope is that factors which affect the country generally are captured by the fixed effect. However, our empirical strategy also implies that if conflict affects growth in the post-conflict period then $\hat{\beta}$ will provide an underestimate of the true economic costs of conflict. We will return to this point below.

3.2.3 Economic Impact at the Ethnic Group Level

As a final piece of evidence, we draw from economic, population and ethnic conflict data compiled in the GROWup dataset. This dataset provides satellite night light data as a proxy for economic activity at the ethnic group level. Conflict incidence in the GROWUp data is defined through more than 25 fatalities, i.e. a fairly low threshold, in ethnic conflicts. This dataset provides information on 502 ethnic groups for the period 1992 to 2013. We run a regression of the following form:

$$g_{it} = \theta_i + \eta_t + \beta \times \text{incidence}_{it} + \epsilon_{it}$$

where g_{it} is night light per capita growth of group i in year t, incidence_{it} is conflict incidence, θ_i are a set of 502 group fixed effects and η_t are year fixed effects. The fixed effects in this context control for group-specific growth trends in the data. Note that we do not use the per capita model in this specification as violence seems to be a pure public bad locally.

Table 3.8 reports the results. We find a negative and statistically significant effect of incidence on per capita light growth. Light per capita growth is reduced by around 7 percentage points in years a group is engaged in conflict. This effect is robust to controlling for lagged light growth in column (2). In columns (4) and (5) we show that these reductions in light per capita growth are composed by two factors; light growth falls dramatically and this is partially offset by a negative effect on population growth. However, column (3) also shows that there is some evidence for negative spill-overs from conflict at the country level. Conflict in other groups in the same country lower light per capita growth for other groups in the same country by 0.7 percentage points.

Again we interpret these estimates in terms of GDP growth following Henderson et al. (2012). If a decrease of night light by 1 percent implies a decrease of GDP of about 0.25 percent, then our result from column (1) in Table 3.8 can be translated into a reduction of GDP growth of about 1.8 percentage points.

Overall this suggests that local effects are important and comparatively large in comparison to spill-overs. This is important for our discussion as they define the challenge that the country faces in the recovery from civil wars. Local effects of violence imply that civil conflict leaves a legacy of regional economic divergence. Another way to understand the size of this relative decline is to perform a second simulation of the GDP loss that focuses on group-level economic performance, following the same steps described above. Figure 3.4 displays the distribution of conflict lengths in the sample for which we also have the night light per capita data (years 1992-2012). We estimated that, for each year a group is in conflict, GDP per capita in its homeland will decrease by 1.8 percentage points. As a result, around one half of the ethnic groups is estimated to have a lost of output that ranges between 10 and 30 percent. If this is not recovered it will lead to regional imbalances on the country level and could amplify conflict.

3.3 Internal Conflicts and Refugees

A particularly serious aspect of internal conflicts is the human suffering it generates. This is not only those who are killed or injured in conflict but the large number of people who are forced to leave their homes. The issue of refugees has received particular attention in Western Media in recent years as refugee flows from the Northern Africa, the Middle East and Afghanistan are increasingly reaching Europe. These refugee streams are linked to a severe humanitarian crisis with considerable funding needs for international donors and heavy strains on host countries. The current refugee crisis, however, is in no way unique. Civil war has always been closely linked to humanitarian crisis and refugee streams are one way to capture this.

In this section we provide a cross-country analysis aimed at investigating how the stock of refugees evolves when a civil conflict hits a country. In the analysis we will focus entirely on showing changes in the stock of refugees across time to illustrate the dimensions involved. We will base our later analysis on these population movements.

We exploit country-level data gathered from several sources. Data about refugees is provided by the UNHCR Population Statistics Database. The database provides information about UNHCR's populations of concern from the year 1951 up to 2014. This database lists seven categories: refugees, asylum-seekers, returned refugees, internally displaced persons (IDPs), returned IDPs, stateless persons and others of concern. For each group the database provides yearly information about their composition by location of residence and origin. We exploit only the data on "refugees".[11] In particular, we are interested in the annual stock of refugees for each country of residence, i.e. how many people with refugees status have left their home country each year. We focus on these numbers as they appear to be the most comparable across time and countries. However, this is likely to capture only the tip of the iceberg in some cases. The number of IDPs is extremely high in some instances but cannot be captured with the same level of confidence as refugees generally.[12]

Cross-country data about conflict is provided by the UCDP/PRIO. As for the index of country-level economic activity, we use again information provided by Penn World Table and World Bank databases.

As mentioned above, our aim is to explore the dynamics of refugees during conflicts. In other words, we attempt to answer several questions. Do people run away from their home country when a conflict breaks out? Does the seriousness of the conflict matter in this decision? In which phase of the conflict do they leave? When do refugees come back to their home country?

In order to answer these questions, we look at the impact of conflict incidence on the yearly stock of refugees. More formally, we run the following regression for country i in year t:

$$\text{refugees}_{it} = \text{incidence}_{it} + \text{phase}_{it} + \delta_i + \delta_t \ + \ \epsilon_{it}$$

Our outcome of interest is refugees_{it}, which is the stock of refugees. The variable incidence_{it} is conflict incidence. In one specification of the model we use the definition from PRIO dataset of at least 25 battle related deaths caused by an internal conflict.

[11] According to the UNHCR definition, this category includes "individuals recognized under the 1951 Convention relating to the Status of Refugees; its 1967 Protocol; the 1969 OAU Convention Governing the Specific Aspects of Refugee Problems in Africa; those recognized in accordance with the UNHCR Statute; individuals granted complementary forms of protection; or those enjoying temporary protection; and people in a refugee-like situation".

[12] The UNHCR Global Trends Report 2014 provides evidence that confirms this hypothesis. About 59.5 million people were forcibly displaced worldwide by the end of year 2014. Among them, 19.5 million were refugees and 38.2 million were IDPs.

In the other specifications we use a more stringent definition of internal conflict which requires at least 1000 battle related fatalities.[13]

Variable phase_{it} is a vector of dummies we use to show the dynamics in the stock of refugees in different phases of the conflict. In one specification we control for the *year before the conflict starts* and *first year of conflict*. Then, we include in succession dummies for the *last year of conflict* and *recovery years* up to the tenth year of recovery.

Finally, δ_i and δ_t are respectively country and year fixed effects: the former control for time invariant country-specific impact on the change of the number of refugees, while year fixed effects capture the average state-invariant influence of each year t on refugees_{it}.

Table 3.9 reports the results. In column (1) conflict incidence coincides with the PRIO definition of at least 25 battle related deaths, i.e variable *25+ battle deaths*. The estimated coefficient indicates that during conflicts more than 144,000 people leave their home country on average. This average rises markedly when we use the more stringent definition of at least 1000 battle related fatalities as conflict incidence variable. In fact, in column (2) the estimated effect of conflict incidence on $refugees_{it}$ more than doubles. Note that in this specification we include the dummy *25-1000 battle deaths*[14] in order to check whether people have an incentive to leave their country when some violence occurred, but the situation has not escalated into a conflict with more than 1000 deaths in a given year. We notice that during these years a significant number of people leave the country, but this average is very low if compared with the average stock of refugees during the violent years with more than 1000 deaths. We therefore focus on dynamics in war period, using as "conflict incidence" the tighter definition of at least 1000 battle related deaths.

Column (3) of table 3.9 adds to the previous model the controls for the initial phase of the conflict, namely *year before the war* and *first year of the war*. Compared to column (2), the estimated coefficient of the variable *1000+ battle related deaths* increases. On average, the stock of refugees during conflict is more than 487,000 people. But in the first year of conflict the average stock is by 384,851 lower, i.e. only 20 percent of the almost 500,000 refugees leave their country during the first year of conflict.[15] Column (3) also tells us that there is there is no significant departure during the year before the conflict. This suggests that forced displacement of people takes off with the start of violence but not before.

Column (4) includes controls for the last phase of the conflict, namely the *last year*

[13]Note, that we now switched to counts of fatalities as our outcome variable is also in absolutes. Weighting both sides by population would not change the results.

[14]This dummy takes value one if in country i during year y the registered battle related deaths are between 25 and 1000.

[15]The number of 20 percent is obtained by computing (487,968-384,851)/487,968.

of conflict and *first year of recovery*. Compared to the specification of column (3), the coefficient of conflict incidence does not change significantly. During the last year of the civil wars, around one half of refugees are still out of their home country (the stock of refugees is approximately 210,000 on average). This amount decreases during the first year of recovery. In fact, the estimated coefficient of the variable *first year of recovery* indicates that, on average, only about 20 percent of the refugees have not come back into their home country in the first recovery year. Columns (5) and (6) add controls for additional years of recovery, i.e. the two specifications includes dummies respectively for five and ten years of recovery. We can observe that the stock of refugees decreases with longer recoveries. The estimated coefficient of the variable *10+ recovery years* in column (6) suggests that around 70,000 refugees who left their home country during civil war have not returned 10 years after the war has ended. These are about 15 percent of all people who leave.[16]

We have also performed the same analysis excluding Afghanistan. Given that the yearly stock of refugees registered in Afghanistan is significantly higher than the stock of any other country, we want to exclude the hypothesis that Afghanistan is driving the results we described above. We compute the equivalent of Table 3.9, that we do not report here. Although coefficients are smaller in size, our main conclusions are left unchanged. The results are also robust to controlling for famine episodes as shown in Table C6 in appendix.[17]

3.4 Concluding remarks

In summary, in this extract of Mueller et al. (2017), using panel data regressions, that economic growth decreases dramatically when a country or region experiences violence. As the economic effects of violence are predominantly local this implies that conflict might lead to or amplify regional imbalances.

When we study the humanitarian crisis triggered by civil war, looking at the relationship between violence and refugees, we drive three main conclusions. First, our analysis suggests that violence leads to a severe humanitarian crisis which forces people to leave their country of origin. Second, residents seem to migrate as refugees when the internal conflict breaks out, in other words, they don't leave the country before the conflict has actually started. After conflict, the large majority of refugees returns very quickly, i.e. within a year after the end of conflict, 4 out 5 refugees have returned. However, the

[16]The number of 15 percent is obtained by computing 73,528/466,418.

[17]Famine and conflict are closely associated in the sense that the risk of a famine more than doubles with conflict. In addition, famine leads to substantial flight. We discuss details in the Appendix.

average return flow then slows markedly and ten years after the conflict ended 3 out 20 refugees remain outside their country of origin.

Bibliography

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: A case study of the basque country. *American economic review*, 113–132.

Acconcia, A., Corsetti, G., & Simonelli, S. (2014). Mafia and public spending: Evidence on the fiscal multiplier from a quasi-experiment. *American Economic Review*, *104*(7), 2185-2209.

Acemoglu, D., Johnson, S., & Robinson, J. A. (2001). The colonial origins of comparative development: An empirical investigation. *American Economic Review*, *91*(5), 1369-1401.

Amodio, F., & Di Maio, M. (2017). Making do with what you have: Conflict, input misallocation and firm performance. *The Economic Journal*, *128*(615).

Angelini, P., & Generale, A. (2008). On the evolution of firm size distributions. *American Economic Review*, *98*(1).

Atkeson, A., & Kehoe, P. J. (2005). Modeling and measuring organization capital. *Journal of Political Economy*, *113*(5), 1026-1053.

Balletta, L., & Lavezzi, M. (2014). Extortion, firm's size and the sectoral allocation of capital. *MIMEO*.

Barone, G., & Narciso, G. (2015). Organized crime and business subsidies: Where does the money go? *Journal of Urban Economics*, *86*(C), 98-110.

Bartelsman, E., Haltiwanger, J., & Scarpetta, S. (2013). Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review*, *103*(1), 305-334.

Bazzi, S., & Blattman, C. (2014). Economic shocks and conflict: Evidence from commodity prices. *American Economic Journal: Macroeconomics*, *6*(4), 1–38.

Besley, T., Fetzer, T., & Mueller, H. (2015). The Welfare Cost Of Lawlessness: Evidence From Somali Piracy. *Journal of the European Economic Association*, *13*(2), 203-239.

Besley, T., & Mueller, H. (2015). *Volatility and political institutions: Theory and application to economic growth* (Tech. Rep.). CEPR Discussion Papers.

Besley, T., & Mueller, H. (2018). Predation, protection, and productivity: A firm-level perspective. *American Economic Journal: Macroeconomics*, *10*(2).

Blattman, C., & Miguel, E. (2010). Civil war. *Journal of Economic Literature*, *48*(01), 3–57. Retrieved from https://www.aeaweb.org/articles.php?doi=10.1257/jel.48.1.3 doi: 10.1257/jel.48.1.3

Blomberg, S. B., & Hess, G. D. (2006). How much does violence tax trade? *The Review of Economics and Statistics*, *88*(4), 599–612.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, *77*(3), 623-685.

Buonanno, P., Durante, R., Prarolo, G., & Vanin, P. (2015). Poor institutions, rich mines: Resource curse in the origins of the sicilian mafia. *The Economic Journal*, *125*(586), F175-F202.

Buonanno, P., & Pazzona, M. (2014). Migrating mafias. *Regional Science and Urban Economics*, *44*(C), 75-81.

Cederman, L.-E., Wimmer, A., & Min, B. (2010, 1). Why do ethnic groups rebel? new data and analysis. *World Politics*, *62*, 87–119. Retrieved from http://journals.cambridge.org/article_S0043887109990219 doi: 10.1017/S0043887109990219

Ciccone, A. (2011). Economic shocks and civil conflict: A comment. *American Economic Journal: Applied Economics*, *3*(4), 215–227.

CITpax. (2012). Actores armados ilegales y sector extractivo en colombia, v informe. Retrieved from http://www.askonline.ch/fileadmin/user_upload/documents/Thema_Wirtschaft_und_Menschenrechte/Bergbau_Rohstoff/Gold/Actores_armados_ilegales_sector_extractivo.pdf

Collier, P. (1999). On the economic consequences of civil war. *Oxford economic papers*, *51*(1), 168–183.

Collier, P., Elliott, V., Hegre, H., Hoeffler, A., Reynal-Querol, M., & Sambanis, N. (2003). *Breaking the conflict trap: Civil war and development policy (washington, dc: World bank).* Oxford University Press.

Daniele, G., & Geys, B. (2015). Organised crime, institutions and political quality: Empirical evidence from italian municipalities. *The Economic Journal*, *125*(586), 233-255.

Donati, C., & Sarno, D. (2015). Are firms in "backward" areas of developed regions more financially constrained? the case of italian sme's. *Industrial and Corporate Change*, *24*(6), 1353.

Galletta, S. (2017). Law enforcement, municipal budgets and spillover effects: Evidence from a quasi-experiment in italy. *Journal of Urban Economics*, *101*, 90 - 105.

Gambetta, D., & Reuter, P. (2000). Conspiracy among the many: the mafia in legitimate industries. In N. G. Fielding, A. Clarke, & R. Witt (Eds.), *The economic dimensions of crime* (pp. 99–120). London: Palgrave Macmillan UK.

Hall, R. E., & Jones, C. I. (1999). Why do Some Countries Produce So Much More Output Per Worker than Others? *The Quarterly Journal of Economics*, *114*(1), 83-116.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. *American Economic Review*, *102*(2), 994-1028. Retrieved from http://www.aeaweb.org/articles.php?doi=10.1257/aer.102.2.994 doi: 10.1257/aer.102.2.994

Hsieh, C.-T., & Klenow, P. J. (2009). Misallocation and Manufacturing TFP in China and India. *The Quarterly Journal of Economics*, *124*(4).

Konrad, K. I., & Skaperdas, S. (1998). Extortion. *Economica*, *65*(260), 461-477.

Ksoll, C., Macchiavello, R., & Morjaria, A. (2009). Guns and Roses: The Impact of the Kenyan Post-Election Violence on Flower Exporting Firms.

Martin, P., Mayer, T., & Thoenig, M. (2008). Civil wars and international trade. *Journal of the European Economic Association*, *6*(2-3), 541–550.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, *57*(5), 995-1026.

Miguel, E., & Satyanath, S. (2011). Re-examining economic shocks and civil conflict. *American Economic Journal: Applied Economics*, *3*(4), 228–232.

Miguel, E., Satyanath, S., & Sergenti, E. (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy*, *112*(4), 725–753.

Mueller, H. (2013). The economic cost of conflict. *IGC Working Paper*.

Mueller, H. (2016). Growth and violence: Argument for a per capita measure of civil war. *Economica*, *83*(331), 473–497.

Mueller, H., LaviniaPiemontese, & Tapsoba, A. (2017). Recovery from conflict : lessons of success. Retrieved from http://documents.worldbank.org/curated/en/832901487083457988/Recovery-from-conflict-lessons-of-success

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the Econometric Society*, 1417–1426.

Olley, G. S., & Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, *64*(6), 1263-97.

Piemontese, L. (2013). The spread of mafia in northern italy: The role of public infrastructure. *Rivista di Politica Economica*(4), 113-140.

Pinotti, P. (2013). Organized crime, violence and the quality of politicians: Evidence from southern italy. In O. M. Philip Cook Stephen Machin & G. Mastrobuoni (Eds.), *Lessons from the economics of crime: What works in reducing offending?* MIT Press.

Pinotti, P. (2015). The economic costs of organised crime: Evidence from southern italy. *The Economic Journal*, *125*(586), 203-232.

Ranasinghe, A. (2017). Property rights, extortion and the misallocation of talent. *European Economic Review*, *98*(C), 86-110.

Ranasinghe, A., & Restuccia, D. (2018). Financial frictions and the rule of law. *Journal of Development Economics*, *134*, 248 - 271.

Restuccia, D., & Rogerson, R. (2008). Policy Distortions and Aggregate Productivity with Heterogeneous Plants. *Review of Economic Dynamics*, *11*(4), 707-720.

Restuccia, D., & Rogerson, R. (2013). Misallocation and productivity. *Review of Economic Dynamics*, *16*(1), 1 - 10.

Scognamiglio, A. (2018). When the mafia comes to town. *European Journal of Political Economy*, *55*(C), 573-590.

Soares, R. R. (2010, Oct). *Welfare Costs of Crime and Common Violence: A Critical Review* (Tech. Rep. No. 581). Department of Economics PUC-Rio (Brazil).

Transcrime. (2013). Progetto pon sicurezza 2007-2013: Gli investimenti delle mafie. rapporto linea 1. Retrieved from www.investimentioc.it

UNODC. (2012). World drug report 2012. Retrieved from https://www.unodc.org/unodc/en/data-and-analysis/WDR-2012.html

Tables and Figures: Chapter 1

Figure 1.1: Sectorial Mafia Presence

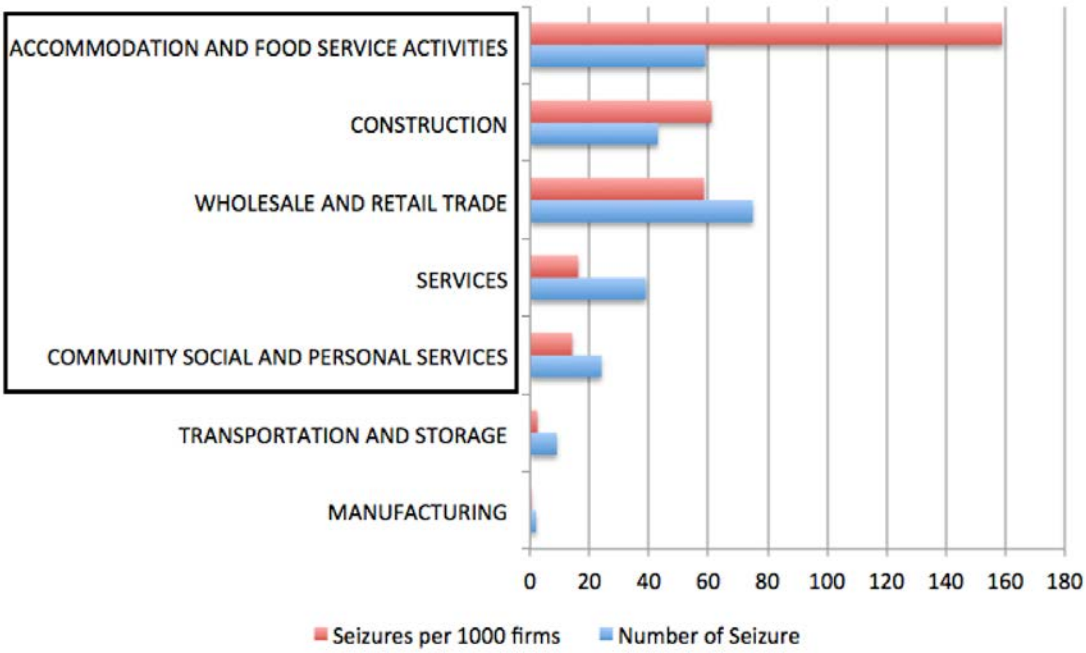


Figure 1.2: Provincial Mafia Presence

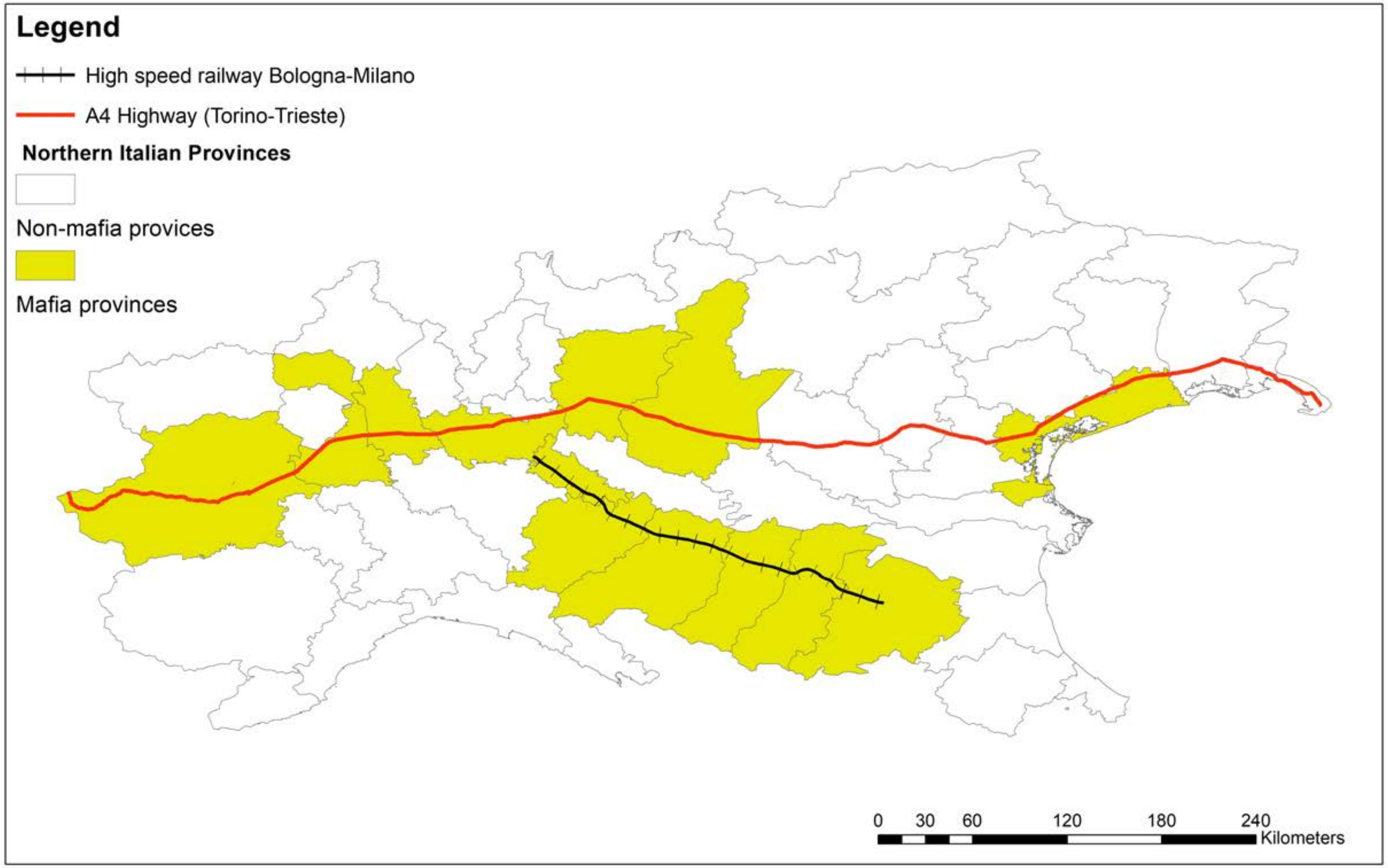


Figure 1.3: Sectorial Mafia Presence

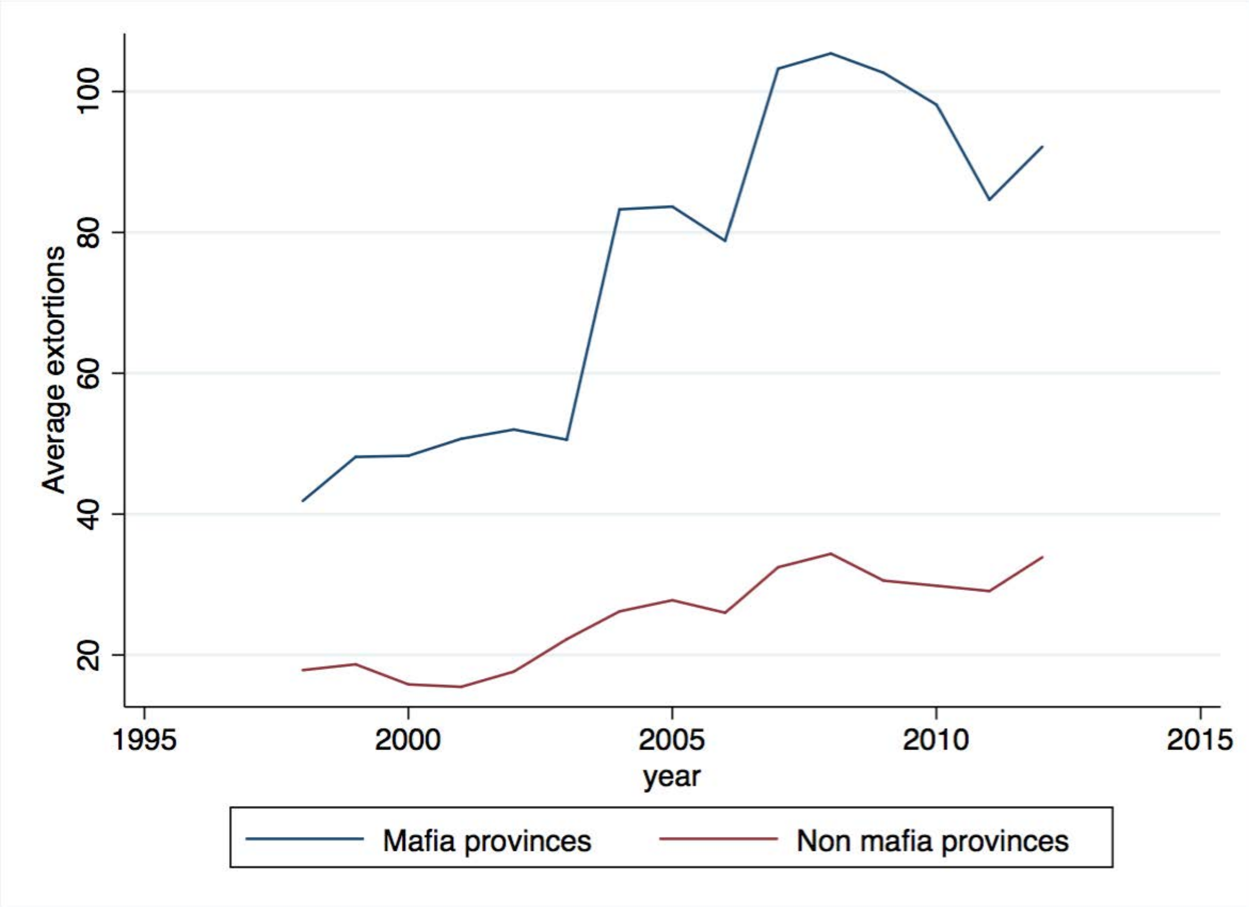
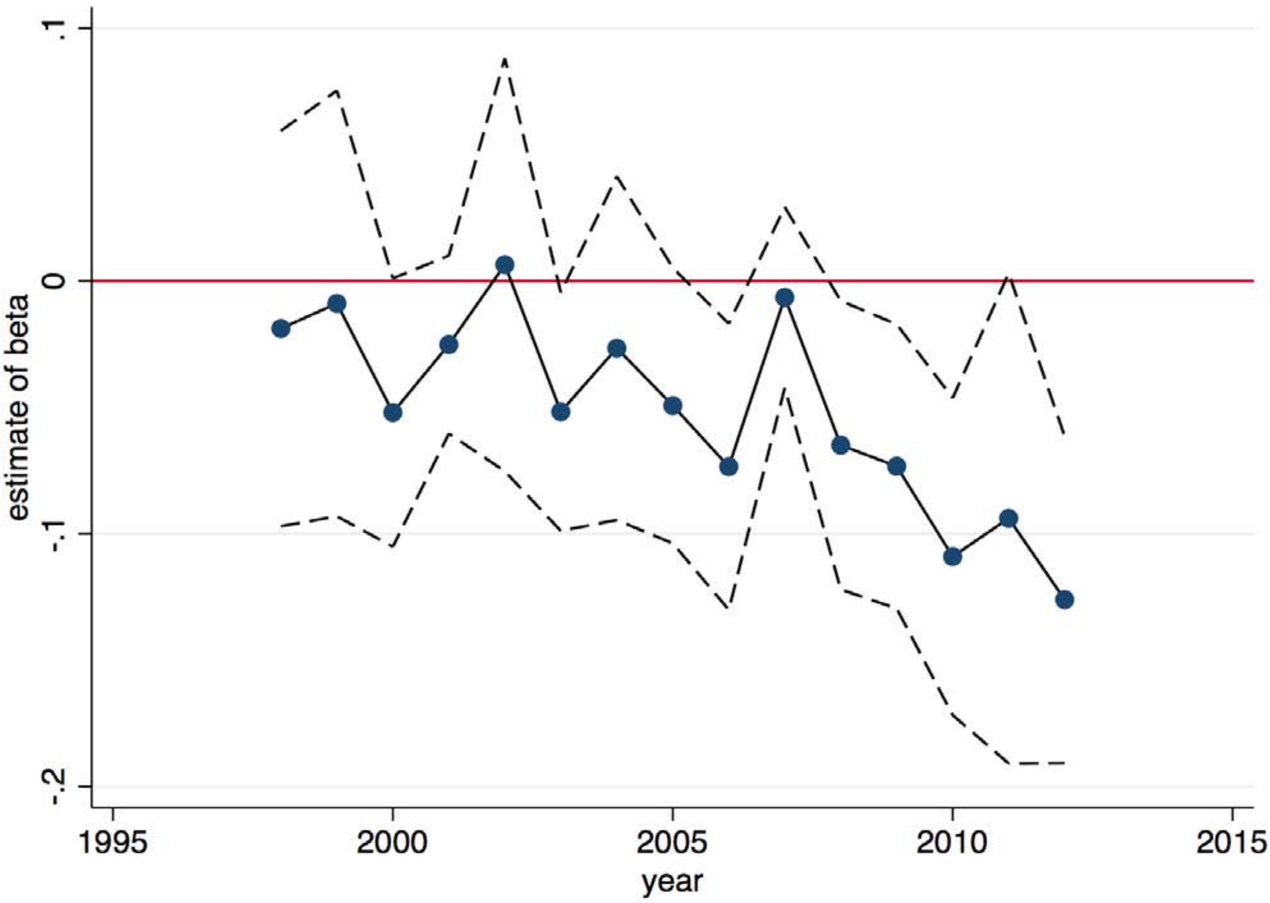


Figure 1.4: OP covariance in mafia infiltrated markets



Tables and Figures: Chapter 2

Figure 2.1: Amount of extortion δ

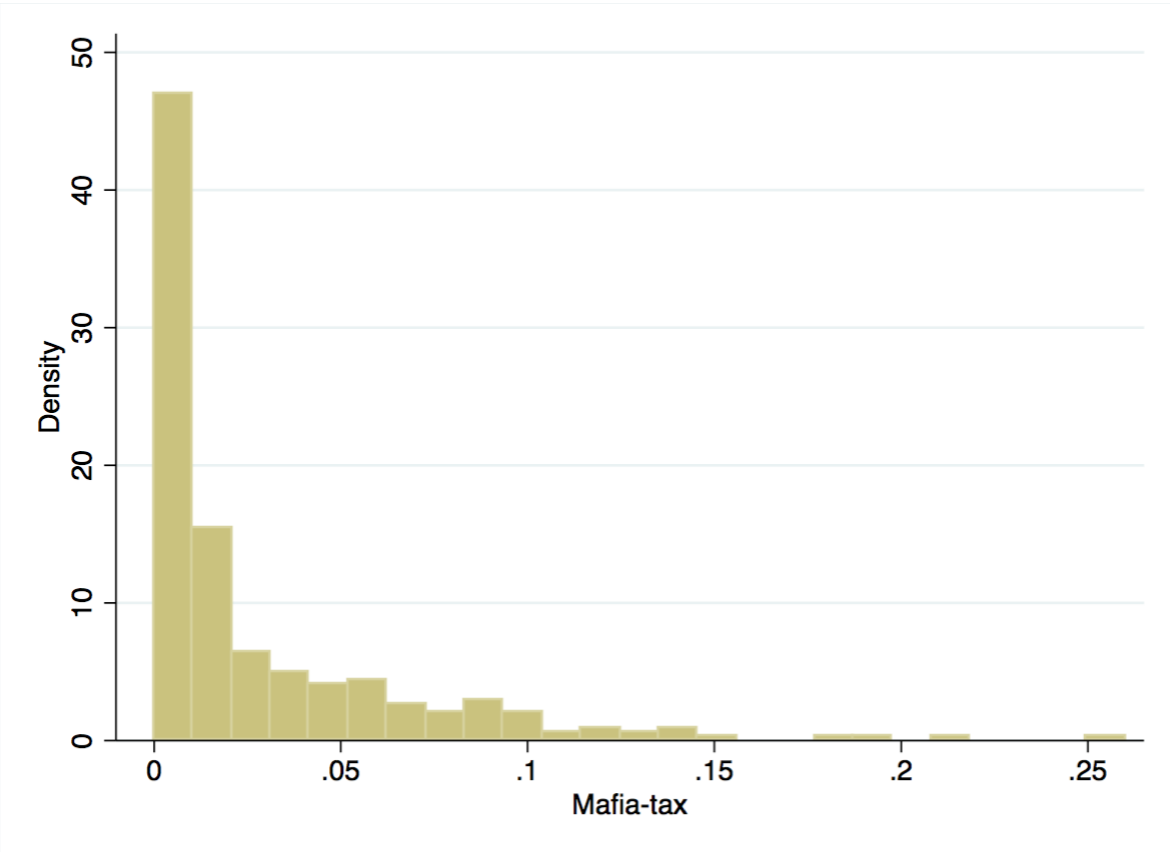


Table 3.1: Mafia-tax on output

Sector	Average by sector	Number of markets
Construction	3%	81
Wholesale and retail trade	1%	81
Accommodation and food services activities	8%	79
Services	3%	81
Community social and personal services	5%	81

Estimates of δ are obtained using data from the *Small and Medium Entrepreneurs Survey* available at ADELE laboratory-ISTAT, data from the yearly Book of Criminal Statistics published by ISTAT and information provided by the National Agency for the Administration and Management of Real Estate and Firms Confiscated from Criminal Organisation (ANBSC). Results come from the application of the Method of Simulated Moments (MSM) through a grid search. δ takes values from a grid bounded between 0 and 0.5.

Table 3.2: Mafia-tax on output

Sector	Average by sector	Number of markets
Food, beverages, tobacco	0.1%	80
Texiles	0.4%	80
Wood, Paper	0.3%	78
Chemicals	0.1%	81
Machinery	0.1%	81
Others	0.4%	80
Transportation and storage	0.2%	78

Estimates of δ are obtained using data from the *Small and Medium Entrepreneurs Survey* available at ADELE laboratory-ISTAT, data from the yearly Book of Criminal Statistics published by ISTAT and information provided by the National Agency for the Administration and Management of Real Estates and Firms Confiscated from Criminal Organisation (ANBSC). Results come from the application of the Method of Simulated Moments (MSM) through a grid search. δ takes values from a grid bounded between 0 and 0.5.

Figure 2.2: Model simulation

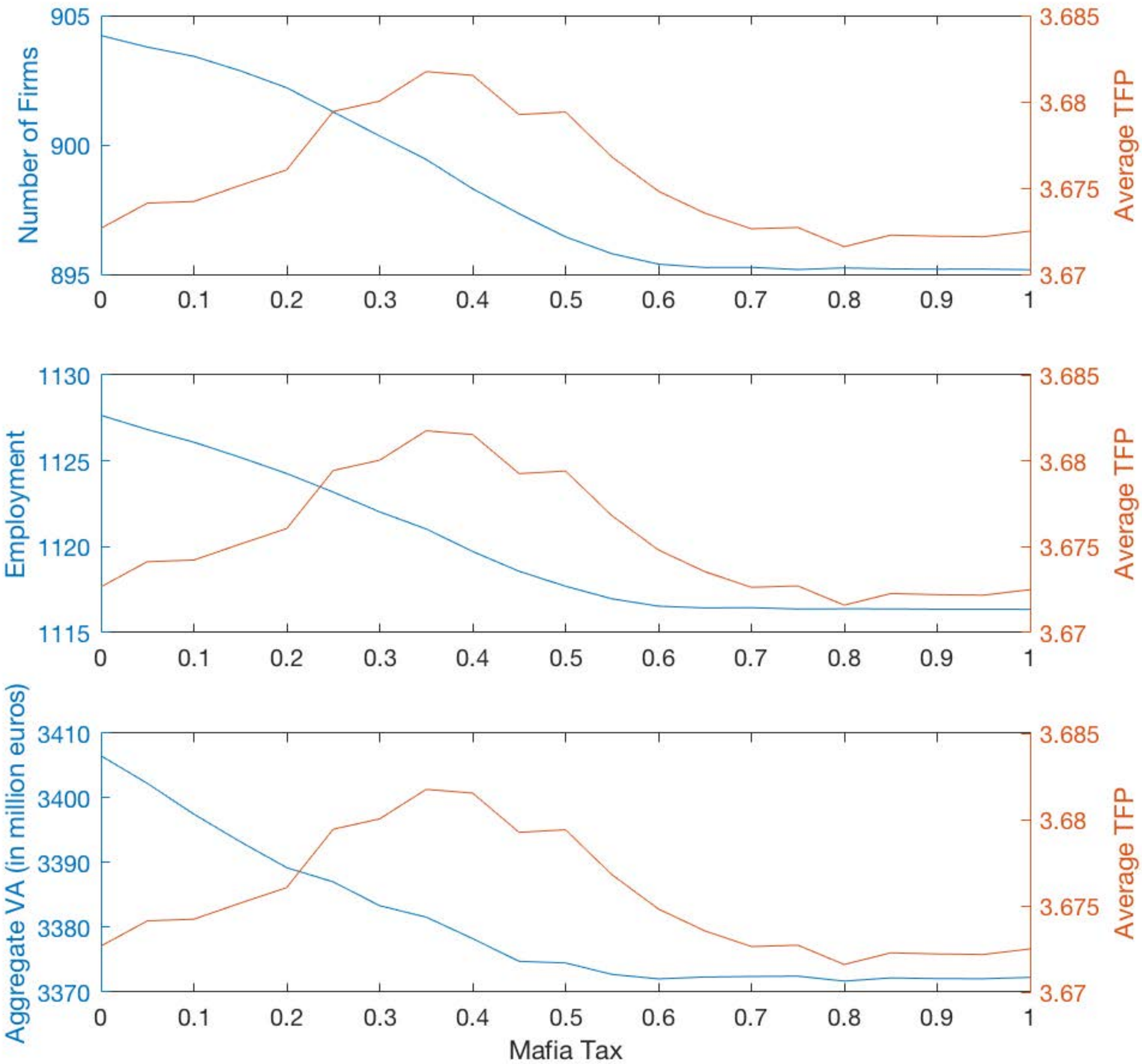


Figure 2.3: Model simulation

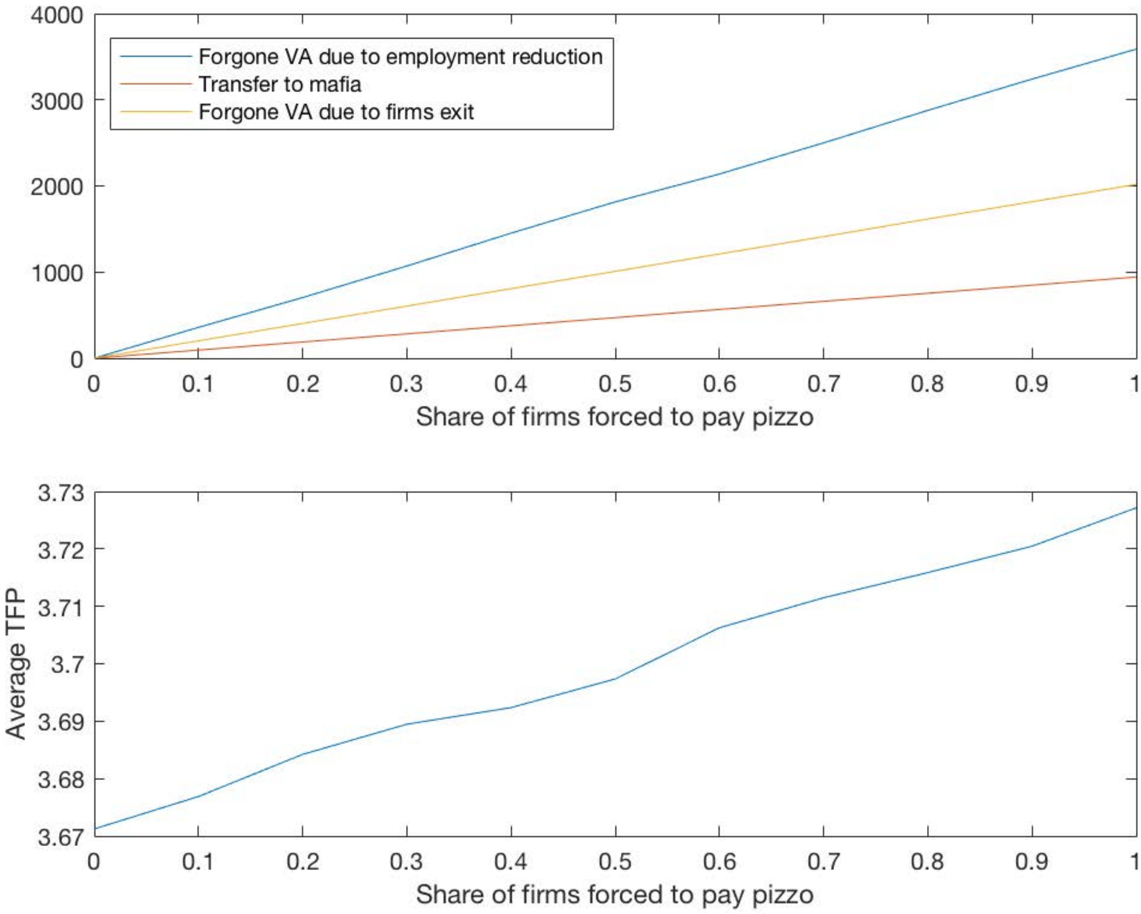


Table 3.3: Loss of value added due to mafia (in million euros)

Sector	Forgone value added	Transfer	Total cost
Construction	311.08	110.44	421.52
Wholesale and retail trade	313.62	107.35	420.97
Accommodation and food services activities	601.18	217.87	819.05
Services	251.15	90.09	341.24
Community social and personal services	379.14	134.60	513.74
Total	1856.17	660.35	2516.52

Results are obtained using data from the *Small and Medium Entrepreneurs Survey* available at ADELE laboratory-ISTAT, data from the yearly Book of Criminal Statistics published by ISTAT and information provided by the National Agency for the Administration and Management of Real Estates and Firms Confiscated from Criminal Organisation (ANBSC). Column "Forgone value added" is computed by aggregating the forgone value added due to employment reduction and to firm exit in each sector. Column "Transfer" collects the *pizzo* payed by each impacted firm in each sector. Column "Total cost" reports the sum of "Forgone value added" and "Transfer".

Table 3.4: Positive correlation between mafia targeting and firm-level TFP

Sector	Average by sector	Number of markets
Construction	2%	81
Wholesale and retail trade	1%	81
Accommodation and food service activities	7%	79
Services	3%	81
Community social and personal services	3%	81

Estimates of δ are obtained using data from the *Small and Medium Entrepreneurs Survey* available at ADELE laboratory-ISTAT, data from the yearly Book of Criminal Statistics published by ISTAT and information provided by the National Agency for the Administration and Management of Real Estates and Firms Confiscated from Criminal Organisation (ANBSC). Results come from the application of the Method of Simulated Moments (MSM) through a grid search. δ takes values from a grid bounded between 0 and 0.5.

Table 3.5: Negative correlation between mafia targeting and firm-level TFP

Sector	Average by sector	Number of markets
Construction	3%	81
Wholesale and retail trade	1%	81
Accommodation and food services activities	9%	79
Services	4%	81
Community social and personal services	5%	81

Estimates of δ are obtained using data from the *Small and Medium Entrepreneurs Survey* available at ADELE laboratory-ISTAT, data from the yearly Book of Criminal Statistics published by ISTAT and information provided by the National Agency for the Administration and Management of Real Estates and Firms Confiscated from Criminal Organisation (ANBSC). Results come from the application of the Method of Simulated Moments (MSM) through a grid search. δ takes values from a grid bounded between 0 and 0.5.

Figure 2.4: Model simulation

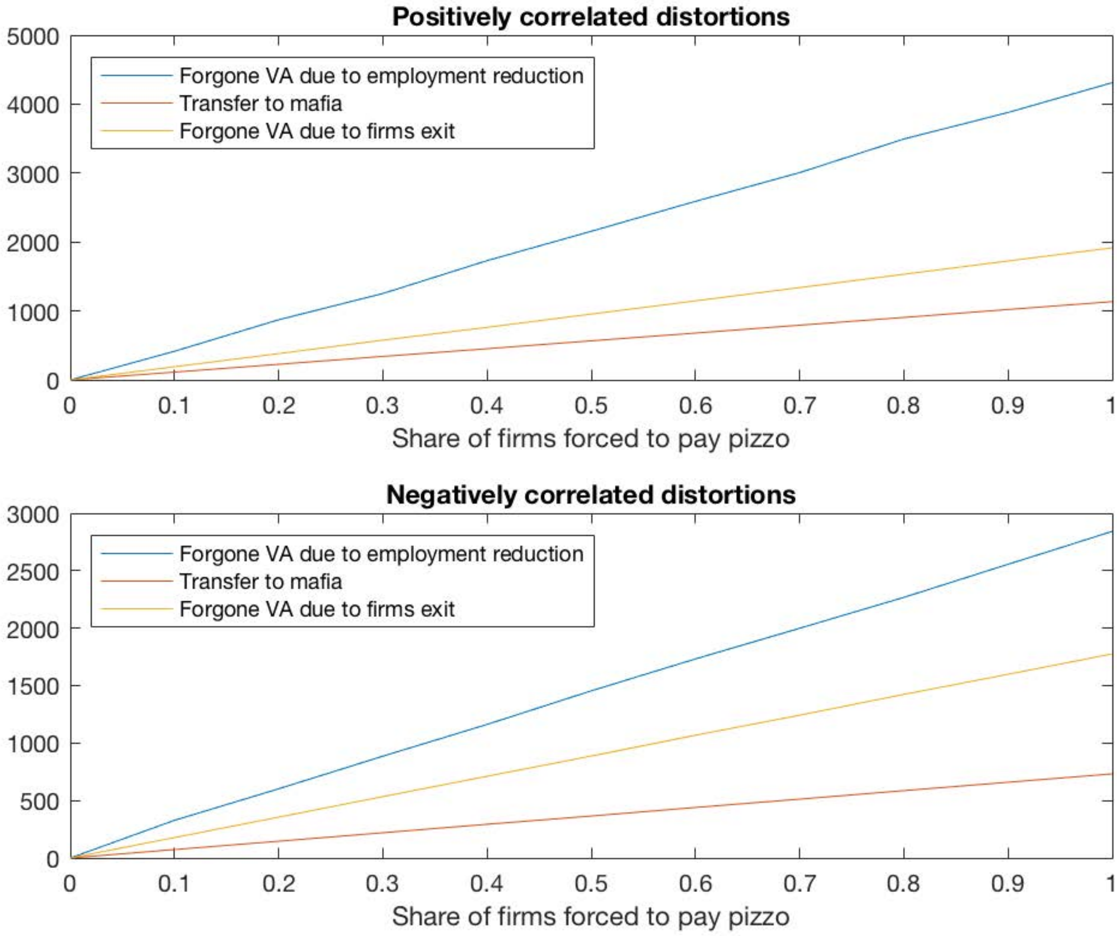


Table 3.6: Loss of value added due to mafia (in million euros)

Sector	Forgone value added	Transfer	Total cost
	Positively correlated distortions		
Construction	515.74	175.35	691.09
Wholesale and retail trade	521.32	172.03	693.35
Accommodation and food service activities	991.56	277.64	1,269.19
Services	418.58	104.65	523.23
Community social and personal services	630.23	220.58	850.81
Total	3,077.47	950.25	4,027.66
	Negatively correlated distortions		
Construction	242.64	67.94	310.57
Wholesale and retail trade	241.48	62.79	304.27
Accommodation and food service activities	456.90	123.36	580.26
Services	195.90	52.89	248.80
Community social and personal services	295.73	82.81	378.54
Total	1,432.73	389.79	1,822.52

Results are obtained using data from the *Small and Medium Entrepreneurs Survey* available at ADELE laboratory-ISTAT, data from the yearly Book of Criminal Statistics published by ISTAT and information provided by the National Agency for the Administration and Management of Real Estate and Firms Confiscated from Criminal Organisation (ANBSC). Column "Forgone value added" is computed by aggregating the forgone value added due to employment reduction and to firm exit in each sector. Column "Transfer" collects the *pizzo* payed by each impacted firm in each sector. Column "Total cost" reports the sum of "Forgone value added" and "Transfer".

Tables and Figures: Chapter 3

Figure 3.1: Violence per Capita and Growth

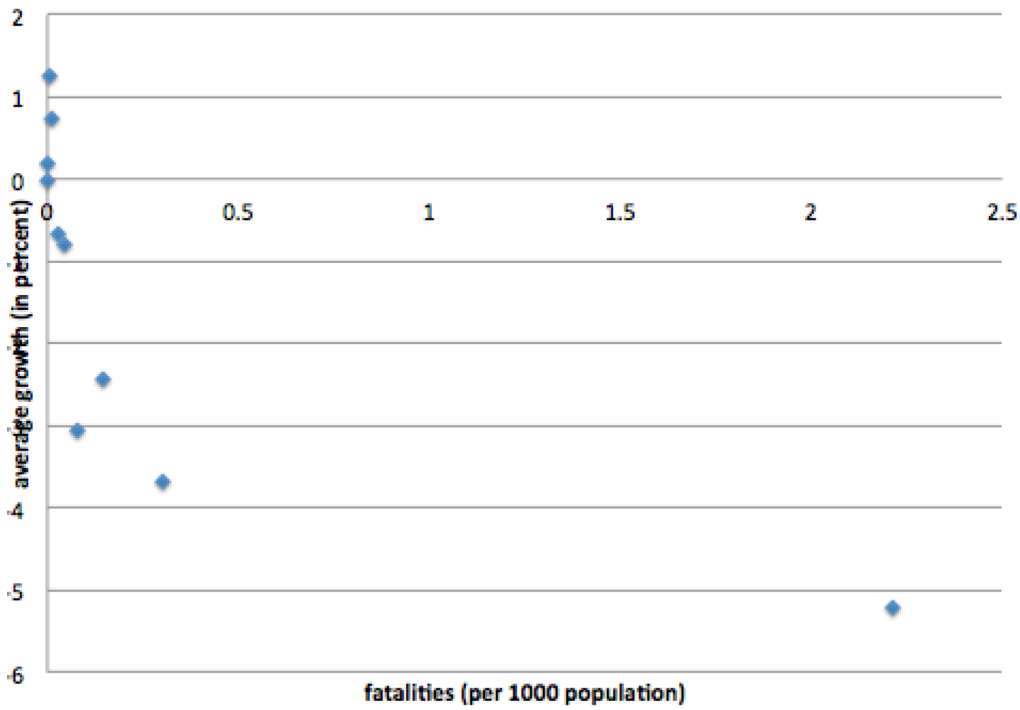


Figure 3.2: Locally Affected Population and Growth (African Countries)

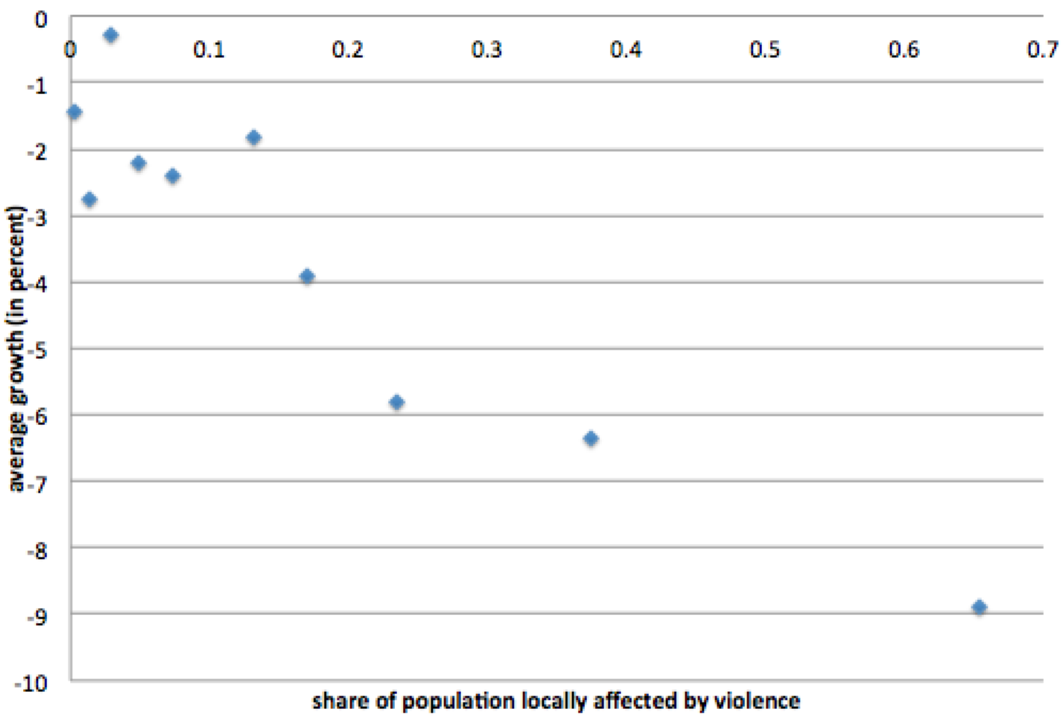


Table 3.7: Conflict Incidence and Economic Performance: Country-level data

Panel A

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.011**	-0.016***	-0.032*
	(0.004)	(0.004)	(0.016)
Observations	8,004	8,076	3,924
R-squared	0.097	0.124	0.267
Country FE	YES	YES	YES
Year FE	YES	YES	YES

Panel B

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.020***	-0.029***	-0.075***
	(0.005)	(0.005)	(0.024)
Observations	8,004	8,076	3,924
R-squared	0.100	0.130	0.268
Country FE	YES	YES	YES
Year FE	YES	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable in columns (1) and (2) is the GDP per capita growth computed using Penn World Table data and World Bank data respectively. Column (3) uses growth of night light per capita form the National Oceanic and Atmospheric Administration (NOAA). Panel A uses as "conflict incidence" a dummy that takes a value of one if in country i at time t the number of battle related deaths is higher than 0. In Panel B "Conflict incidence" dummy takes a value of one if the number of battle deaths is above a threshold of 0.008 fatalities per 1000 population.

Figure 3.3: Estimated output loss in the period 1946-2010 (cross-country analysis)

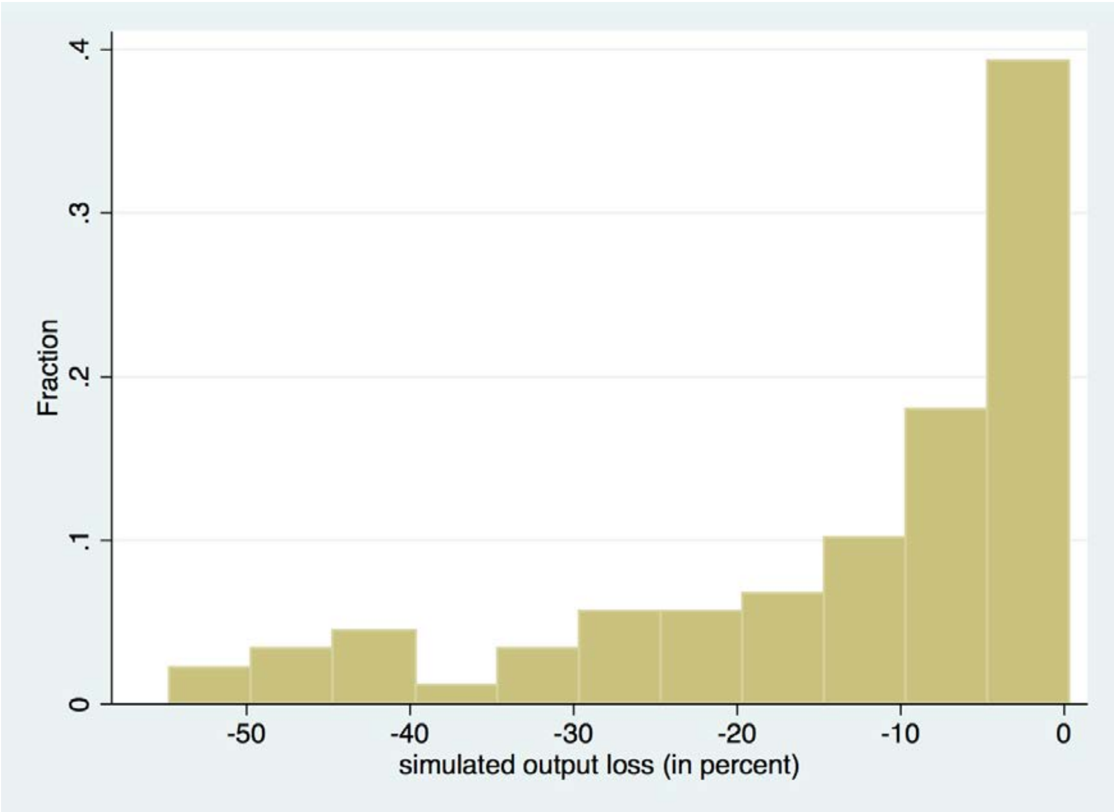


Table 3.8: Conflict Incidence and Country Economic Performance: Ethnic-level data

	(1)	(2)	(3)	(4)	(5)
	light per capita growth	light per capita growth	light per capita growth	light growth	population growth
Conflict Incidence	-0.074**	-0.069**	-0.063**	-0.077***	-0.003**
	(0.029)	(0.034)	(0.031)	(0.029)	(0.001)
N. of groups in conflict in the country			-0.0078*		
			(0.00424)		
Group FE	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES
Lagged growth control	NO	YES	NO	NO	NO
Observations	7,949	7,425	7,949	7,949	7,949
R-squared	0.268	0.473	0.268	0.267	0.045
Number of groups	502	499	502	502	502

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Errors in column (3) are clustered at the country level. We use data over the period 1992-2013. Unit of observation are ethnic groups. “Light per capita growth” and “Population growth” are the log difference of the variable to the previous year. “Conflict incidence” is a dummy defined through more than 25 fatalities in ethnic conflicts. “Number of group in conflict” counts the number of ethnic group involved in conflict episodes in a given country during year t.

Figure 3.4: Estimated Output Loss in the period 1992-2012 (ethnic group level)

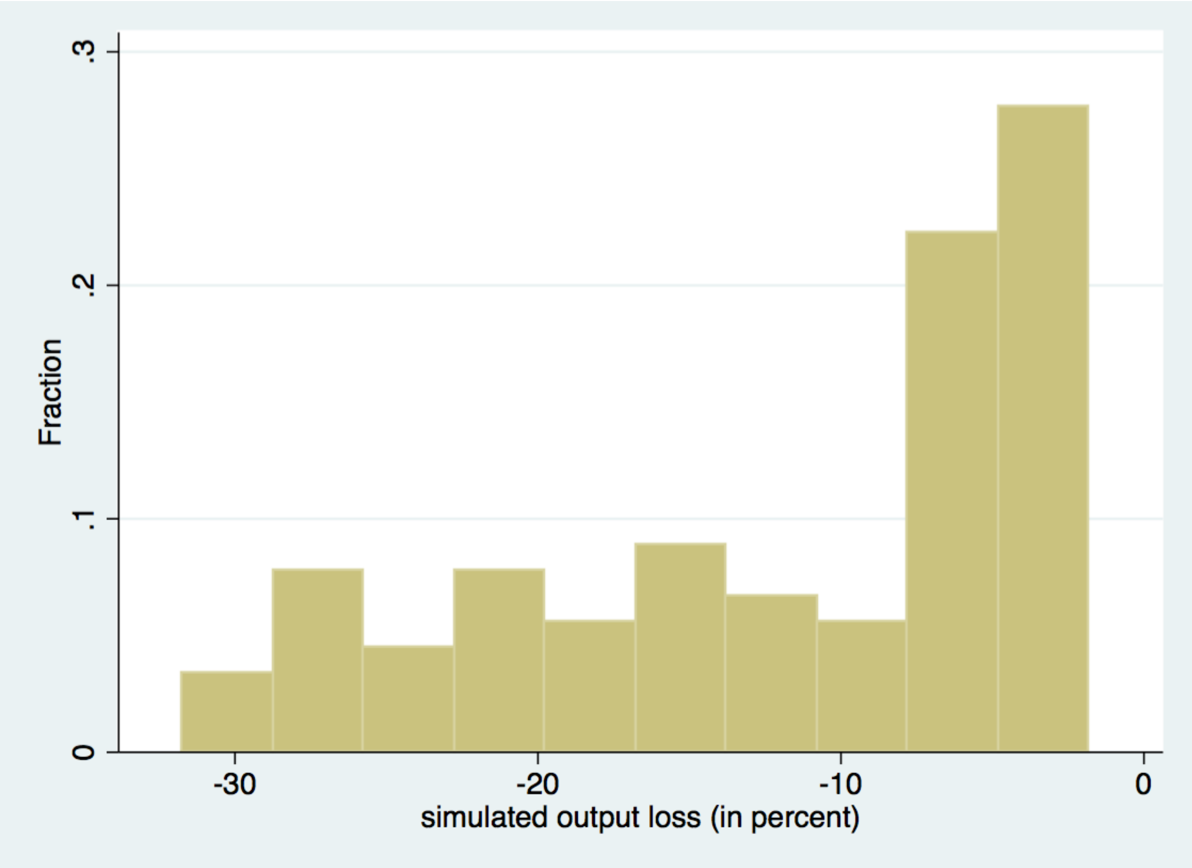


Table 3.9: Impact of Civil Conflict on the Stock of Refugees: Country-level Analysis

VARIABLES	(1) Refugees	(2) Refugees	(3) Refugees	(4) Refugees	(5) Refugees	(6) Refugees
25+ battle deaths	143,789***					
	(11,963)					
1000+ battle deaths		337,498***	487,968***	424,115***	463,596***	466,418***
		(31,004)	(45,835)	(41,811)	(43,736)	(43,790)
25-1000 battle deaths		66,436***	66,873***	57,496***	41,575***	38,586***
		(7,018)	(7,118)	(6,759)	(6,406)	(6,419)
year before war			-5,246			
			(20,679)			
first year of war			-384,851***			
			(52,151)			
last year of war				-259,338***	-256,393***	-255,965***
				(53,001)	(52,639)	(52,618)
1^{st} recovery year				103,277***	156,552***	161,668***
				(27,380)	(28,291)	(28,274)
2^{nd} recovery year					123,404***	128,224***
					(20,514)	(20,493)
3^{rd} recovery year					113,294***	117,929***
					(20,348)	(20,339)
4^{th} recovery year					102,866***	107,279***
					(19,990)	(19,991)
5^{th} recovery year on					76,984***	
					(8,624)	
5+ recovery years						116,326***
						(22,593)
6^{th} recovery year						109,938***
						(23,690)
7^{th} recovery year						109,283***
						(24,808)
8^{th} recovery year						88,871***
						(15,669)
9^{th} recovery year						84,465***
						(16,136)
10+ recovery years						73,528***
						(8,684)
Observations	12,784	12,784	12,763	12,784	12,784	12,784
R-squared	0.435	0.467	0.489	0.478	0.483	0.485
Country FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Country-level data on yearly stocks of refugees are gathered from UNCHR database. In column (1) "25+ battle deaths" is the definition from PRIO dataset of at least 25 battle related deaths in a given year. In column (2)-(6) internal conflict, i.e. variable "1000+ battle deaths", is defined as a year in which violence involves more than 1000 fatalities. Variable "25-1000 battle deaths" controls for years in which fightings cause a number of deaths between 25 and 1000. "Year before war" is a dummy that controls for years with no violence or violence that involves between 25 and 1000 fatalities, followed by a year with at least 1000 battle related deaths. Variables "first year of war" and "last year of war" are dummies that take a value of one during the first and the last year of civil war respectively. Dummies "1^{st} recovery year", "2^{nd} recovery year", etc. take a value of 1 in the x^{th} year after the war ends.

Appendix: Chapter 1

A.1 Proofs

This appendix contains the proof of Proposition 1 introduced in Section 1.2.

Proposition 1 *Consider the vector ω containing N firm-level productivity ranked as follows $\omega_1 > \omega_2 > \omega_3 > ... > \omega_N$. Consider the vector θ containing N of firm-level size ranked as follows $\theta_I > \theta_{II} > \theta_{III} > ... > \theta_N$.*
If aggregate productivity is defined as the sum of firm-level productivity weighted by firm-level size, the way of maximising it is to have the ranking of firm-level productivity and firm-level size perfectly aligned. In other words, aggregate productivity is maximised when ω_1 is matched to θ_I, ω_2 is matched to θ_{II}, and so forth. Consider two elements of ω, $\omega_1 > \omega_2$, and two elements of θ, $\theta_I > \theta_{II}$. To prove Proposition 1, Equation A.1 must be true:

$$\omega_1\theta_I + \omega_2\theta_{II} > \omega_1\theta_{II} + \omega_2\theta_I \tag{A.1}$$

Given that $\omega_1 > \omega_2$, ω_1 can be written as follows: $\omega_1 = \omega_2 + \Delta\omega$ (with $\Delta\omega > 0$). For the same reasoning $\theta_I = \theta_{II} + \Delta\theta$ (with $\Delta\theta > 0$). Plugging these expressions into Equation A.1 and rearranging we obtain:

$$(\omega_2 + \Delta\omega)(\theta_{II} + \Delta\theta) + \omega_2\theta_{II} > (\omega_2 + \Delta\omega)\theta_{II} + \omega_2(\theta_I I + \Delta\theta)$$

$$\omega_2\theta_{II} + \Delta\omega\Delta\theta + \Delta\omega\theta_{II} + \Delta\theta\omega_2 + \omega_2\theta_{II} > \omega_2\theta_{II} + \Delta\omega\theta_{II} + \omega_2\theta_{II}\Delta\theta\omega_2$$

That leads to:

$$\Delta\omega\Delta\theta > 0$$

which is always true.
Pick two other elements from vector ω and vector θ, such that $\omega_A > \omega_B$ and $\theta_C > \theta_D$.

Given the computations provided above, we know that:

$$\omega_A\theta_C + \omega_B\theta_D > \omega_A\theta_D + \omega_B\theta_C \tag{A.2}$$

therefore:

$$\omega_1\theta_I + \omega_2\theta_{II} + \omega_A\theta_C + \omega_B\theta_D > \omega_1\theta_{II} + \omega_2\theta_I + \omega_A\theta_D + \omega_B\theta_C \tag{A.3}$$

We can apply this reasoning to every element of vector ω and vector θ.

A.2 Robustness checks

This appendix introduces some robustness checks of the motivated evidence presented in Section 1.4, where I show that mafia *presence* correlates negatively with allocative efficiency. The main point here is to show that mafia *intensity* correlates negatively with allocative efficiency. To show this point I implement the following IV strategy. Equation A.4 describes the core specification of the model.

$$\text{OP covariance}_{spt} = \alpha + \beta \text{Mafia intensity}_{spt} + \eta_p + \theta_{st} + u_{spt} \tag{A.4}$$

The outcome variable is the OP covariance between size and productivity computed for every sector s province p and year t. Mafia intensity is measured through the interaction between "extortion cases per 1000 firms", reported in each province p and year t, and a dummy variable that equals one if sector s is classified as mafia-appealing. Mafia intensity is instrumented with a dummy that takes a value of one if the market is defined as mafia-infiltrated, according to the definition introduced in Section **??**.[1] The core specification of the model includes province fixed effects η_p and sector-year fixed effects θ_{st}. In order to control for a higher degree of unobserved heterogeneity, two additional specifications include sector-year and province-year fixed effects and sector-year, province-year and province-sector fixed effects. The three model specifications exclude the construction sector in order to avoid the bias introduced by the potential violation of the exclusion restriction assumption. In fact, the approval of renewal works for public infrastructure is likely to have an impact per se on the allocative efficiency of the construction sector.

Table A1 shows the results. Column (1) presents the results of the main specification of the model. The coefficient of mafia intensity suggests that an increase of mafia intensity correlates negatively with the allocative efficiency of mafia-appealing sectors located in the provinces where the mafia succeeded in infiltrating, after mafia arrival. Results reported in Column (2) support this idea. In Column (3), the coefficient on mafia intensity is negative, but it loses significance. This might be due to the quite limited variation brought by the high number of fixed effects.

These results suggest that the extortive behaviour of mafia groups, which increased significantly in some northern provinces because of investment in public infrastructure, negatively biased the OP covariance between productivity and size characterising the mafia-infiltrated markets. This outcome is line with the suggestive evidence presented in Section 1.4, where I show a decreasing trend of the OP covariance in the mafia-appealing

[1] Sector s located in province p observed at time t is mafia-infiltrated if, in province p and year t, public works for the A4 highway or the high speed rail have been approved and sector s is mafia-appealing.

sectors located in mafia-infiltrated provinces after the arrival of the mafia.

Table A1: Mafia intensity and allocative efficiency

	(1) OP covariance	(2) OP covariance	(3) OP covariance
Mafia Intensity	-0.091**	-0.099***	-0.128
	(0.036)	(0.032)	(0.231)
Observations	7,359	7,359	7,359
R-squared	0.138	0.202	0.359
Province FE	YES	NO	NO
Year FE	NO	NO	NO
Sector-Year FE	YES	YES	YES
Province-Year FE	NO	YES	YES
Province-Sector FE	NO	NO	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable in columns (1)-(3) is OP covariance for log-labour productivity and share of employment computed using data from the *Small and Medium Entrepreneurs Survey* available at ADELE laboratory-ISTAT. Variable "Mafia intensity" is obtained interacting "extortion cases per 1000 firms" and "mafia-appealing sector" dummy, i.e. a dummy that takes a value of one if sector s is labelled as mafia appealing. Data on reports of extortion is provided by the Yearly Book of Criminal Statistics published by ISTAT. "Mafia intensity" is instrumented by a dummy that takes a value of one if sector s located in province p observed at time t is defined as mafia-infiltrated.

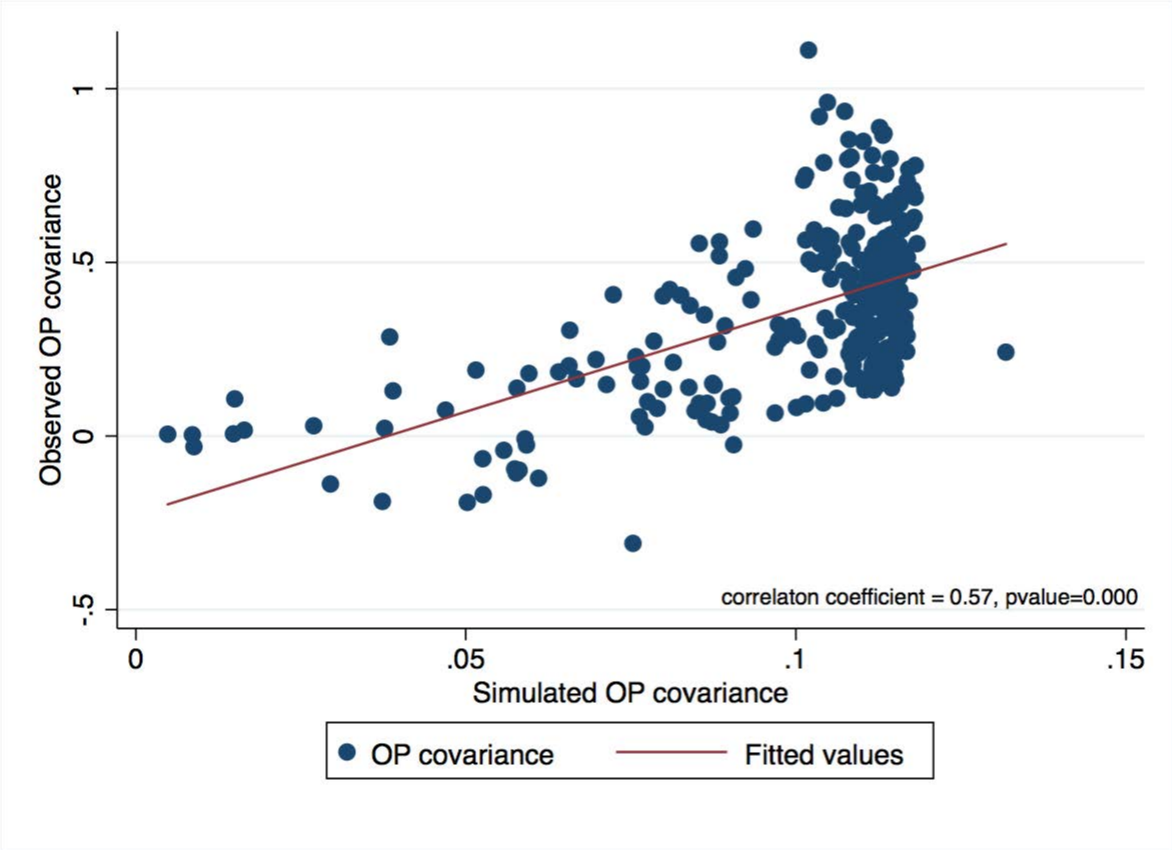
Appendix: Chapter 2

B.1 Fit of the model

A brief comment on the performance of the model is worthwhile. Figure B1 provides some insights about the relationship between observed OP covariance of each mafia-infiltrated market and its simulated counterpart. The correlation between observed and simulated OP covariance is equal to 0.57. However, it can be noticed that the two variables have highly different scales. In particular, the simulated OP covariance is systematically lower than the observed one. This might be due to the presence in the actual data of few outliers in terms of productivity that scale up the observed OP covariance. These outliers cannot be simulated because of the distributional assumptions imposed in the model. This suspicion is corroborated by the comparison of the observed correlation between productivity and size to its simulated counterpart. The computation of the correlation coefficient requires the variance of firms' productivity at the denominator, the presence of the outliers would increase the denominator and lead to a lower correlation coefficient. Conversely, given that simulated firm-level data does not produce these outliers, the simulated correlation coefficient is not scaled down. This is what happens in the present context: the observed correlation between size and coefficient is systematically lower than the simulated one.[1]

[1]Graphs that shows the difference between observed and simulated correlation between firm productivity and size are available upon request.

Figure B1: OP covariance



Appendix: Chapter 3

C.1 Robustness Check on Conflict Incidence and Economic Performance

In this appendix we show several robustness checks of the models described in equation (1). In Table C1 we add to the model described in equation (1) the first lag of GDP/night light per capita growth rate and results are robust to this.[1] In order to exclude the hypothesis that the measured impact is driven by time varying characteristics of the country-level economy rather than conflict itself, we ran specifications that include country-specific time trends. Table C2 reports the results. Again, the estimated coefficients of the "conflict incidence" variable are very close to the ones reported in Table 1.

Tables C3 and C4 report results obtained from a subsample of countries, i.e. data used in Miguel & Satyanath (2011) . Table C3 Panel A shows results of the regression model of equation (1), while Panel B includes the rainfall shocks variables used in Miguel & Satyanath (2011). Specifications reported in Table C4 Panel A and B include country-specific time trends. Again, results are completely robust. This suggests that when controlling for the most common instrument for GDP shocks directly nothing happens to the correlation between violence and GDP growth. This to us indicates that our results stem from a causal chain that must run the other direction.

[1] However, this specification is not correct, strictly speaking, as it is affected by the so-called "Nickel bias". From Nickell (1981) we learn that the demeaning process implemented in fixed effect regressions creates a correlation between regressors and error.

Table C1: Controlling for Lagged Economic Performance

Panel A

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.009**	-0.014***	-0.047***
	(-0.004)	(-0.004)	(-0.015)
Lagged growth	0.139***	0.193***	-0.148**
	(-0.036)	(-0.048)	(-0.075)
Observations	7,887	7,884	3,737
R-squared	0.117	0.158	0.290
Country FE	YES	YES	YES
Year FE	YES	YES	YES

Panel B

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.017***	-0.025***	-0.067***
	(-0.005)	(-0.005)	(-0.018)
Lagged growth	0.137***	0.190***	-0.147*
	(-0.036)	(-0.048)	(-0.075)
Observations	7,887	7,884	3,737
R-squared	0.118	0.162	0.291
Country FE	YES	YES	YES
Year FE	YES	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable in columns (1) and (2) is the GDP per capita growth computed using Penn World Table data and World Bank data respectively. Column (3) uses growth of night light per capita form the National Oceanic and Atmospheric Administration (NOAA). Variable "Lagged growth" controls for one period lagged GDP per capita growth in columns (1) and (2), and one period lagged night light per capita growth in column (3). Panel A uses as "conflict incidence" a dummy that takes a value of one if in country i at time t the number of battle related deaths is higher than 0. In Panel B "Conflict incidence" dummy takes a value of one if the number of battle deaths is above a threshold of 0.008 fatalities per 1000 population.

Table C2: Controlling for Country-specific Time Trend

Panel A

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.010**	-0.015***	-0.034**
	(-0.004)	(-0.004)	(-0.015)
Observations	8,004	8,076	3,924
R-squared	0.116	0.151	0.307
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-time trend	YES	YES	YES

Panel B

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.020***	-0.028***	-0.078***
	(-0.005)	(-0.005)	(-0.027)
Observations	8,004	8,076	3,924
R-squared	0.119	0.156	0.309
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-time trend	YES	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable in columns (1) and (2) is the GDP per capita growth computed using Penn World Table data and World Bank data respectively. Column (3) uses growth of night light per capita form the National Oceanic and Atmospheric Administration (NOAA). Panel A uses as "conflict incidence" a dummy that takes a value of one if in country i at time t the number of battle related deaths is higher than 0. In Panel B "Conflict incidence" dummy takes a value of one if the number of battle deaths is above a threshold of 0.008 fatalities per 1000 population.

Table C3: Using Miguel & Satyanath (2011) Data

Panel A

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.039**	-0.034***	-0.145**
	(-0.01)	(-0.01)	(-0.048)
Observations	1,189	1,128	697
R-squared	0.120	0.153	0.336
Country FE	YES	YES	YES
Year FE	YES	YES	YES

Panel B

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.039***	-0.034***	-0.145**
	(-0.011)	(-0.008)	(-0.056)
Rainfall (t)	0.027**	0.028***	-0.02
	(-0.012)	(-0.010)	(-0.051)
Rainfall (t-1)	-0.026*	-0.012	-0.008
	(-0.015)	(-0.012)	(-0.067)
Rainfall (t-2)	-0.015	-0.024**	-0.165
	(-0.014)	(-0.011)	(-0.107)
Observations	1,073	1,014	592
R-squared	0.141	0.224	0.346
Country FE	YES	YES	YES
Year FE	YES	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable in columns (1) and (2) is the GDP per capita growth computed using Penn World Table data and World Bank data respectively. Column (3) uses growth of night light per capita form the National Oceanic and Atmospheric Administration (NOAA). Variable "conflict incidence" is a dummy that takes a value of one if the number of battle deaths is above a threshold of 0.008 fatalities per 1000 population. Panel B adds rainfall variation variables: variable "rainfall(t)" measures the log of rain precipitation in country i at time t, variables "rainfall(t-1)" and "rainfall(t-2)" are its first and second lag respectively.

Table C4: Using Miguel & Satyanath (2011) Subsample and Country-specific Time Trends

Panel A

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.038***	-0.031***	-0.153***
	(-0.01)	(-0.01)	(-0.056)
Observations	1,189	1,128	697
R-squared	0.147	0.186	0.386
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-time trend	YES	YES	YES

Panel B

	(1) GDP growth	(2) GDP growth	(3) Night light growth
Conflict Incidence	-0.038***	-0.031***	-0.160**
	(-0.011)	(-0.008)	(-0.069)
Rainfall (t)	0.032**	0.039***	-0.006
	(-0.013)	(-0.011)	(-0.057)
Rainfall (t-1)	-0.027*	-0.011	-0.038
	(-0.015)	(-0.012)	(-0.094)
Rainfall (t-2)	-0.014	-0.026**	-0.172
	(-0.014)	(-0.012)	(-0.114)
Observations	1,073	1,014	592
R-squared	0.174	0.261	0.399
Country FE	YES	YES	YES
Year FE	YES	YES	YES
Country-time trend	YES	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The dependent variable in columns (1) and (2) is the GDP per capita growth computed using Penn World Table data and World Bank data respectively. Column (3) uses growth of night light per capita form the National Oceanic and Atmospheric Administration (NOAA). Variable "conflict incidence" is a dummy that takes a value of one if the number of battle deaths is above a threshold of 0.008 fatalities per 1000 population. Panel B adds rainfall variation variables: variable "rainfall(t)" measures the log of rain precipitation in country i at time t, variables "rainfall(t-1)" and "rainfall(t-2)" are its first and second lag respectively.

C.2 Robustness Check on Internal Conflicts and Refugee Flows

In this appendix, we consider famine episodes as alternative source of refugee movements and we check robustness of the estimates in table 3.9 when accounting for it. In the data, we indeed observe a strong correlation between presence of conflict and famine. Descriptive statistics shown in table C5 suggest that we have famine episodes 0.4% of all country-year observations without conflict versus 4.2% for country-year observations with conflict (10 times higher).

This difference also holds within countries. When regressing famine incidence on conflict controlling for country fixed effects, the coefficient obtained is positive and significant. This means that even within the same country, famine episodes are more likely to occur during conflict years and this may lead to substantial bias in estimates shown in table 3.9.

However, even after controlling for famine incidence, table C6 shows that these estimates do not change substantially suggesting that the direct effect of famine on refugee flows is orthogonal to the effect of conflicts on refugee flows.

Table C5: Descriptive Statistics: Conflicts and Famine Episodes

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Famine incidence without conflict	11470	0.004	0.061	0	1
Famine incidence with conflict	1314	0.042	0.200	0	1

Table C6: Civil Conflict and Refugee Stock: Robustness to Famine Episodes

VARIABLES	(1) Refugees	(2) Refugees	(3) Refugees	(4) Refugees	(5) Refugees	(6) Refugees
25+ battle deaths	142,234***					
	(11,994)					
1000+ battle deaths		335,925***	487,063***	422,675***	462,213***	465,089***
		(31,232)	(46,288)	(42,036)	(43,964)	(44,026)
25-1000 battle deaths		66,257***	66,828***	57,385***	41,491***	38,514***
		(7,049)	(7,135)	(6,775)	(6,414)	(6,425)
Incidence of Famine Event	66,109*	27,663	12,274	26,516	25,411	24,149
	(35,806)	(34,261)	(35,128)	(34,463)	(34,155)	(34,123)
year before war			-5,904			
			(20,702)			
first year of war			-384,260***			
			(52,456)			
last year of war				-259,647***	-256,688***	-256,248***
				(52,898)	(52,540)	(52,521)
1st recovery year				102,460***	155,753***	160,886***
				(27,452)	(28,373)	(28,363)
2nd recovery year					123,301***	128,104***
					(20,554)	(20,534)
3rd recovery year					112,489***	117,143***
					(20,410)	(20,407)
4th recovery year					103,265***	107,637***
					(19,980)	(19,979)
5th recovery year on					77,023***	
					(8,617)	
5 + recovery years						116,015***
						(22,658)
6th recovery year						109,942***
						(23,734)
7th recovery year						108,840***
						(24,807)
8th recovery year						89,189***
						(15,645)
9th recovery year						84,825***
						(16,099)
10 + recovery years						73,544***
						(8,679)
Observations	12,784	12,784	12,763	12,784	12,784	12,784
R-squared	0.436	0.467	0.489	0.478	0.484	0.485
Country and Year FE	YES	YES	YES	YES	YES	YES

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Country-level data on yearly stocks of refugees are gathered from UNCHR database. In column (1) "25+ battle deaths" is the definition from PRIO dataset of at least 25 battle related deaths in a given year. In column (2)-(6) internal conflict, i.e. variable "1000+ battle deaths", is defined as a year in which violence involves more than 1000 fatalities. Variable "25-1000 battle deaths" controls for years in which fightings cause a number of deaths between 25 and 1000. "Year before war" is a dummy that controls for years with no violence or violence that involves between 25 and 1000 fatalities, followed by a year with at least 1000 battle related deaths. Variables "first year of war" and "last year of war" are dummies that take a value of one during the first and the last year of civil war respectively. Dummies "1^{st} recovery year", "2^{nd} recovery year", etc. take a value of 1 in the x^{th} year after the war ends.

C.3 Data

C.3.1 Uppsala Conflict Data Program Battle-Related Dataset

The Uppsala Conflict Data Program (UCDP) collects information on a large number of aspects of armed conflicts occurred since 1946. The UCDP provides several datasets that allow to explore different features of armed conflicts. In this report we make a large use of the UCDP Battle-Related Dataset. This dataset provides yearly information on country-level number of fatalities related to combat, the time period covered is 1946-2014.

The data on battle-related deaths are collected through the use of news sources. All reports which contain information about individuals killed or injured in fightings are gathered and coded manually into an event-year level dataset. For every event, several details are recorded and translated into variables: the date and location of the event, the reporting source, the primary source, the actors involved, what happened, and three estimates of fatalities caused by the event (low, high, and best estimate).

For the purpose of the present report, we focus on three variables provided by the UCDP Battle-Related Dataset that we describe below. These variables describe the *type* of the violent event, its *location* and the estimate of the *number of fatalities*.

Type of conflict

The UCDP/PRIO Armed Conflict Dataset identifies four different types of conflict. For each conflict event coded in the dataset we can distinguish among:

1. Extra-systemic conflict: occurs when a government of a state is fighting to retain the control of a territory outside the state system.

2. Interstate conflict: occurs between two or more different states.

3. Internal conflict: occurs when the government of a state fight against one or more internal groups. In this type of conflict there is no intervention from other states.

4. Internationalized internal conflict: occurs when the government of a state fight against one or more internal groups and external states intervene to support one or both sides.

Given that in this report we analyze civil conflicts, we only focus on two types of conflict reported in this dataset, namely internal conflict and internationalized internal conflict, i.e. conflict of types 3 and 4.

Location

In order to assign a location to conflict events, we make use of the variable *locationinc*, that reports "the name of the country/countries whose government(s) has a primary claim to the object in dispute". In case that more than one country name is included, the location is assigned to the country where the conflict is fought. If the territory over which the conflict is fought covers more than one country (e.g. the fightings take place on the border between two countries), the same estimate is assigned to each country that is covered by the territory.

Number of battle-related deaths

The UCDP Battle-related deaths Dataset provides three estimates of the number of fatalities that each violent event implies. These variables are:

- *bdlow*: this variable provides the low estimate of the occurred battle related deaths for each conflict event and year. This estimate is the results of the aggregation of low estimates for all the fatalities related to battle-related incidents.

- *bdhigh*: this variable results from the aggregated high estimates for all battle-related incidents in a given conflict event and year.

- *bdbest*: the estimate consist of the aggregated most reliable numbers for all battle-related fatalities in a given conflict event and year. If different reports provide different estimates, the estimate provided by the most reliable source is provided. If no such distinction can be made, the lowest among these numbers is used.

In this report we mainly use variable *bdbest*.

Recall that in the analyses where we use this data the unit of observation is country-year. Since UCDP Battle-related deaths Dataset provides information at conflict event-year level, we aggregate *bdlow*, *bdhigh* and *bdbest* by summing them for each country year.

To sum up, we make a three step re-coding of the dataset. We first exclude the observations that are not related to internal conflicts. We then locate each observation to one (or more) country. Finally we sum the estimates battle related deaths by country-year. We end up with a panel that reports three estimates of battle related deaths.

C.3.2 Geographical Research On War, Unified Platform ($Grow^{up}$) Dataset

$Grow^{up}$ federated data platform provides access to disaggregated, integrated and spatially explicit conflict related data. It offers research-ready data on ethnic groups and intrastate conflict compiled from various sources and provided in group-year and country-year format.

The sample universe of ethnic groups in the RFE group-level data is adopted from the EPR (Ethnic Power Relations) Core dataset (Cederman et al., 2010). It covers all countries between 1946 and 2013 except failed states, overseas colonies and countries with less than 500'000 inhabitants. Newly independent states are included in the dataset beginning with the year of independence. From this country-year list, EPR Core dataset defines ethnicity as "any subjectively experienced sense of commonality based on the belief in common ancestry and shared culture". Only politically relevant ethnic groups are included in the dataset. An ethnic group is classified as relevant if "at least one political organization claims to represent it in national politics or if its members are subjected to state-led political discrimination". This yields 817 politically relevant ethnic groups in 141 countries in the EPR Core dataset.

Ethnic Conflicts

The information on ethnic conflicts in the RFE group-level data is compiled from two different sources: The ACD2EPR dataset (**?**), the UCDP Actor Dataset (Uppsala Conflict Data Program 2014), and the Uppsala/PRIO Armed Conflict Database (ACD) (**?**). The conflict data focuses on ethnic civil wars (Internal, and Interna- tionalized Internal Conflicts). For all conflict onset and incidence variables in the RFE group- level dataset, a conflict episode is only considered terminated if there is no conflict-related activity in the following two calender years where "Conflict-related activity" refers to the UCDP threshold of at least 25 battle deaths per annum.

Geographical Data

Geographical data (night light intensity and population) at ethnic group level are raster derived data, created by overlaying the GeoEPR 2014 settlement polygons with geospatial raster datasets. Night light intensity data is taken from DMSP-OLS [2] Nighttime Lights Time Series (Average Visible, Stable Lights, and Cloud Free Coverages). All nightlights within group polygon are aggregated.

[2] Defense Meteorological Satellite Program

C.3.3 Polity IV Dataset

The Polity datasets aim at coding authority characteristics of states in the world system for purposes of quantitative analysis. The Polity IV dataset covers all major (around 167 countries), independent states in the global system over the period 1800-2014.

Location of the data

This dataset covers a very extended time period, thus it provides information about countries that do not exist anymore. In particular, in this report we deal with three countries (confederation of countries) that separated in the early 90's, namely Soviet Union, Yugoslavia and Czechoslovakia. For each country belonging to these confederation, we do not have distinct information until year 1990-1991. For example, until year 1990, we have observations for Czechoslovakia, but not for Czech Republic or Slovakia. We deal with this issue in a very simple way: we impute to each member state the observations made on the confederation, until the last year the confederation existed. Hence, until year 1991, we assign to Czech Republic and Slovakia observations made on Czechoslovakia. We apply the same procedure for each member state of Soviet Union and Yugoslavia. Hence, until year 1990-1991 observation made on Soviet Union will be assigned to Armenia, Azerbaijan, Belarus, Estonia, Georgia, Kazakstan, Kyrgyzstan, Latvia, Lithuania, Republic of Moldova, Russia, Tajikistan, Ukraine, Uzbekistan and information on Yugoslavia will be imputed to Croatia, Macedonia, Serbia and Montenegro and Slovenia.

Variable *xconst*

In this report we use variable *xconst* to build our measure of strong executive constraint as in Besley & Mueller (2015). This variable is on a seven point scale and the manual explains its construction as follows: "Operationally, this variable refers to the extent of institutionalized constraints on the decision making powers of chief executives, whether individuals or collectivities. Such limitations may be imposed by any "accountability groups". In Western democracies these are usually legislatures. Other kinds of accountability groups are the ruling party in a one-party state; councils of nobles or powerful advisors in monarchies; the military in coup-prone polities; and in many states a strong, independent judiciary. The concern is therefore with the checks and balances between the various parts of the decision-making process." [p. 24, Polity IV Dataset Users' Manual 2010]

We create a dummy that takes value 1 when *xconst* equal 7 following Besley & Mueller (2015). The highest score of the variable *xconst* is only allocated if important legislation can be initiated by a parliament which holds the executive to account.

C.3.4 Refugees data

We exploit country-level data gathered from several sources. Data about refugees is provided by the UNHCR Population Statistics Database. The database provides information about UNHCR's populations of concern from the year 1951 up to 2014. This database lists seven categories: refugees, asylum-seekers, returned refugees, internally displaced persons (IDPs), returned IDPs, stateless persons and others of concern. For each group the database provides yearly information about their composition by location of residence and origin. We exploit only the data on refugees.

According to the UNHCR definition, refugees are "individuals recognized under the 1951 Convention relating to the Status of Refugees; its 1967 Protocol; the 1969 OAU Convention Governing the Specific Aspects of Refugee Problems in Africa; those recognized in accordance with the UNHCR Statute; individuals granted complementary forms of protection; or those enjoying temporary protection; and people in a refugee-like situation".

In particular, we are interested in the annual stock of refugees for each country of origin, i.e. how many people with refugees status have left their home country each year. Hence, for each country of origin we sum the stock of refugees reported in each host country. The resulting number will be the variable *refugees*.