UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche “Marco Fanno”

CRIME AND SOCIAL SANCTION

PAOLO BUONANNO
Università di Bergamo

GIACOMO PASINI
University of Utrecht - Università di Venezia

PAOLO VANIN
Università di Padova

March 2008

“MARCO FANNO” WORKING PAPER N.71

Crime and Social Sanction

Paolo Buonanno*
University of Bergamo

Giacomo Pasini†
University of Utrecht
University of Venice

Paolo Vanin‡
University of Padua

March 12, 2008

Abstract

Social sanctions may be a strong deterrent of crime. This paper presents a formal model that relates crime and social sanction to social interaction density. We empirically test the theoretical predictions using a provincial level panel dataset on different crimes in Italy between 1996 and 2003. We exploit detailed demographic and geo-morphological information to develop exogenous measures of social interaction density. We estimate a spatial panel model by means of a GMM procedure and we find that provinces with denser social interactions display significantly and substantially lower rates of property crime, but not of violent crime.

JEL Classification: A14, K42, Z13.
Keywords: Crime, Spatial Panel, Social Interaction, Social Sanction.

*Department of Economics, University of Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy. E-mail: paolo.buonanno@unibg.it.

†Utrecht School of Economics, Janskerkhof 12, 3512 BL Utrecht, Netherlands. E-mail: g.pasini@econ.uu.nl

‡Department of Economics, University of Padua, Via del Santo 33, 35123 Padova, Italy. E-mail: paolo.vanin@unipd.it

1 Introduction

Starting from Becker (1968), crime economics has always emphasised the deterrent effect of expected punishment. Economists have devoted most attention to formal, legal sanctions (see, e.g., Polinsky and Shavell, 2000). They have devoted much less attention to informal, social sanctions, which might range from simple disapproval to strong forms of socio-economic ostracism or retaliation.[1]

Among the few recent contributions tackling this issue, Sherman et al. (1992) show evidence that in some cases legal punishment exerts a deterrence effect only for those who are also exposed to social sanction; Rasmusen (1996) and Funk (2004) observe that formal conviction may lead to stigmatisation, in the sense that other people are reluctant to interact with a convicted criminal (yet it may be noticed that social sanctions need not follow public conviction, since they are based on information transmission through social networks and need not rely on formal proofs); Weibull and Villa (2005) argue that the effectiveness of social norms against crime, due to both guilt and shame, is decreasing in crime rates; and Falk et al. (2005) provide experimental evidence supporting the idea that retaliation may be the most important motive behind certain forms of informal sanctions.

Most important for the present investigation, recent work emphasises that denser social networks might raise aggregate crime levels. This may be due to different mechanisms, like know-how sharing among criminals (Calvó-Armengol and Zenou, 2004) or imitation of peer behaviour (Glaeser et al., 1996; Patacchini and Zenou, 2005).[2] Yet the density of social interaction might also work as a deterrent for criminal activity. The reason is that I might want to refrain from damaging other people if I am very likely to interact in the future either with them or with people they know, for the fear that my future contacts may retaliate. A similar argument, relating the strength of social sanction to social network closure (the property of social networks of being interconnected) is already spelled out by Coleman (1988). The idea that social sanction works better in small communities is also emphasised by Posner (1997).[3]

[1]This may be surprising, given the fact that the relationship between legal and informal social control is an old theme in social theory. Prominent historical voices include the eighteenth-century Italian economist and criminologist Cesare Beccaria, who provided important insights on the role of state-imposed punishments, and the nineteenth-century French sociologist Émile Durkheim, who emphasised the role of shame and social sanction in deterring crime.

[2]See also Haynie (2001) and Silverman (2004).

[3]As already argued by Elster (1989), whereas the threat of social sanctions may make it individually rational to abstain from crime, a coherent theory should also clarify why other individuals are willing to impose social sanctions to those who deviate. In small communities, repeated interaction is likely to provide the adequate incentives.

We present a simple model formalising this argument (we speak indifferently of network closure and of social interaction density) and then use it to guide our empirical investigation, which is based on provincial level panel data on different crimes in Italy between 1996 and 2003. According to the theoretical model, social sanctions are better able to enforce honesty where social interaction density is higher. Recent empirical work by Allcott et al. (2007) confirms that network closure decreases in community size and that a substantial part of the negative correlation between community size and pro-social behaviour is mediated through the closure channel. Building on this evidence, we exploit detailed demographic and geo-morphological information to develop an exogenous and reliable measure of small and isolated communities, characterised by denser social interaction, and then use it as a proxy for the strength of social sanction.

One advantage of our approach is that we explicitly take spatial correlation into account. The relevance of spatial correlation has recently been emphasised by Millo and Pasini (2006) and by Patacchini and Zenou (2007). Omitting to account for possible correlation across provinces could yield inconsistent estimates. Spatial panel estimation techniques, first outlined in Elhorst (2003), have not yet become a standard in the economic literature because of computational difficulties. Based on the Generalized Spatial Two Stage Least Squares (GS2SLS) estimator of Kelejian and Prucha (1998), we estimate the model by means of GMM. In particular, our approach allows to explicitly account for endogeneneity of the effect of crime rates in neighbouring provinces.

Controlling for standard determinants of crime and for spatial correlation, we find that our measures for social interaction density have a substantial and significant negative effect on property crime rates, but not on violent crimes. We interpret this as an indication that social sanction is important for property but not for violent crime. Our results are robust to big city controls, which makes us confident that we are indeed capturing the effect of social sanction and not other forms of aggregate increasing returns to crime (Glaeser and Sacerdote, 1999).

Our results are related to a few recent economic contributions on the impact of socio-cultural variables on crime. In particular, Buonanno, Montolio, and Vanin (Buonanno et al.) find that history-predicted association density and altruistic norms, proxied by blood donations, significantly reduce property crime, possibly because associational networks increase returns to non-criminal activities and raise detection probabilities, and altruistic norms attach guilt and shame to criminal behaviour, thus increasing its opportunity cost. In turn, civic norms, proxied by referenda turnout, are not significant.[4] Heaton (2006) finds that most of the negative correlation between religious participation and crime is indeed due to an effect of the latter on the former, and not the other way around. Although related to our investigation, neither of these works focuses on social sanction. Further, to avoid endogeneity, they resort to historical instruments. Here, to the contrary, we exploit demographic and geo-morphological information to develop exogenous measures of social interactions density.[5]

The remainder of the paper is organised as follows. Section 2 presents the theoretical model and its key prediction. Section 3 provides a discussion of the data on crime and its potential determinants, while the following section is dedicated to the definition of an empirical measure for social sanction and social interaction density. Sections 5, 6 and 7 present our empirical methodology, our results and their robustness, respectively. Section 8 concludes.

2 A theoretical model

The main intuition of why crime rates should be lower in communities with denser social interaction is straightforward. Crime may be deterred either by the threat of formal punishments, imposed by the judiciary system, or by the threat of 'informal' punishment, coming through social interaction (e.g., refusal to hire a worker for a job or refusal to spend time together). In both cases detection probability is crucial, but whereas formal conviction requires formal proofs, social sanction only requires that awareness of the committed crime is transmitted to the next interaction partners. The denser the patterns of social interaction, the more likely it is that I will interact tomorrow with an individual who, if I commit crime today, will be aware of it and therefore will be in a position to impart me a social sanction.[6] In one word, in less dense interaction environments individuals discount more heavily the prospective social sanction they may have to incur if they commit a crime. To make this intuition formal, we adapt a model originally developed by Dixit (2003) to explain the extent of honest

[4] In a cross-country study, Lederman et al. (2002) analyse the effect of social capital on violent crime finding that only the component of social capital measured by trust in community members has the effect of reducing the incidence of violent crimes.

[5] Related contributions by sociologists and criminologists on the role of social capital and civic engagement for crime also include Rosenfeld et al. (2001), who find a negative impact of social capital on homicide rates; Messner et al. (2004), who find that homicide rates decrease in social trust, but, surprisingly, increase in community and political activism, and Chamlin and Cochran (1997), who find that social altruism, proxied by charitable donations, has a negative and significant impact on both property and violent crime in a sample of U.S. cities.

[6] Social sanction takes here the form of refusal of a mutually advantageous interaction opportunity. Individuals are willing to pay the cost of sanctioning others for fear of being victimised themselves.

trade.[7]

2.1 Model setup

Consider an economy populated by a continuum of agents, uniformly distributed along a circumference of length $2L$, with mass one per unit arc distance. There are two periods. In period 1 each agent visits a location on the circle. There he finds a criminal opportunity, say stealing, worth $W > 0$. He can choose whether to take it or to forgo it (actions S and H, respectively). The returns to honest behaviour are normalised to 0. Movements along the circle are localised, in the sense that agents are more likely to visit locations close to them than far away. Specifically, the probability density of visiting a location at distance $x \in [0, L]$ is $f(x) = \frac{\alpha e^{-\alpha x}}{2(1-e^{-\alpha L})}$. By distance we mean the shorter arc distance. The same $f(x)$ applies to both locations at distance x, clockwise and counterclockwise. The denominator grants that $2\int_0^L f(x)dx = 1$. The parameter $\alpha > 0$ captures traveling costs and therefore the degree of localization: the distribution of movements converges to uniform as $\alpha \to 0$ and is more localised the higher α.

In period 2 each agent visits a location, with the same distribution as in period 1. There he is matched to one random resident of the visited location for possible interaction. The visited agent may either accept or reject interaction (actions A and R, respectively). Before deciding, the visited agent perfectly observes his visitor's movements, both present and past, but may or may not know what his visitor did in period 1. Specifically, there is a localised communication technology, by which, if my current visitor previously visited a location at distance z from myself, with probability $e^{-\beta z}$ I receive (accurate) news of what he did there, and with probability $1 - e^{-\beta z}$ I receive no news.

In most cases, that is when two normal types (N types) meet, interaction is mutually profitable: if the visited agent chooses A, then both he and his visitor get $I > W$; if the visited agent chooses R, then both get 0. A small fraction ϵ of the population is made of dangerous types (D types). Both N and D types are uniformly distributed on the circle. D types make no choice: they always steal in period 1, always accept interaction with visitors in period 2, and interaction with them yields a negative payoff

[7] The main difference from Dixit (2003) is in the payoff structure and the matching technology. In a pioneering study, Kandori (1992) already shows that frequent social interaction allows community enforcement (a situation in which dishonest behaviour against one agent is sanctioned by other members in the society) and that what is crucial for community enforcement is information transmission among community members (see also Ellison, 1994) and Vega-Redondo (2006).

$-K$. Therefore, if I expect my visitor to be a D type, the rational response is to choose R.

In this model, it is the fear of been confused with D types, and therefore to be excluded from beneficial future interaction, that induces honest behaviour in period 1. Moreover, D types play the technical role of pinning down out of equilibrium beliefs. To allow N types to accept interaction with the average visitor, we assume $\epsilon \in \left(0, \frac{I}{I+K}\right]$.

The parameter $\beta > 0$ captures the key aspect of our analysis: the density of social interaction. We say that social interaction is dense if β is small, that is, if I am very likely to receive news of what other people did, as it is typically the case in small and isolated communities, where everybody is aware of what everybody else did. Conversely, a high β is suited to describe a big city, where anonymity means lower information diffusion. To simplify later exposition we assume $\beta \neq \alpha$.

2.2 Equilibrium

A (weak) perfect Bayesian equilibrium may be summarised by a distance $X \in [0, L]$ (the extent of honest behaviour), such that N types behave according to the following rule:

- in period 1 play H within distance X and S afterwards;
- in period 2 play R if you know that your current visitor previously went closer than X and chose S, and play A otherwise.

Let $\omega_L \equiv \frac{\alpha\left[e^{-\beta L} - e^{-\alpha L}\right]}{(\alpha-\beta)[1-e^{-\alpha L}]}$ and $\omega_H \equiv \frac{\alpha\left[1-e^{-(\alpha+\beta)L}\right]}{(\alpha+\beta)[1-e^{-\alpha L}]}$. This equilibrium is characterised as follows.

Proposition 1 ***(Equilibrium extent of honesty)***

- *If* $\frac{W}{I} > \omega_H$*, then stealing is optimal at any distance and* $X = 0$.
- *If* $\frac{W}{I} < \omega_L$*, then honesty is optimal at any distance and* $X = L$.
- *If* $\frac{W}{I} \in [\omega_L, \omega_H]$*, then* $\exists!\ X^* \in [0, L]$ *(the maximum extent of honesty) such that* X *is an equilibrium if and only if* $X \in [0, X^*]$.

Proof First consider period 2 beliefs, denoted by μ. Call y the distance previously covered by an N type's current visitor. Denote $\mathcal{S}$, $\mathcal{H}$ and $\mathcal{O}$ the events of having received news that the visitor previously chose S, of having received news that he chose H and of having received no news, respectively. Equilibrium strategies imply the following beliefs: if $y \le X$, then $\mu(D|\mathcal{S}) = 1,\ \mu(D|\mathcal{H}) = 0,\ \mu(D|\mathcal{O}) = \epsilon$; and if $y > X$, then $\mu(D|\mathcal{S}) = \mu(D|\mathcal{O}) = \epsilon,\ \mu(D|\mathcal{H}) = 0$.

Choosing R if $y \le X$ and $\mathcal{S}$ is then clearly optimal. Choosing A otherwise is optimal if $\epsilon \le \frac{I}{I+K}$, which we assumed above.

Consider now period 1 choices. Call x the distance traveled by an N type. The expected payoff to honest behaviour is $\pi(H|x) = I,\ \forall x \in [0, L]$. The expected payoff to stealing is $\pi(S|x) = W + I$ if $x > X$; if instead $x \le X$, it is[8]

$$\begin{aligned}
\pi(S|x) &= W + \frac{\alpha I}{2\left(1 - e^{-\alpha L}\right)} \left\{ \int_0^x e^{-\alpha(x-y)} \left(1 - e^{-\beta y}\right) dy + \right. \\
&\quad + \int_0^{L-x} e^{-\alpha(x+y)} \left(1 - e^{-\beta y}\right) dy + \int_{L-x}^{L} e^{-\alpha(2L-x-y)} \cdot \\
&\quad \left. \cdot \left(1 - e^{-\beta y}\right) dy + \int_x^L e^{-\alpha(y-x)} \left(1 - e^{-\beta y}\right) dy \right\} = \\
&= I + \Delta(x),
\end{aligned}$$

where

$$\Delta(x) = W + \frac{\alpha I}{1 - e^{-\alpha L}} \left\{ \frac{\beta e^{-\alpha x} - \alpha e^{-\beta x} + e^{-(\alpha+\beta)L} \left[\alpha e^{\beta x} - \beta e^{\alpha x}\right]}{\alpha^2 - \beta^2} \right\}.$$

It is then obvious that stealing is the optimal choice at any distance $x > X$. To show the optimality of honesty at any $x \le X$, notice that $\Delta(x)$, the expected payoff advantage to stealing (over being honest) at distance $x \le X$, is a strictly increasing function of x.[9] This is intuitive: the incentive to steal is increasing in distance, because when an agent steals at a location further away, he is less likely to encounter in the future another agent who is aware of his stealing.

Choosing H at any $x \le X$ is optimal if and only if $\forall x \in [0, X],\ \Delta(x) \le 0$. It is immediate to show that $\Delta(L) > 0 \iff \frac{W}{I} > \omega_L$ and $\Delta(0) < 0 \iff \frac{W}{I} < \omega_H$. Monotonicity of $\Delta(x)$ then implies the three cases specified in the proposition. If $\frac{W}{I} > \omega_H$, then stealing is optimal at any distance and $X = 0$ is indeed an equilibrium. If $\frac{W}{I} < \omega_L$, then it is optimal to behave honestly at any distance and $X = L$ is indeed an equilibrium. The interesting case is $\frac{W}{I} \in [\omega_L, \omega_H]$. In this case $\exists!\ X^* \in [0, L] :\ \Delta(x) < 0\ \forall x \in [0, X^*]$ and $\Delta(x) > 0\ \forall x \in [X^*, L]$. X^* is the

[8] To be precise, both $\pi(H|x)$ and $\pi(S|x)$ should be multiplied by $(1 - \epsilon)$, but this is immaterial for choices.

[9] This is easily established by showing that $\Delta'(x)$ is equal in sign to $\left[1 - e^{-(\alpha+\beta)(L-x)}\right]$, independently of whether $\alpha > \beta$ or $\beta > \alpha$.

maximum extent of honesty and any $X \in [0, X^*]$ is indeed an equilibrium. ■

To give honesty its best chance, let us focus on the maximum extent of honesty X^*. For the purpose of our empirical investigation, the main result of this model is the following.

Proposition 2 ***(Honesty increases in social interaction density)***
Suppose $\frac{W}{I} \in [\omega_L, \omega_H]$. *Then* $\frac{\partial X^*}{\partial \beta} < 0$.

Proof Although this cannot be proved analytically, it can be established numerically (as in Dixit, 2003) by showing that $\frac{\partial \Delta(x)}{\partial \beta} > 0$[10]. ■

This is again intuitive: the denser social interaction, and thus the easier it is for information to spread over (the lower is β), the stronger the threat of social punishment and therefore the broader the maximum extent of honesty. Figure 1 gives a visual image of how the maximum extent of honesty declines in β. It plots X^* against β for the following parameter values: $L = 1$, $W = 1$, $I = 2$. The four lines correspond to different values of α. Notice that X^* is the (maximum) extent of honesty that may be sustained by social interaction alone, without any intervention of the state.[11]

To take the prediction of Proposition 2 to the data, we need to control for the other determinants of crime and we need good empirical proxies for α and β.

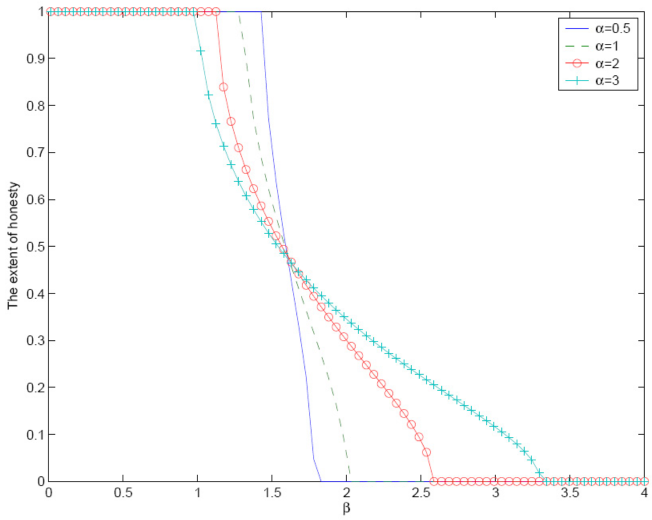
3 Data and potential determinants of crime

Our panel dataset comprises annual observations at the provincial level for the 103 Italian provinces, corresponding to level 3 in the NUTS (Nomenclature of Territorial Units for Statistics) classification by Eurostat, over the period from 1996 to 2003. We control for several variables, which are usually considered determinants of crime, thus minimising the risk that our results are driven by omitted variables. To further control for unobserved heterogeneity, for instance due to differences in cultural tolerance toward organised crime, we introduce macro-regional dummies.

[10] We have verified this with a fine grid for $L \in (0, 100]$, $\alpha \in (0, 10]$, $\beta \in (0, 10]$.

[11] Besides the fact that X^* is decreasing in β, one can also notice from Figure 1 that an improvement in traveling technology (a reduction in α) has opposite effects in societies characterised by dense and by coarse social interaction: in the former ones, say small towns, it expands the range of honest behaviour; in the latter ones, say big cities, it shrinks the range of honesty.

Figure 1: The maximum extent of honesty X^* is increasing in the density of social interaction (decreasing in β)



3.1 Crime Variables

Crime data, representing the dependent variable, are collected from Judicial Statistics, a yearly statistical publication of the Italian Statistics Institute (ISTAT). From these supplements we collect information about the evolution of different crimes for the 103 Italian provinces between 1996 and 2003.[12] In particular, we distinguish among different forms of violent crime (*murder*, *assault*, *rape*) and of property crime (*robbery*, *common theft*, *aggravated theft*, *car theft*, *bag snatch and pickpocketing*, *breaking*).[13]

[12]ISTAT publishes two crime statistics, those reported by the police (*Polizia di Stato*, *Carabinieri* and *Guardia di Finanza*) to the judiciary authority (*Le Statistiche della Delittuosità*), and those for which the judiciary authority starts the penal prosecution (*Le Statistiche della Criminalità*). The two statistics differ because the judiciary activity is delayed with respect to the time the crime has been committed, and moreover crimes reported by the police have a finer disaggregation. For these reasons we opt for using crime as measured by the police.

[13]Report rates vary a lot with the type of crime. The last available victimisation survey coducted by ISTAT in 2002 shows that car theft report rate is 94.5%, breaking 69.1%, bag snatch 54.4%, robbery 49.6%, pickpocketing 48.7%, theft 26.8%, violent assault 21.8% and rape 7.4%. Our control for macroareas smaller than the usual North, Centre and South hopefully captures part of the unobserved effect due to report rates. Clearly, the fact that report rates are available only at national level and for one year (2002) determines their unobservability in our provincial panel and may induce an unavoidable bias.

3.2 Socioeconomic, Demographic and Geographic Controls

Our dataset comprises a set of socioeconomic and demographic variables that are likely to be correlated with crime rates. The explanatory variables are separated into three groups: deterrence, demographic and socio-economic variables. Deterrence variables determine the expected returns to crime. In particular, we proxy for the probability of apprehension with the clear-up rate (*Clear*), defined as the ratio of the number of crimes cleared up by police to the total number of reported crimes.

We include three variables that measure the distribution of the regional male population across age categories (15-19, 20-24 and 25-29 years). Young men are said to be more prone to engage in criminal activities than the rest of the population (Freeman, 1991; Grogger, 1998).

We complete our dataset by including a set of economic variables: logarithm of real GDP per capita (*lnGDP*) and the unemployment rate for people aged 15-64 years (*Unemployment*). These factors proxy for the general level of prosperity in each province and thus for legitimate and illegitimate earning opportunities (Ehrlich, 1973).

Our list of control variables is likely to be incomplete, since it is impossible to control for all factors affecting crime. Thus, to control for unobserved factors, we exploit the panel structure of our dataset. We include year dummies in order to adjust for the effect of factors that cause exhogenous shocks in crime rates that are common to all provinces. Moreover, as a robustness check we add to our econometric specification both linear and quadratic province–specific time trends to control for variation in within-province crime rates due to factors that are province specific over time.

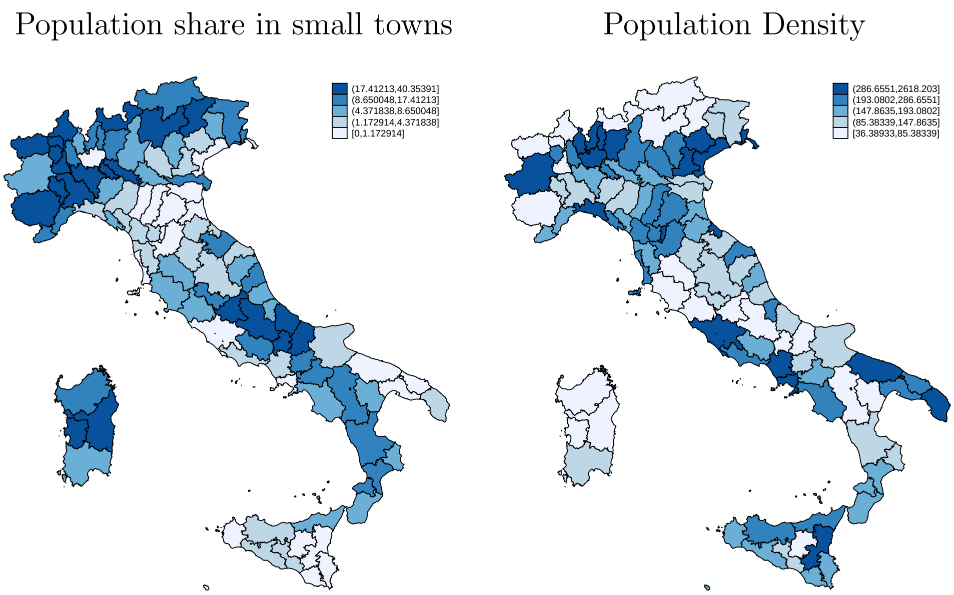
4 Social sanction and social interaction density

Our dataset does not contain direct information on social sanction. We thus exploit the theoretical model, which links social sanction to the key parameters α and β, capturing travel costs for individuals and for information, respectively.

We use five proxies, taken from demographic and geo-morphological information for each province: the proportion of population living in towns with less than 2,000 inhabitants (*C2K*), the proportion living in cities with more than 250,000 inhabitants (*BC250K*), the overall population density (*Density*), and the fraction of territory constituted by mountains (*Mountain*) and by hills (*Hill*).

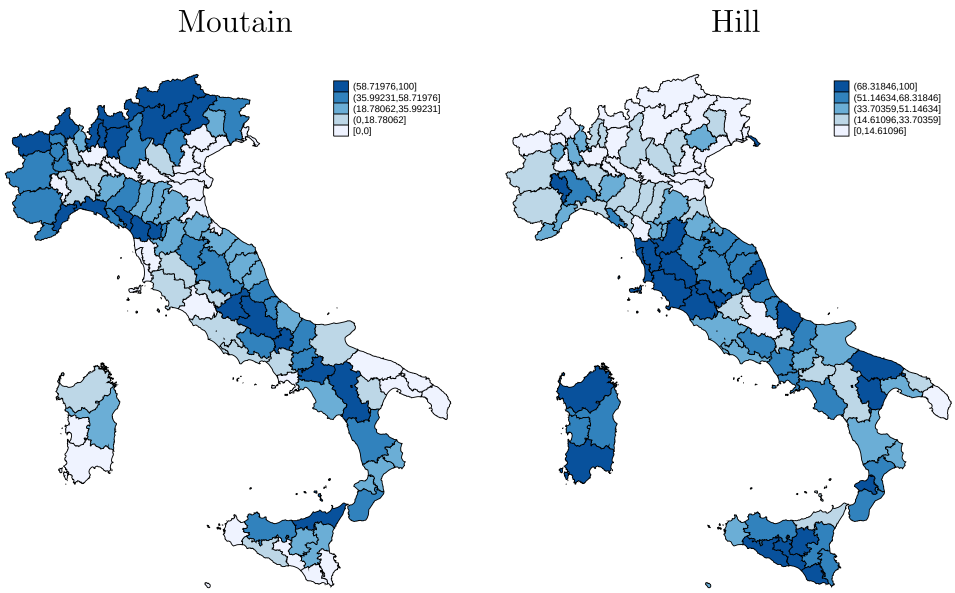
BC250K serves to isolate the role of social sanction from the notorious effect of big cities on crime (Glaeser and Sacerdote, 1999). *C2K* is introduced to capture social interaction density. In a similar vein, Goldin and Katz (1999) use County level data for Iowa (USA) in 1915 and use the proportion of County population living in small towns (less than 1,700 inhabitants) as a proxy for closed and cohesive social networks. Yet in 1915 Iowa the overall population density was very low, therefore living in a small village meant at the same time living kilometers far away from other towns. Nowadays Italy on the contrary is characterized by a very high population density. This means that living in a small town does not necessarily mean living in an isolated place. An example is the Po valley in Northern Italy: towns can be really small, below 2,000 inhabitants, but they often happen to be one beside another with no free land in the middle. This means that the percentage of population living in small towns alone does not necessarily identify isolated communities, where everybody knows each other. Therefore, we need to simultaneously control for overall population density and for the presence of mountains and hills. The geographical distribution of *C2K*, *Density*, *Mountain* and *Hill* is shown in Figures 2 and 3.

Figure 2: Geographical distribution of the population share living in small towns and of population density in 1996



Recall that what matters in the theoretical model is the probability that tomorrow I will encounter somebody who is aware of what I did today. We believe that, once we explicitly control for population density and for geo-morphological characteristics, we can interpret *C2K* as a good proxy of social interaction density.

Figure 3: Geographical distribution of mountainous and hill territory



In order to be suitable proxies for social interacions density and big city effects, *C2K*, *BC250K* and *Density* should be exogenous. One potential source of endogeneity is reverse causality: while our model predicts that social interactions density determines crime rates, it may be that people decide where to live as a consequence of criminality. Thus, higher crime rates in big cities and lower in isolated community would induce to relocate from main cities to smaller towns. Given the long panel structure of our sample, we can test whether *C2K* and *BC250K* time series follows crime rates changes along time. As an example, an increase in crime rates in a given province would induce (with some time lag) a reduction in the fraction of people living in big cities and an increase in the fraction of those living in small towns.

Table 1 presents the analysis of variance (ANOVA) for the three demographic variables used in the empirical part. It clearly emerges that *C2K*, *BC250K* and *Density* do not vary significantly over time. Indeed, the standard deviation within provinces assumes very low values and is not statistically different from zero,[14] while the standard deviation between provinces is significantly higher suggesting a relevant heterogeneity across provinces in our demographic measures. Since the same exercise run on crime rates reveals substantial within province variation, the time invariance of the

[14] Variance of within standard deviations and the relative significance tests are not reported, but are available upon request.

Table 1: Anova analysis of demographic and crime variables

Variable	Mean	Std. dev. (between)	Std. dev. (within)
C2K	9.628	9.602	0.398
BC250K	5.119	14.484	1.765
Density	241.373	328.957	3.345
Murder (rate)	0.013	0.012	0.008
Assault (rate)	0.509	0.226	0.165
Rape (rate)	0.037	0.010	0.018
Theft (rate)	19.154	8.579	2.536
Agg. theft (rate)	3.905	1.399	0.855
Robbery (rate)	0.377	0.367	0.114
Breaking (rate)	3.529	1.313	0.838
Car theft (rate)	2.614	2.308	0.699
Bag snatch (rate)	1.882	2.140	0.550

geo–demographic variables rules out potential endogeneity due to reverse causality.

5 Empirical methodology

Our dataset is a balanced panel: we have 103 observations (one for each province) observed over eight years, from 1996 to 2003. A pooled OLS is likely to be inefficient, since the IID hypothesis on the error terms is inappropriate in panel data settings. Thus, the choice remains open between a fixed effects (FE) and a random effects (RE) specification. In our case we are forced to choose RE: FE estimators are based on within–group heterogeneity, i.e. they require all the explanatory variables to vary within each group (in our case, within each province). Two of our key explanatory variables are based on the shape of a province's territory, which is clearly invariant. This is not the only time–invariant regressor: in order to exploit cross–sectional variations and address unobserved heterogeneity we include macro–regional dummies for areas which are likely to be homogeneous along relevant unobserved characteristics, namely the North, the Centre and the South of Italy. To go even further, we split Southern Italy in three further macro-regions, distinguishing between areas plagued by organized crime (Sicily, Calabria, Campania and Puglia), mainland areas which are not, and Sardinia. Excluding all those variables from the regressors in order to perform a FE estimation would leave us just one proxy for social interaction density, *C2K*. We already mentioned that this proxy is hard to interpret if used alone. Moreover, as previously discussed and reported in table 1, *C2K*, *BC250K* and *Density* do not vary significantly over time, hence a FE estimation would be affected by quasi–collinearity.

The econometric model to be estimated in its most general form is the following error components model:

$$Y_{it} = \boldsymbol{X}_{it}\boldsymbol{\beta} + \nu_i + \epsilon_{it} \quad i = 1, \ldots, 103; \; t = 1, \ldots, 8 \tag{1}$$

where $\boldsymbol{X}$, ν_i and ϵ_{it} are independent of each other and both uncorrelated with the explanatory variables. Y_{it} is the log of per capita crimes in province i in year t. We will estimate the same model for each type of crime of interest defined in the data section, and then compare the results.

Defining $\xi_{it} = \nu_i + \epsilon_{it}$, the assumption that shocks are independent can be rewritten as

$$\begin{array}{rcll} \mathrm{Var}(\xi_{it}) & = & \sigma_\nu^2 + \sigma_\epsilon^2 & \\ \mathrm{Cov}(\xi_{it}, \xi_{is}) & = & \sigma_\nu^2 & \forall t \neq s \\ \mathrm{Cov}(\xi_{it}, \xi_{js}) & = & 0 & \forall t \neq s, i \neq j \end{array}$$

Under the RE specification, homoskedasticity in both ν_i and ϵ_{it} and no serial correlation in ϵ_{it}, the variance-covariance matrix of the errors becomes

$$V = {\sigma_\nu}^2(I_N \otimes \boldsymbol{i}_T\boldsymbol{i}'_T) + {\sigma_\epsilon}^2(I_N \otimes I_T) \tag{2}$$

where I_N is the $N \times N$ identity matrix and $\boldsymbol{i}_N$ is a $N \times 1$ vector of 1. Therefore, V is block-diagonal with

$$V = I_N \otimes \begin{bmatrix} \sigma_\epsilon^2 + \sigma_\nu^2 & \sigma_\nu^2 & \ldots & \sigma_\nu^2 \\ \sigma_\nu^2 & \sigma_\epsilon^2 + \sigma_\nu^2 & \ldots & \vdots \\ \ldots & & \ddots & \sigma_\nu^2 \\ \sigma_\nu^2 & & & \sigma_\epsilon^2 + \sigma_\nu^2 \end{bmatrix} \tag{3}$$

Observations regarding the same province share the same ν_i effect, thus the relative errors are autocorrelated, with $Corr(v_{is}v_{it}) = \frac{\sigma_\nu^2}{(\sigma_\epsilon^2+\sigma_\nu^2)}$. Ordinary least squares estimates for β in model (1) are therefore inefficient, though consistent. Generalized least squares (GLS) are the efficient solution if Ω is known. Various feasible GLS procedures exist drawing on consistent estimators of Ω. The standard approach to RE panels is to assume (3). This is not feasible in our case: the economic model explicitly accounts for the role of distance between a potential criminal's place of residence and the place where interaction takes place. Therefore, in order not to have spurious effects on our parameters of interest it is relevant to control for spatial heterogeneity.

5.1 Spatial structure

Spatial heterogeneity is a typical problem econometricians have to account for when dealing with province level data. There are good reasons to think that criminal activity may not follow provincial administrative boundaries: there may be spillover effects from one province to another, or simply the province may not be a relevant aggregation for the phenomena we are observing. Distance between points in space is usually characterized by means of a *proximity matrix*, say W, containing a measure of proximity for every pair of data points and, by convention, setting the diagonal to zero. Hence a *spatial lag operator* is defined such that Wy, the *spatial lag* of y, stands for "the average value of y at *neighbouring* locations".[15] Anselin (1988) warns about the relevant consequences on estimation (and, to a lesser extent, on inference) of the choice of W. Here we resort to a proximity matrix, where each entry w_{ij} is the inverse of the traveling distance between provinces i and j (measured as the distance between their capital cities).[16]

The two standard specifications for spatial effects in regression models are the *spatial lag* (SAR) model:

$$y = \rho W y + X\beta + \epsilon \tag{4}$$

and the *spatial error* (SEM) model:

$$\begin{aligned} y &= X\beta + e \\ e &= \lambda W e + \epsilon \end{aligned} \tag{5}$$

The consequences on estimation of omitting the lagged dependent variable are inconsistency and bias of parameter estimates. Neglecting a spatial error structure has less serious consequences: estimates, while still consistent, are inefficient. Therefore, we concentrate our analysis on a SAR extension of our panel random effects model

[15] See Anselin (1988), Ch.3, for a classic treatment.

[16] A common practice in spatial econometrics is to row–standardize the matrix, and to set proximity to 0 for w_{ij} smaller than a certain threshold. Both transformations help convergence in maximum likelihood estimation: since we are going to use a GMM estimator we are not forced to do them. Further on, row–standardization is typically used when the W matrix is a 0/1 matrix, i.e. a matrix that sets weights to 1 for neighboring locations and 0 otherwise. In our case, row–standardizing a coordinates-based matrix would change proximities w_{ij} from absolute distance measures to relative ones, thus loosing part of the information.

and we bootsrap standard errors in order to accout for unmodelled heterogeneity. This is particularly relevant because it comprises arbitrary spatial structures on the error term, i.e. any SEM specification on ϵ. Following Elhorst (2003), stacking the data as one cross section for every point in time, the panel RE version of (4) becomes

$$y = \rho(I_T \otimes W)y + X\beta + (i_T \otimes \nu) + \epsilon \tag{6}$$

We estimate the model with a GMM estimator, that does not impose any further restriction on the idiosyncratic error term ϵ. In particular, we use the Generalized Spatial two Stage Least Squares (GS2SLS) estimator of Kelejian and Prucha (1998). If we do not assume a SEM structure on ϵ, the GS2SLS estimator is just the usual 2SLS estimator where instrumental variables for Wy must be chosen among the vectors $WX, W^2X, \ldots, W^nX$: we restrict our search to WX, i.e. the weighted output variable y is instrumented by the weighted exogenous Xs. The panel version of a two stage least squares estimators that allows to account for endogenous regressors in a Random Effects panel data model has been developed by Balestra and Varadharajan-Krishnakumar (1987).

6 Discussion of the results

Table 4 reports estimation results for property crimes. Our main indicator for the strength of social sanctions, the percentage of province population living in towns with less than 2000 inhabitants (*C2K*), is always significant and negative, confirming the empirical implications of the model outlined in section 2.1. The magnitude of such an effect is relevant: its absolute value lies between 0.0104 for breaking and 0.0316 for bag snatch and pickpocketing. This means that an increase by 10% of population living in small cities, reduces crime rates between 11% and 31%.

The other proxies for community isolation (*Mountain* and *Hill*) are not individually significant and so is *Density*. Their interactions with *C2K* are so small that marginal effects are barely changed.[17] Nevertheless, as we argued above, the inclusion of population density and of the morphological variables is fundamental for *C2K* to be a correct proxy for social interaction density, and therefore for the strength of social sanction.

We also argued that, as Glaeser and Sacerdote (1999) point out, there may be different causal effects of city size on crime rates, which are not related to social sanction. First, big cities may offer higher expected returns to crime because of higher income

[17]The full estimation results, including interactions, are available upon request.

and wealth levels and because of smaller detection probability. Second, if travel costs are high for criminals, we would expect criminals to be disproportionately concentrated in big cities. Third, big cities might attract crime-prone individuals (or make them crime-prone). If we did not explicitly control for the population share living in big cities, the possible action of all these effects would cast serious doubts on our interpretation of *C2K*. Including among regressors *BC250K*, the percentage of people living in cities with more than 250 thousands inhabitants, we find that it is significant and positive for thefts, car thefts and bag snatches. Its absolute value is always smaller than that of *C2K*. Since we control for isolation and population density, we can reasonably think the two variables measure two different determinants of crime: *C2K* captures the effect of social sanction, whereas *BC250K* controls for the various other effects of big cities. Our results suggest that social sanction in small communities has a higher marginal effect than urban life in big cities.

Social sanctions seem not to matter for violent crimes (see table 5): the parameter of interest is not significant for murder, nor for rapes or serious assaults. This is not surprising given the different nature of property and violent crime: the model we set up is likely to fit the motives behind the former much better than those behind the latter.

Concerning the performance of the other variables listed in Subsection 3.2, clearance rate generally exerts negative effects on crime rates with the exception of bag snatch and murder. Economic variables, such as GDP per capita and unemployment rate, do not present consistent patterns and are sensitive to the type of crime rate considered. These findings are in line with those of the literature, that has failed to reach a consensus on the relationship between GDP per capita, unemployment and crime. For instance, Buonanno (2006) shows that unemployment exerts no significant effect on property crimes in Italy as a whole, whereas it has a large and positive effect on crime rates in southern regions. Finally, the effect of the share of young males depends on the type of crime.

The spatial autoregressive term is significant for five crime types out of nine, and in all those cases (but for robbery) the sign is positive as we expected. While interesting per se, such a result confirms that controlling for spatial autocorrelation is necessary to obtain consistent estimates. As explained in section 5.1 and confirmed by Hausman tests (see table 6), Wy is endogenous. We use the same instruments for all our specification: *WDensity* and *WC2K*. Given that they have no direct economic interpretation - remember they are selected within the set W^nX, $\forall n$ thanks to the GS2SLS estimator we use - they are selected in order to be statistically relevant and exogenous: as table 6 reports they are relevant (F–test of joint significance on the first stage estimates) and exogenous (Hansen J–test for overidentification).

7 Robustness checks

In this section we perform a set of alternative regressions to provide further evidence that the results obtained in the previous section are robust to different specifications and to the inclusion of additional regressors.

First, we differently define the variable used to capture the big city effects. In particular, we replace the percentage of population living in cities with more than 250 thousands inhabitants (*BC250K*) with the population living in cities with more than 100 thousands (*BC100K*) and 500 thousands (*BC500K*) inhabitants. Tables 7 and 8 do not show any significant difference due to different definitions of the big city size.[18]

Second, we exploit the panel aspects of our data and we add to our econometric specification both linear and quadratic province-specific time trends to control for variation in within-province crime rates due to factors that are province specific over time (Raphael and Winter-Ebmer, 2001). The magnitudes of the effects of both *C2K* and *BC250K* are very stable across specifications. The results, presented in Table 9, indicate that the size of the coefficient on *C2K* slightly decreases when linear and quadratic province-specific trends are included, with the exception of bag snatch.

8 Conclusion

Social sanctions may be a very strong deterrent of crime. We may be unwilling to commit crime against those people we are very likely to interact with in the future, as well as against those who are close to them (in a social network sense). The reason is that these people are likely to be aware of our actions and they are in a position to impart us a social sanction. Since this risk is higher in smaller and isolated communities, in which social networks are closed and the density of social interaction is high, we might expect less crime in such places. Yet some recent contributions of crime economics emphasise that crime rates might actually be higher where social interaction is denser, for instance due to higher possibilities of know-how sharing among criminals.

[18]Moreover, as additional control we also use the percentage the percentage of population living in cities with more than 50 thousands inhabitants (*BC50K*) and with more than 80 thousands (*BC80K*). Results, available upon request, do not show any significant difference with respect to our preferred specification.

We present a formal model that relates crime and social sanction to social interaction density. The main testable implication is that the degree of honesty enforceable by social sanction is increasing in social interaction density. We test the model using a panel dataset of the 103 Italian provinces between 1996 and 2003. We exploit detailed demographic and geo-morphological information to develop exogenous measures of social interactions density. Moreover, we carefully take into account the spatial correlation of crime between contiguous provinces and estimate a spatial panel model by means of GMM procedure. In line with the theoretical model, our findings show that the negative relation between social sanctions and crime is indeed there, it is significant, substantial in magnitude and robust across different specifications.

References

Allcott, H., D. Karlan, M. M. Möbius, T. S. Rosenblat, and A. Szeidl (2007). Community size and network closure. *American Economic Review 97*(2), 80–85.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

Balestra, P. and J. Varadharajan-Krishnakumar (1987). Full information estimations of a system of simultaneous equations with error component structure. *Econometric Theory 3*(2), 223–246.

Becker, G. S. (1968). Crime and punishment: an economic approach. *Journal of Political Economy 76*(2), 169–217.

Buonanno, P. (2006). Crime and labour market opportunities in Italy (1993-2002). *Labour 20*(4), 601–624.

Buonanno, P., D. Montolio, and P. Vanin. Does social capital reduce crime? *Journal of Law & Economics*. Forthcoming.

Calvó-Armengol, A. and Y. Zenou (2004). Social networks and crime decisions: the role of social structure in facilitating delinquent behavior. *International Economic Review 45*(3), 939–958.

Chamlin, M. B. and J. K. Cochran (1997). Social altruism and crime. *Criminology 35*(2), 203–228.

Coleman, J. S. (1988). Social capital in the creation of human capital. *The American Journal of Sociology 94*(Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure), S95–S120.

Dixit, A. (2003). Trade expansion and contract enforcement. *Journal of Political Economy 111*(6), 1293–1317.

Ehrlich, I. (1973). Participation in illegitimate activities: a theoretical and empirical investigation. *Journal of Political Economy 81*, 521–565.

Elhorst, J. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review 26*(3), 244–268.

Ellison, G. (1994). Cooperation in the prisoner's dilemma with anonymous random matching. *The Review of Economic Studies 61*(3), 567–588.

Elster, J. (1989). Social norms and economic theory. *The Journal of Economic Perspectives 3*(4), 99–117.

Falk, A., E. Fehr, and U. Fischbacher (2005). Driving forces behind informal sanctions. *Econometrica 73*(6), 2017–2030.

Freeman, R. B. (1991). Crime and the employment of disadvantaged youths. *NBER working paper N. 3875*.

Funk, P. (2004). On the effective use of stigma as a crime deterrent. *European Economic Review 48*(4), 715–728.

Glaeser, E. L. and B. Sacerdote (1999). Why is there more crime in cities? *The Journal of Political Economy 107*(6/2), 225–258.

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

Goldin, C. and L. F. Katz (1999). Human Capital and Social Capital: the rise of secondary schooling in America, 1910 to 1940. *The Journal of Interdisciplinary History 29*(4), 683–723.

Grogger, J. T. (1998). Market wages and youth crime. *Journal of Labor Economics 16*, 756–791.

Haynie, D. L. (2001). Delinquent peers revisited: does network structure matter? *American Journal of Sociology 106*(4), 1013–1057.

Heaton, P. (2006). Does religion really reduce crime. *Journal of Law & Economics 49*(1), 147–172.

Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic Studies 29*(1), 63–80.

Kelejian, H. H. and I. R. Prucha (1998). A Generalized Spatial Two–Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics 17*(1), 99–121.

Lederman, D., N. Loayza, and A. M. Menendez (2002). Violent crime: Does social capital matter? *Economic Development and Cultural Change 50*(2), 509–539.

Messner, S. F., E. Baumer, and R. Rosenfeld (2004). Dimensions of social capital and rates of criminal homicide. *American Sociological Review 69*(6), 882–903.

Millo, G. and G. Pasini (2006). Does social capital reduce moral hazard? a network model for non-life insurance demand. *Venice University Economics Department WP N. 59*.

Patacchini, E. and Y. Zenou (2005). Crime and conformism. *CEPR Discussion Papers N. 5331*.

Patacchini, E. and Y. Zenou (2007). Ethnicity and spatial externalities in crime. *CEPR Discussion Paper N. 6130*.

Polinsky, M. A. and S. Shavell (2000). The economic theory of public enforcement of law. *Journal of Economic Literature 38*(1), 45–76.

Posner, R. A. (1997). Social norms and the law: An economic approach. *American Economic Review 87*(2), 365–369.

Raphael, S. and R. Winter-Ebmer (2001). Identifying the effect of unemployment on crime. *Journal of Law & Economics 44*(1), 259–284.

Rasmusen, E. (1996). Stigma and self-fulfilling expectations of criminality. *Journal of Law & Economics 39*(2), 519–544.

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

Sherman, L. W., D. A. Smith, J. D. Schmidt, and D. P. Rogan (1992). Crime, punishment, and stake in conformity: Legal and informal control of domestic violence. *American Sociological Review 57*(5), 680–690.

Silverman, D. (2004). Street crime and street culture. *International Economic Review 45*(3), 761–786.

Vega-Redondo, F. (2006). Building up social capital in a changing world. *Journal of Economic Dynamics & Control 30*(11), 2305–2338.

Weibull, J. and E. Villa (2005). Crime, punishment and social norms. *SSE/EFI Working Paper N. 610*.

Table 2: Variables description

Variable	Description
Theft	Logarithm of common thefts per 1,000 inhabitants at the province level
Agg. theft	Logarithm of aggravated thefts per 1,000 inhabitants at the province level
Robbery	Logarithm of robberies per 1,000 inhabitants at the province level
Breaking	Logarithm of breakings per 1,000 inhabitants at the province level
Car theft	Logarithm of car thefts per 1,000 inhabitants at the province level
Bag snatch	Logarithm of bag snatchs and pickpocketings per 1,000 inhabitants at the province level
Murder	Logarithm of murders per 1,000 inhabitants at the province level
Assault	Logarithm of serious assaults per 1,000 inhabitants at the province level
Rape	Logarithm of rapes per 1,000 inhabitants at the province level
Clear	Clear–up rate for each province and crime category
C2K	Percentage of population living in towns with less than 2,000 inhabitants at the province level
BC250K	Percentage of population living in cities with more than 250,000 inhabitants at the province level
BC50K	Percentage of population living in cities with more than 50,000 inhabitants at the province level
BC80K	Percentage of population living in cities with more than 80,000 inhabitants at the province level
BC100K	Percentage of population living in cities with more than 100,000 inhabitants at the province level
BC500K	Percentage of population living in cities with more than 500,000 inhabitants at the province level
Mountain	Percentage of mountainous territory at the province level
Hill	Percentage of hill territory at the province level
Density	Population density, inh. per sq. Km
lnGDP	Logarithm of real GDP per capita at the province level
Unemployment	Unemployment rate for population between 15 and 64 years old, at the province level
Male1519	Percentage of young males aged 15-19 in the population at the province level
Male2024	Percentage of young males aged 20-24 in the population at the province level
Male2529	Percentage of young males aged 25-29 in the population at the province level
Wy	Spatial autoregressive term

Table 3: Summary statistics

Variable	Mean	Std. dev.	Min	Max
Theft	2.857	0.435	1.349	4.152
Agg. theft	1.272	0.437	-0.371	2.434
Robbery	-1.256	0.703	-3.565	1.32
Breaking	1.164	0.451	-0.488	2.401
Car theft	0.629	0.798	-1.241	2.685
Bag snatch	0.113	1.043	-3.526	2.579
Murder	-5.152	1.79	-10.925	-2.063
Assault	-0.809	0.525	-2.735	0.747
Rape	-3.454	0.631	-10.106	-2.148
Clear (theft)	6.223	2.725	1.505	21.024
Clear (agg. theft)	7.972	3.011	1.922	28.283
Clear (robbery)	32.18	13.069	3.448	96.463
Clear (breaking)	5.779	2.953	1.216	27.273
Clear (car theft)	6.364	10.117	0.589	100
Clear (bag snatch)	3.967	3.6	0	36.364
Clear (murder)	62.128	35.724	0	100
Clear (assault)	82.150	13.182	21.875	100
Clear (rape)	84.430	15.094	0	100
Mountain	31.915	30.208	0	100
Hill	41.952	27.996	0	100
Density	241.37	327.57	35.87	2,626.23
C2K	9.628	9.569	0	42.559
BC50K	27.265	19.817	0	87.680
BC80K	18.344	20.325	0	87.680
BC100K	14.110	20.405	0	87.680
BC250K	5.120	14.531	0	70.489
BC500K	2.930	12.358	0	70.489
Unemployment	10.33	7.553	1.4	33.3
Male1519	2.754	0.585	1.697	4.46
Male2024	3.309	0.515	1.939	4.628
Male2529	3.814	0.265	2.851	4.728

Table 4: Property crimes: baseline regressions

	Theft	Agg. theft	Robbery	Breaking	Car theft	Bag snatch
C2K	-0.0117***	-0.0114**	-0.0190***	-0.0104**	-0.0176**	-0.0316***
	(0.0040)	(0.0050)	(0.0071)	(0.0050)	(0.0076)	(0.0094)
BC250K	0.0067***	0.0024	0.0092	0.0013	0.0125**	0.0200***
	(0.0022)	(0.0023)	(0.0063)	(0.0017)	(0.0059)	(0.0055)
Mountain	0.0008	0.0007	-0.0012	0.0009	0.0008	0.0003
	(0.0015)	(0.0017)	(0.0024)	(0.0018)	(0.0027)	(0.0032)
Hill	0.0022	-0.0002	0.0003	0.00002	0.0018	0.0098***
	(0.0019)	(0.0019)	(0.0029)	(0.0020)	(0.0030)	(0.0036)
Density	0.0001	0.00008	0.0007	-0.0002	0.0005	0.0006
	(0.0003)	(0.0003)	(0.0005)	(0.0002)	(0.0006)	(0.0006)
Clear	-0.0350***	-0.0339***	-0.0044***	-0.0285***	-0.0008	-0.0231***
	(0.0060)	(0.0039)	(0.0013)	(0.0045)	(0.0011)	(0.0053)
lnGDP	0.1189	0.2081	-0.2116	0.1985	0.1263	-0.1795
	(0.2275)	(0.2295)	(0.4092)	(0.2423)	(0.4465)	(0.5619)
Unemployment	-0.0012	-0.0073*	-0.0073	-0.0081**	0.0005	0.0017
	(0.0030)	(0.0038)	(0.0114)	(0.0035)	(0.0061)	(0.0131)
Male1519	0.0461	-0.0150	0.1286	-0.0994	0.3404**	-0.3255
	(0.1059)	(0.1128)	(0.1623)	(0.1215)	(0.1601)	(0.2910)
Male2024	-0.0244	-0.0897	-0.5139***	-0.0108	-0.0037	0.6529***
	(0.0851)	(0.0954)	(0.1343)	(0.1064)	(0.1592)	(0.2493)
Male2529	0.1304	0.0811	0.0602	0.0748	0.0775	-0.5376***
	(0.0822)	(0.0977)	(0.1184)	(0.1113)	(0.1452)	(0.2034)
Wy	-0.1640	0.2147	-1.7927*	0.1133	2.8937***	2.9501**
	(0.1601)	(0.4695)	(0.9874)	(0.4796)	(1.1172)	(1.2897)

Notes: This table presents the results of spatial estimates on a panel of yearly observations for all 103 Italian provinces during the period 1996-2003. The dependent variables are the logs of the number of crimes per 1,000 inhabitants reported by the police (source: ISTAT). The first column refers to the common thefts. The other columns refer to specific types of crimes, which are reported on top of each column. The variable $C2K$ is the percentage of provincial population living in towns with less than 2,000 inhabitants (source: ISTAT). Other regressors not included in the table: constant, interaction terms between $C2K$ and Mountain, Hill and Density (namely C2K*Mountain, C2K*Hill and C2K*Density). All estimates include macro regional dummies for the North, the Centre, and the South of Italy, where the South is further split in areas plagued by organised crime (Sicily, Calabria, Campania and Puglia), mainland areas which are not (Abruzzo, Molise and Basilicata), and Sardinia. Year fixed-effects are included in all specifications. Bootstrapped standard errors (1,000 replications) are presented in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%. GMM random effects panel spatial autoregressive model, with weighting matrix based on the inverse of road travelling distance between provinces.

Table 5: Violent crimes: baseline regressions

	Murder	Assault	Rape
C2k	-0.0049	0.0048	0.0066
	(0.0111)	(0.0054)	(0.0046)
BC250K	0.0042	-0.0019	0.0017
	(0.0057)	(0.0035)	(0.0023)
Mountain	0.0028	0.0047**	-0.0006
	(0.0043)	(0.0022)	(0.0019)
Hill	0.0044	0.0027	0.0032**
	(0.0040)	(0.0024)	(0.0016)
Density	0.0003	0.00008	0.00006
	(0.0005)	(0.0003)	(0.0003)
Clear	0.0295***	-0.0085***	0.0020
	(0.0022)	(0.0019)	(0.0032)
lnGDP	1.3063	0.1779	0.0032
	(0.9811)	(0.4082)	(0.3378)
Unemployment	0.0365	0.0149*	-0.0058
	(0.0243)	(0.0089)	(0.0084)
Male1519	0.8184*	-0.0400	-0.2066
	(0.4861)	(0.1791)	(0.2280)
Male2024	-0.4976	-0.2058	-0.4314**
	(0.6458)	(0.2066)	(0.2084)
Male2529	0.4609	0.8825***	0.2316
	(0.4457)	(0.2040)	(0.1965)
Wy	-0.0473	2.3397**	0.5906***
	(0.2638)	(0.9143)	(0.1483)

Notes: This table presents the results of spatial estimates on a panel of yearly observations for all 103 Italian provinces during the period 1996-2003. The dependent variables are the logs of the number of crimes per 1,000 inhabitants reported by the police (source: ISTAT). The first column refers to the murders. The other columns refer to specific types of crimes, which are reported on top of each column. The variable *C2K* is the percentage of provincial population living in towns with less than 2,000 inhabitants (source: ISTAT). Other regressors not included in the table: constant, interaction terms between *C2K* and *Mountain*, *Hill* and *Density* (namely C2K*Mountain, C2K*Hill and C2K*Density). All estimates include macro regional dummies for the North, the Centre, and the South of Italy, where the South is further split in areas plagued by organised crime (Sicily, Calabria, Campania and Puglia), mainland areas which are not (Abruzzo, Molise and Basilicata), and Sardinia. Year fixed-effects are included in all specifications. Bootstrapped standard errors (1,000 replications) are presented in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%. GMM random effects panel spatial autoregressive model, with weighting matrix based on the inverse of road travelling distance between provinces.

Table 6: Instrumental variables diagnostics

	Theft	Agg. theft	Robbery	Breaking	Car theft	Bag snatch
	(1)	(2)	(3)	(4)	(5)	(6)
F-test of joint significance of instruments						
F-test	166.18	149.53	143.61	113.86	115.06	35.59
p–value	0.000	0.000	0.000	0.000	0.000	0.000
Hansen J–tests for overidentification						
J–stat	0.133	0.726	0.459	0.740	0.363	2.373
p–value	0.716	0.394	0.498	0.389	0.547	0.123
Hausman test for Wy exogeneity						
χ^2 stat	7.16	2.76	7.77	4.38	2.19	0.03
p–value	0.008	0.097	0.005	0.036	0.139	0.854

	Murder	Assault	Rape
	(5)	(6)	(7)
F-test of joint significance of instruments			
F-test	172.05	132.38	183.31
p–value	0.000	0.000	0.000
Hansen J–tests for overidentification			
J–stat	0.007	0.031	0.960
p–value	0.933	0.859	0.327
Hausman test for Wy exogeneity			
χ^2 stat	4.09	0.19	1.29
pval	0.043	0.661	0.256

Notes: WDensity and *WC2K* are used as instruments for *Wy*. The F-statistic refers to the null hypothesis that the coefficients on the excluded instruments are jointly equal to zero in the 1rst stage. The Hansen J-test is a test of overidentifying restrictions, distributed as chi-square under the null of instrument validity. Finally, the Hausman test compares IV versus OLS under the null hypothesis that the OLS estimator is consistent and efficient.

Table 7: Property crimes: robustness checks

Panel A: including *BC100K*						
	Theft	Agg. theft	Robbery	Breaking	Car theft	Bag snatch
C2K	-0.0103***	-0.0110**	-0.0155**	-0.0103**	-0.0131*	-0.0270***
	(0.0040)	(0.0048)	(0.0069)	(0.0050)	(0.0075)	(0.0087)
BC100K	0.0065***	0.0020	0.0108***	0.0008	0.0134***	0.0191***
	(0.0017)	(0.0019)	(0.0034)	(0.0017)	(0.0037)	(0.0041)
Mountain	0.0005	0.0007	-0.0018	0.0008	0.0003	-0.0004
	(0.0015)	(0.0017)	(0.0024)	(0.0017)	(0.0026)	(0.0033)
Hill	0.0007	-0.0007	-0.0020	-0.0002	-0.0011	0.0060*
	(0.0019)	(0.0020)	(0.0030)	(0.0019)	(0.0031)	(0.0037)
Density	0.00009	0.00007	0.0006	-0.0002	0.0004	0.0005
	(0.0003)	(0.0003)	(0.0004)	(0.0003)	(0.0005)	(0.0005)
Wy	-0.2259	0.1586	-1.5348	0.0788	2.3346*	2.9013**
	(0.1609)	(0.4645)	(1.0341)	(0.4575)	(1.2390)	(1.3943)

Panel B: including *BC500K*						
	Theft	Agg. theft	Robbery	Breaking	Car theft	Bag snatch
C2K	-0.0120***	-0.0113**	-0.0184***	-0.0105**	-0.0175**	-0.0329***
	(0.0040)	(0.0048)	(0.0068)	(0.0051)	(0.0086)	(0.0095)
BC500K	0.0082**	0.0037	0.0158	0.0015	0.0198*	0.0199
	(0.0037)	(0.0030)	(0.0116)	(0.0042)	(0.0115)	(0.0148)
Mountain	0.0009	0.0007	-0.0012	0.0009	0.0006	0.0006
	(0.0016)	(0.0016)	(0.0024)	(0.0018)	(0.0028)	(0.0034)
Hill	0.0024	-0.00004	0.0009	0.00005	0.0021	0.0103**
	(0.0020)	(0.0019)	(0.0028)	(0.0021)	(0.0030)	(0.0041)
Density	0.00008	0.00005	0.0005	-0.0002	0.0003	0.0005
	(0.0003)	(0.0003)	(0.0005)	(0.0003)	(0.0006)	(0.0007)
Wy	-0.1042	0.2802	-2.0841**	0.1330	2.9454**	3.3829**
	(0.1774)	(0.4720)	(1.0068)	(0.4992)	(1.1709)	(1.4089)

Notes: See note to Table 4. Other regressors not included in the table: *Clear*, *lnGDP*, *Unemployment*, *Male1519*, *Male2024*, *Male2529*, constant and the interaction terms between *C2K* and Mountain, Hill and Density (namely C2K*Mountain, C2K*Hill and C2K*Density).

Table 8: Violent crimes: robustness checks

Panel A: including *BC100K*			
	Murder	Assault	Rape
C2K	-0.0045	0.0060	0.0075
	(0.0114)	(0.0059)	(0.0049)
BC100K	0.0034	0.0007	0.0024
	(0.0043)	(0.0028)	(0.0021)
Mountain	0.0026	0.0045**	-0.0007
	(0.0043)	(0.0021)	(0.0020)
Hill	0.0035	0.0028	0.0028*
	(0.0043)	(0.0024)	(0.0016)
Density	0.0003	0.00005	0.00004
	(0.0005)	(0.0004)	(0.0003)
Wy	-0.0191	2.2460**	0.6050***
	(0.2655)	(0.9310)	(0.1494)

Panel B: including *BC500K*			
	Murder	Assault	Rape
C2K	-0.0060	0.0038	0.0061
	(0.0116)	(0.0056)	(0.0046)
BC500K	0.0017	-0.0060	0.0003
	(0.0095)	(0.0068)	(0.0035)
Mountain	0.0029	0.0048**	-0.0006
	(0.0042)	(0.0021)	(0.0019)
Hill	0.0041	0.0022	0.0030*
	(0.0041)	(0.0023)	(0.0016)
Density	0.0003	0.0002	0.00007
	(0.0005)	(0.0003)	(0.0003)
Wy	-0.0319	2.6113***	0.6046***
	(0.2658)	(0.9357)	(0.1515)

Notes: See note to Table 5. Other regressors not included in the table: *Clear*, *lnGDP*, *Unemployment*, *Male1519*, *Male2024*, *Male2529*, constant and the interaction terms between *C2K* and Mountain, Hill and Density (namely C2K*Mountain, C2K*Hill and C2K*Density).

Table 9: Robustness checks: linear and quadratic trend

	Linear trend	Linear & quadratic trend
	C2K	*C2K*
Theft	-.0116***	-.0102***
	(.0021)	(.0027)
Agg. theft	-.0081***	-.0080**
	(.0025)	(.0036)
Robbery	-.0144***	-.0160***
	(.0042)	(.0057)
Breaking	-.0069**	-.0070*
	(.0028)	(.0040)
Car theft	-.0128***	-.0153***
	(.0035)	(.0044)
Bag snatch	-.0424***	-.0423***
	(.0055)	(.0072)
Murder	-.0258*	-.0275
	(.0149)	(.0253)
Assault	.0025	.0121*
	(.0047)	(.0064)
Rape	.0049	-.0055
	(.0065)	(.0106)

Notes: The dependent variables are the logs of the number of crimes per 1,000 inhabitants reported by the police (source: ISTAT). The first column include linear province-specific time trend, while the second column includes both linear and quadratic province-specific time trend. The variable *C2K* is the percentage of provincial population living in towns with less than 2,000 inhabitants (source: ISTAT). Other regressors not included in the table: *BC250K*, *Mountain*, *Hill*, *Density*, *Clear*, *lnGDP*, *Unemployment*, *Male1519*, *Male2024*, *Male2529*, constant and the interaction terms between *C2K* and *Mountain*, *Hill* and *Density* (namely C2K*Mountain, C2K*Hill and C2K*Density). All estimates include macro regional dummies for the North, the Centre, and the South of Italy, where the South is further split in areas plagued by organised crime (Sicily, Calabria, Campania and Puglia), mainland areas which are not (Abruzzo, Molise and Basilicata), and Sardinia. Year fixed-effects are included in all specifications. Bootstrapped standard errors (1,000 replications) are presented in parenthesis. * significant at 10%; ** significant at 5%; *** significant at 1%. GMM random effects panel spatial autoregressive model, with weighting matrix based on the inverse of road travelling distance between provinces.