Non-linearity between Crime and Education: Evidence from Italian Regions

Paolo Buonanno
University of Milan-Bicocca

Leone Leonida*
University of Calabria

June 30, 2005

Abstract

This paper analyses the effects of education on crime in Italy. We propose a theoretical framework to determine the effects of education and past criminal activity on crime rates. We show both the necessary conditions under which education reduces crime, and suggest that such conditions may be violated for high levels of education. In other words, we suggest that the relationship between education and crime may be non-linear. We test such prediction using Italian regions over the period 1980-1995. In line with our theoretical model, empirical results suggest that crime is negatively correlated to education for low and medium levels of education, and that criminality displays persistence over time. However, as expected, crime is positively correlated to education for high levels of education, a results that seems to be driven by a white collar effect. Our results are robust to model specification, endogeneity, changes in the typology of crimes and finally, to alternative definitions of education.

Key words: Crime; Education, Dynamics, GMM

JEL Classification: I2; J24; K42

* *Corresponding Author.* Address: Dipartimento di Economia e Statistica, Universita' degli Studi della Calabria, Via P. Bucci, 87036, Arcavacata di Rende, Cosenza, Italy. Tel: 0039 0984 492452. *Email*: l.leonida@unical.it. We would like to thank Piergiovanna Natale, Giulio Fella, Daniel Montolio and Giovanni Caggiano for useful comments and discussions about this paper. Financial supports from the University of Milan-Bicocca and University of Calabria are gratefully acknowledged.

1 Introduction

It is believed that criminals tend to be less educated than non criminals (Wilson and Herrnstein, 1985; Freeman, 1991). Such a belief is confirmed by looking at simple statistics: for example, in the US two-thirds of the incarcerated men had not graduated from high school (Freeman, 1996); analogously, more than 75% of the overall convicted population in Italy had not attained this level of education (see Official Statistics of Ministero della Giustizia, 2001). Despite this evidence few papers have studied the role of education and its impact on crime.[1]

Clearly, from a pure theoretical perspective, education may reduce the incentive to engage in criminal activities through several channels. Prior research suggests that higher levels of education are associated with higher returns in the labour market, increasing the opportunity cost of criminal behaviour. In an intertemporal setting, this means that education increases the cost associated with incarceration, since more educated individuals will experiences greater losses in earnings while in jail (Lochner, 2004). It is also suggested that education may alter personal preferences in a way that affects decisions to engage in crime - a sort of "civilization" effect. More specifically, Fajnzylber *et al.* (2002) suggest that education, incorporating a civic component, may increase the individual's moral stance, and then affects the individuals' perception of crime. Usher (1997) stresses that education perpetuates the values of society, enculturates people to serve their communities, and promotes the virtues of hard work and honesty. Lochner and Moretti (2004) claim that schooling generates benefits beyond the private return received by individual. Moreover, school enrolment alone, independently of the level of educational attainment, reduces the time available for participating in the crime activity (Witte and Tauchen, 1994).

These theoretical implications have been found to be empirically relevant. The empirical tests especially for the US (Lochner, 2004; Lochner and Moretti, 2004). Lochner (2004) and Lochner and Moretti (2004), using data from the National Longitudinal Survey of Youth (NLSY), found that high school graduation directly lowers criminal propensity even after controlling for market returns.

The models which have been proposed and the empirical results discussed above are

[1] Usher, 1997; Lochner, 1999 and 2004; Lochner and Moretti, 2004 represent few notable exceptions.

able to explain much of the relationship between education and crime; however, they are partially inconsistent with some of the events that have characterized recent years. In particular we refer to white collar crime: scandals such as the ones regarding Enron, Worldcom or Parmalat are only some examples of frauds that can be committed by skilled individuals, and especially by those working in large burocracies. A similar argument is made by Lochner (2004): while it is reasonable to hypothesize that unskilled crime - as theft, robbery, larceny - decreases as long as education increases, this does not need to be true for skilled crimes (e.g. fraud), for which education may increase the returns to crime more than return to legal work.

This, in turn, suggests that there may exist a non-linearity between the crime rate and education. In other words, the effects of education on delinquency may be conditional on the level education itself: when education is low, increasing education reduces crime rates; instead, when education is high, increasing education may lead to an increase in the criminal activities for certain categories of crime (e.g. fraud). The purpose of this paper is both to show the theoretical conditions under which such non-linearity may arise, and to test such hypothesis for the Italian regions.

To this aim, we propose a simple model of criminal behaviour, in which individuals decide how to allocate their disposable time between education, crime and work in the legal sector. Our model provides some insights into why education may affect criminal behaviour and allows us to identify some channels through which this happens. The model does not mean to be a complete description of criminal behaviour; rather, it wants to clearly state the conditions under which education may be positively correlated to crime rate. Education may increase the return to crime more than the returns to legal work for certain typologies of crimes such as fraud, forgery and embezzlement, hence suggesting that we should observe a non-linear relationship between crime rates and education. Our model has other interesting testable hypothesis. We show that crime depends on past crime experience, and that wage rate and law enforcement are important determinants of criminal behaviour.

We test the theoretical predictions of our model by using a panel data set for the twenty Italian regions for the period 1980 to 1995. In doing so, we focus our attention on property crimes, simply because it is recognised that are more likely to depend on economic motivations than violent crimes (i.e. murder, assault, rape, etc.). Our results suggest, as expected, that the effects of education on crime is non-linear, and that past criminal experience matters. More specifically, we show that the non-linearity is due to a white collar effect. We perform a number of exercises in order to deeply study the robustness of our results: first, we study the correlation between education and, separately, property crime rate, theft rate and total crime rate, always controlling for labour market opportunities (wage rate and employment rate), GDP per capita, growth rate of the GDP and a set of deterrence variables (police forces, quickness of conclusion of judicial proceedings and percentage of crime committed by unknown offenders). Also, we adopt different measures of education (the percentage of population with high school diploma, percentage of population with university degree and average years of schooling of the population). Finally, we consider inertial effects of crime (i.e. persistence over time) and control for endogeneity: we employ an instrumental variable approach for panel data using the GMM-system estimator that allows us to control for the joint endogeneity between crime rate and the explanatory variables and, also, for measurement errors in crime rates.

We contribute to the relevant literature in many repects. First, we propose a theoretical framework which shows the effects of education, past criminal activity, wage and law enforcement on crime rate. Second, our framework allows us, and to our knowledge we are the first to do so, to describe both the conditions under which education reduces crime, and when these conditions are likely to be violated, suggesting the possibility of a non-linear relationship between education and crime. Third, and again we are the first to do so, we study the relationships between education and crime in Italy, and assess the empirical relevance of the predicted non-linearity. Fourth, we deeply study the empirical reasons beyond such non-linearity, performing a number of robustness checks, and comparing results from our GMM-system model with results coming from a standard static panel data framework.

The paper proceeds as follows. Section 2 presents the existing literature on crime and the theoretical model we propose. Section 3 develops our econometric model. Section 4 describes our dataset and empirical results. Section 5 concludes.

2 The effects of education on crime: previous studies and motivation

2.1 Previous studies

The idea that education may have a negative effect on crime through its effect on wage rate is not a new one. As previously asserted in the introduction theory suggests several channels through which educational attainment may affect criminal behaviour. Most of these contributions stress how education raises individuals' skills and abilities, thus increases the returns from legitimate work, raising the opportunity costs of illegal behaviour. But there exist benefits from education that are not taken into account by individuals, this implies that the social return of education is higher than its private return (Lochner and Moretti, 2004).

Lochner (2004) explores the relationship between education and crime within the standard Becker (1964) and Ben-Porath (1967) investment model of human capital formation. In this paper he develops a theoretical framework in which he models individuals' decision to work, to be engaged in crime and to invest in education. Using microdata from the National Longitudinal Survey of Youth (NLSY) Lochner estimates the relationship between educational attainment and both property and violent crime controlling for a set of socioeconomic and demographic variables, such as age, race, unemployment rate and parental education. His results show a strong negative effect of education both on unskilled property crime and violent crime.

Lochner and Moretti (2004) explore the importance of the relationship between schooling and criminal participation using three different data sources: individual-level data from the Census on incarceration, state-level data on arrest from the Uniform Crime Reports (UCR) and sel-report data on crime and incarceration from NYLS. Education significantly reduces criminal activity, this finding is robust to different identification stategies and measures of criminal activity.

Usher (1997) analyses the relationship between education and criminal activity from a different point of view. He argues that education may also have a "civilization" effect, tending to reduce the incidence of criminal activity. Education conveys a civic externality, a benefit to society over and above the benefit to the student in enhancing his future earning power. Students are taught not only to be productive, but to be law abiding and loyal to their country. The civic externality is incorporated into an "anarchy" model where people choose to be farmers or bandits, and schooling inculcates a distaste for a life of crime. Estimates of the return to education are biased down when the civic externality is overlooked.

The Italian empirical literature on crime is relatively small. Crime has received little attention and has been almost neglected by Italian economists. Very few papers has been produced on this topics and most of them have been published in Italian, while from our point of view criminal activity in Italy should be more widely and carefully analyze. Campiglio (1990), using 1981 Italian census data at county level, finds a positive relationship between robberies and unemployment. Scorcu and Cellini (1998) analyze the long-run relationships between economic determinants and crime rates by using the tools provided by cointegration analysis in Italy over the period 1951 to 1994. Their main conclusion is that the level of real per capita consumption better explains homicides and robbery rates, while unemployment rate is the best economic explanatory factor for theft.

Marselli and Vannini (1997), using panel dataset at regional level over the period 1980-1990, study the socioeconomic and demographic determinants for four different types of crime: murder, theft, robbery and fraud. Furthermore, to take into account specific regional characteristics related to the presence of organized crime, they exploit the panel structure with unobservable individual components. The regressors considered in their estimate are: probability of being apprehended, severity of the final judgment, regional real consumption per capita, regional rate of unemployment, number of employed in the service sector, social security benefit, average monthly salary, public works, percentage of all males in the age group 14 to 29 and the number of students who have completed secondary and high schools relative to the population. Their main findings suggest that the probability of punishment is relatively more effective than severity of punishment; and that proxies for the relative returns from legal and illegal activities, such as unemployment rate, the value of public works started by government and the proportion of people employed in the service sector, have a significant effect on crime. In their following contribution, Marselli and Vannini (2000) specifically focus their attention on the role of unemployment in determining crime rates. Their analysis does not allow to establish a robust correlation between unemployment rate and crime, even if it "overall supports the hypothesis that unemployment significantly affects crime rate and thus this suggests to pay attention in implementing control police not only to the expected punishment but also to employment possibilities" (p. 296).

2.2 Motivation

The literature presented suggests that education has a twofold effect on crime: it raises skills and abilities and then improves labour market perspectives thus implying a higher opportunity cost of crime and it has a non-market effect that affects the preferences of individuals. However, none of these studies recognises the possibility that education may increase the returns to certain crimes such as frauds. This sentence is implicitly recognised by Lochner (2004), that claims that while unskilled crimes should be negatively correlated with education, this does not necessarily hold for white collar crimes.

To this purpose, consider a representative economy where individuals decide how to allocate their available time each period to education, crime and work in the legal sector with the aim of maximizing their expected disposable income. Individuals are endowed with an initial level of ability h_0, that represents the level of ability acquired during primary school, and learning ability ε. Denote level of ability at time t by $h_t = h(s_{t-1}, \varepsilon)$, the time spent for education as s_t, and time for committing crime as d_t. Total time each period is normalized to 1; hence, time spent working is $l_t = 1 - s_t - d_t$. Each period individuals earn $w_t h_t l_t$ from legal work, where w_t is the average wage rate. If an individual is engaged in criminal activity, she obtains with probability $(1 - \pi_a)$ a return $R(d_t, h_t)$, which depends on the time devoted to crime activities and individual ability.

Returns from crime are assumed to be strictly increasing and concave in d_t and non decreasing in h_t. With probability π_a a criminal is apprehended and punished. An apprehended criminal goes to jail for the entire period in which she is apprehended[2] and received a punishment $P(d_t)$, which is increasing in d_t.

[2] Our analysis is general and does not change if we allow the apprehended criminal to go in jail for a fraction of her disposable time.

The individual's maximization problem is:

$$\max_{s,d} \sum_{t=1}^{T} \beta^t y_t \tag{1}$$

under time constraints $l_t + s_t + d_t = 1$, $l_t, s_t, d_t \geq 0$, where β is the intertemporal discount rate.

Total disposable income is defined as

$$y_t = \begin{cases} w_t h_t l_t + R(d_t, h_t) & \text{with prob} \quad (1-\pi_a) \\ \overline{c} - P(d_t) & \text{with prob} \quad \pi_a \end{cases} \tag{2}$$

where $\overline{c}$ the level of consumption of a convicted criminal. For simplicity we consider the case in which $T = 2$ and $s_2 = 0$.[3] Thus, we can rewrite the maximization problem as:

$$\max_{s_1,d_1,d_2} \left\{ \begin{array}{l} (1-\pi_a)w_1h_1(1-s_1-d_1) + (1-\pi_a)R(d_1,h_1) + \pi_a\overline{c} - \pi_a P(d_1) \\ +\beta\left[(1-\pi_a)w_2h_2(1-d_2) + (1-\pi_a)R(d_2,h_2) + \pi_a\overline{c} - \pi_a P(d_2)\right] \end{array} \right\} \tag{3}$$

The first order conditions with respect to s_1, $d_1$and d_2 for an interior solution are:

$$d_{i,1} : (1-\pi_a)w_1h_1 = (1-\pi_a)R'(d_1,h_1) - \pi_a P'(d_1) \tag{4}$$

$$s_{i,1} : (1-\pi_a)w_1h_1 = \beta\left[(1-\pi_a)w_2h_2'(1-d_{i,2}) + (1-\pi_a)R'(d_2,h_2)h_2'\right] \tag{5}$$

$$d_{i,2} : (1-\pi_a)w_2h_2 = (1-\pi_a)R'(d_2,h_2) - \pi_a P'(d_2) \tag{6}$$

The simple model above provides us with interesting empirical implications. First, the higher the fraction of time spent committing crimes in $t = 1$, the higher the fraction of time committing crime in $t = 2$. In fact, a higher level of time spent committing crimes in the first period implies less time dedicated to schooling in the same period. This reduces expected returns in the legal sector in $t = 2$, which, in turn, lowers the opportunity cost of commiting crime in $t = 2$. In this sense, our model predicts criminal inertia, which is the first testable hypothesis of our model. This is in line with the empirical study by Fainzylber *et al.* (2002).

Second, our model allows us to study the effect of wage on crime. However, we need to distinguish between first and second period. An increase of wage in the first

[3] Individuals in the second period do not invest in education due that they live the last period of their life.

period implies that education and crime are more costly in terms of foregone income, by choosing less education individuals will have a lower wage in the second period, then the level of crime will be higher being the opportunity cost of crime lower. On the other hand, a higher wage in the second period reduces unambiguously time spent in crime in $t = 2$.

Third, prevention and effective law enforcement policies reduce crime rate: an increase in the probability of apprehension (π_a) corresponds to a reduction of the expected return from illegal activities, and this leads to a reduction in the level of time spent in committing crime.

Apart from the predictions which we just addressed, our model clarifies also the conditions under which education may reduce criminal activity. Indeed, equation (5) allows us to study the costs and returns of education. On the one hand, a higher education implies higher returns both from work and crime; on the other hand, an individual with a high level of education if apprehended and convicted experiences a greater earnings loss. Thus, more time invested in education in the first period is associated with higher expected returns in the legal sector in the second period, this corresponds to a higher opportunity cost of crime and then to a lower level of crime in the second period. However, while equation (4) suggests that education increases individual returns from work, thereby increasing the opportunity costs of crime, equation (6) shows that education affects the net marginal returns to crime.

Notice that as long as education increases the marginal returns to work more than crime ($w_t h_t^{'}(s_{t-1}) > R_{h_t} h_t^{'}(s_{t-1})$), crime is decreasing in education, as suggested by Usher (1997), Lochner (2004) and Lochner and Moretti (2004). However, while education is likely to have a small effect on the returns of unskilled property crimes, this need not to be true for white collar crimes. In our opinion, it is possible that for skilled crimes the inverse applies: education may increase marginal returns to crime more than to legal work ($w_t h_t^{'}(s_{t-1}) < R_{h_t} h_t^{'}(s_{t-1})$). In other words, because the difference between marginal returns from legal activity and marginal returns from crime cannot be unambiguosly signed, the relationship between education and crime can present non-linearities. Such non-linearity is conditional on the level of education itself: when education is low, education reduces criminal activity; when education is high, it may be the case that education increases criminal activity for certain crimes (e.g. fraud). The possibility of non-linearities in the relationship between education and crime together with the fact that very little attention has been devoted to the empirial analysis of the relationship between crime and education motivate our empirical study.

3 Data and empirical strategy

3.1 Data

In this section we provide an extensive discussion about the data used in our empirical analysis. Our panel dataset comprises annuals observations for the 20 Italian regions (NUTS2) [4] over the period 1980 to1995. Table A1 describes all variables used in our estimations. Crime data are taken from CRENoS (Centre for North South Economic Research). In particular, we consider three different crime rates (*Crime*): property crime rate, theft rate and total crime rate. Crime rate is obtained normalizing the total number of crime in each category by resident population in each region, population is taken from ISTAT. The explanatory variables are separated into three groups: education, deterrence variables and socioeconomic variables.

We use different measures of education: percentage of population with high school diploma (*High School*), percentage of population with university degree (*University*) and average years of schooling of the population (Avg years School). Population with high school diploma and population with university degree are taken from Quarterly Labour Forces by ISTAT, while the variable average years of schooling is our own calculation on the basis of the Quarterly Labour Forces. The variable *Secondary*, taken from ISTAT, is defined as the ratio between students enrolled in high-school and university and the total population of each region.

We use four deterrence variables: quickness of conclusion of Istruttoria and Primo Grado (*Istruttoria*) and Appello and Cassazione (*Appello*) measured as inverse of average length of judicial process respectively for Istruttoria and Primo Grado and for Appello and Cassazione, [5] percentage of crimes committed by unknown offenders (*Un-*

[4] Valle d'Aosta has been aggregated to Piemonte due to its small dimensions.

[5] Istruttoria and Primo grado represents the first stage of the entire judicial proceeding. Appello and Cassazione represent respectively the second and the third stage of judicial proceeding.

known) as proxy of the probability of apprehension, measured as the ratio of crimes committed by unknown offenders to all recorded crimes in each category and, finally, the extent of police forces (*Police*), normalized by population.[6] All these four variables are taken from CRENoS crime database. Deterrence variables (i.e. clear-up rate, efficiency of the judicial system and police force) determine the expected returns from crime activity.

We complete the data set by adding a set of socioeconomic variables taken from ISTAT. In particular, the rate of employment (*Employment*) defined as the ratio between the number of employed and the population for each region, he average regional wage (*Wage*) at 1990 constant price, Gdp per capita (*Gdp*) at 1990 constant prices and the growth rate in the Gdp at 1990 constant prices (*Growth*). Employment rate and wage capture the labour market opportunities of the region that, as shown in out theoretical model, may represent two important determinants of crime behaviour. Following the analysis made by Ehrlich (1973) we can consider the GDP per capita and the growth rate of the GDP as proxies for the general level of prosperity in the provinces, then as indicators of illegal income opportunities.

Finally, we consider dynamics in delinquency. In fact, past experience in criminal activity affects in several ways the decision to commit a crime (Sah, 1991; Glaeser *et al.*, 1996; Fajnzylber *et al.*, 2002); in other words, higher crime today is associated with higher crime tomorrow (i.e. persistence over time). Criminals can learn-by-doing and acquire an adequate criminal know-how level; this acquisition, in turn, makes the costs of carrying out criminal acts to decrease over time (Case and Katz, 1991). Convicted criminals have fewer opportunities of legal employment and a lower expected wage (Grogger, 1995). These arguments strongly suggest the possibility of criminal hysteresis or inertia.

In Table 1 we present the behaviour of property crimes per 1.000 inhabitants over the period 1980-1995. In order to facilitate the analysis we have aggregated the twenty Italian regions in northern regions, central regions and southern regions. As it clearly appears from Table 1 property crimes and thefts present an upward trend over the considered period. From 1980 to 1990 they have substantially remained stable, but

[6] Italian police force is composed by Carabinieri, Polizia and Guardia di Finanza. All these three bodies of the Italian police force operate at national level.

after 1990 they peaked again. The trend is similar for all Italian regions. Total crimes per capita present a behaviour similar to that of property crimes. Both for total crimes and property crimes the trend is overall increasing in Italy, even if over the period 1986-1990 southern regions have experienced a reduction, while central and northern regions have experienced a slight increase. Starting from 1991 a sharp increase has affected all Italian regions.

Table 1 also presents the rate of education (high school and university) defined as the number of high school and university students normalized by regional population. Percentage of population with high school diploma considerably increased over the considered period, and the same is true also for percentage of population with university degree.

3.2 The empirical procedure

Our empirical models are as follows:

$$Crime_{it} = \beta_1 Crime_{it-1} + \beta_2 Edu_{it} + \beta_3 Edu_{it}^2 + \sum_{j=1}^{s} \beta_j X_{jit} + \eta_i + \eta_t + \varepsilon_{it} \qquad (7)$$

where the subscripts i and t represent region and time period, respectively; η_i is a region fixed effect, η_t is a time fixed effect, $Crime_{it}$ is the number of crimes per region residents, Edu_{it} is the education variable, X_{it} is the set of explanatory variables defined in the previous section and ε_{it} is the error term.

From an econometric perspective, there are several estimation problems that may arise in estimating these empirical models. First, using a panel dataset it is well-known that OLS coefficients are biased both in the case that unobservable province-specific effects (η_i) are statistically significant, and in the case that regressors and these effects are correlated. Second, our theoretical model suggests that there exists a significant relationship between crime rates in t and $t-1$; for this reason, we include the lagged dependent variable ($Crime_{i,t-1}$) in our model. In such a framework, OLS results in inconsistent estimates since $Crime_{i,t-1}$ and η_i are necessarily correlated, even if the idiosyncratic component of the error term is serially uncorrelated. An obvious solution to these problems is to eliminate the term η_i by taking first-differences. However, OLS still does not consistently estimate the parameters of interest because first-differencing introduces correlation between the lagged dependent variable and differenced error terms, i.e. $Crime_{it-1}$ and ε_{it} are correlated trough the terms $Crime_{it-1}$ and ε_{it-1}. The alternative to first differences transformation is the within transformation; however, and although controlling for fixed effects, the within transformation leads to consistent estimates only under the hypothesis of strictly exogenous regressors. Third, it is unlikely that explanatory variables are strictly exogenous; the relationship between crime rates and their determinants is often characterized by a two-way causality. Fourth, it is very likely that crime data may be subject to measurement errors, which induce biases in the estimates.

The econometric problems presented above suggest the use of an instrumental variables procedure applied to a dynamic model of panel data. This paper therefore employs the GMM estimator that uses the dynamic properties of the data to generate proper instrumental variables (Arellano and Bond, 1991; Arellano and Bover, 1995). The GMM technique allows to control for (weak) endogeneity by using the instrumental variables, which consist of appropriate lagged values of the explanatory variables. To deal with the fact that measurement errors are likely to be determined not only by random errors but by specific and persistent characteristics of each region we employ the GMM-system estimator (Arellano and Bover, 1995; Blundell and Bond, 1998) which joins into a single system the regression equation in both differences and levels.

The consistency of the parameters obtained by means of the GMM estimator depends crucially on the validity of the instruments. We therefore consider two specification tests suggested by Arellano and Bond (1991) and Arellano and Bover (1995). The first test is the Sargan test of overidentifying restrictions, which tests the null hypothesis of overall validity of the instruments used. Failure to reject this null hypothesis gives support to the choice of the instruments. We also report the test for serial correlation of the error term, which tests the null hypothesis that the differenced error term is first and second order serially correlated. Failure to reject the null hypothesis of no second-order serial correlation implies that the original error term is serially uncorrelated and the moment conditions are correctly specified. Appendix A presents the econometric methodology employed in this paper in details.

4 Empirical results

In Table 2, Columns (a) present the results obtained when we use the percentage of population with high school diploma and the percentage of population with university degree as measure for education, in columns (b) we report the results when the average years of schooling of population is used. We report the Hausman test (random model vs fixed effects models), the Wald test of joint significance of the time and individual dummies, the R^2 and the R^2 adjusted for the degrees of freedom - both statistics show that our model adapt well to the set of observations we are using.

It is worth to notice the average years of schooling, the percentage of people having a high school diploma, the Gdp per capita, the percentage of crime committed by unknown offenders and especially enrolment rates have significant coefficients with the expected signs for both property crimes, and total crimes. The wage rate is negatively correlated to crime rate, even if it is significant only in models when we use average years of schooling as a proxy for education. Notice, also, that the percentage of people having a university degree is insignificant, and it is wrongly signed in two models out of three. We come extensively back to this point later.

Table 3 reports the results we obtain when we use the dynamic frameworks. In these model all variables, including lagged crime rates, are treated as endogenous. We report four test statistics: (i) the Wald test of joint significance of the time dummies; (ii) Sargan test of overidentifying restrictions; (iii) and (iv) tests for first and second order serial correlation.

Notice that all coefficients drop, which suggest that our previous results are affected by endogeneity; however, the main message of our results remain unchanged: education measured both by the average years of schooling of population and by percentage of population with high school diploma has a negative and significant effect on the three different measures of crime rates used (property crime rate, theft rate and total crime rate) even after controlling for socio-economic variables and in particular for labour market returns and opportunities. Furthermore, enrolment rate has a negative effect on crime rates, stressing the incapacitation role of schooling. Moreover, GDP per capita is positively and significantly correlated to crime rate as suggested by the literature, in fact following the analysis made by Erlich (1973) we can consider the GDP per capita as a proxy for the general level of prosperity in the region, then as an indicator of illegal income opportunities. Third, wage rate has a negative and significant effect on crime rates, even if its significance depends on the measure of education used in the estimates. Fourth, with the exceptions of total crime rate, the percentage of crime committed by unknown offenders is positively and significantly correlated to crime rate. This variables allows us to capture the effect of deterrence and law enforcement, a higher level of the percentage of crime committed by unknown offenders is associated to a higher expected returns from crime. Finally, the dynamic specification we are using allows us to show that crime display persistence over time. The coefficient on the lagged dependent variables is between 0.3 and 0.5, confirming that crime rates show a sizeable degree of inertia. In particular, total crime rate show a bigger degree of inertia compared to property crime rate and thefts rate.

In order to test the possible non linearity between education and crime rates, we have added the squared values of our indexes of education to all models we have proposed in Table 3. Results are reported in Table 4. As a general results, the evidence suggests that in all model where our proxies of education are statistically significant, their squared values are significant as well. More importantly, in model where total crimes are used as a proxy for crime rates, the university degree is now statistically significant, and the relationship results to be non-linear. All the evidence suggests that the relationships between crime and education is U-shaped.

In order to investigate the sources of such non-linearity, we have studied the relationship between fraud and education. So far, in our analysis we have considered the effects of education on unskilled crime and the empirical investigation gives support to the negative correlation between education and crime rates. But, as presented in our theoretical model and as stresses also by Lochner (2004) the correlation between schooling and white collar crime rate could be not statistically significant or even positive. Results are reported in Table 4. As expected the relationship between fraud and education is positive, and statistically significant when the proxy of education we employ is university. In both the static and the dynamic settings, we found a positive and significant correlation between percentage of population with university degree and fraud rate. The other variables used, apart from the percentage of fraud committed by unknown offenders that is positively and significantly correlated to fraud rate, are not significant. This relationship is consistent with our theoretical model of crime and education as long as white collar crimes (fraud) offer a higher return to education than unskilled crime.

5 Conclusions

In this paper, we study whether education exerts a non-market effect on crime rate. For our empirical analysis we use a panel data set for the twenty Italian regions over the period 1980 to 1995. We employ an instrumental variables approach to investigate this relationship and we consider different measures of education to test the robustness of our results. Our results show that education, independently on the measure used, is negatively and significanlty related to crime rate even after controlling for labour market opportunities (employment rate and wage rate). Furthermore, crime rate display persistence over time and GDP per capita, used to proxy illegal income opportunities, is positively related to crime rate.

We extend our analysis to study whether the relationship between education and crime may be characterized by non linearity. Our results suggest, as expected, that the effects of education on crime is non-linear, and that past criminal experience matters. More specifically, we show that the non-linearity is due to a white collar effect. We perform a number of exercises in order to deeply study the robustness of our results: first, we study the correlation between education and, separately, property crime rate, theft rate and total crime rate, always controlling for labour market opportunities (wage rate and employment rate), GDP per capita, growth rate of the GDP and a set of deterrence variables (police forces, quickness of conclusion of judicial proceedings and percentage of crime committed by unknown offenders). Also, we adopt different measures of education (the percentage of population with high school diploma, percentage of population with university degree and average years of schooling of the population). Finally, we consider inertial effects of crime (i.e. persistence over time) and control for endogeneity: we employ an instrumental variable approach for panel data using the GMM-system estimator that allows us to control for the joint endogeneity between crime rate and the explanatory variables and, also, for measurement errors in crime rates.

References

Arellano, M., and Bond, S., 1991. Some Test of Specifications for Panel Data: Monte Carlo Evidence and an Application to Employment Equation. Review of Economic Studies 58, 277-297.

Arellano, M., and Bover, O., 1995. Another Look at the Instrumental Variable Estimation of Error-Components Models. Journal of Econometrics 68, 29-51.

Becker, G.S., 1968. Crime and Punishment: An Economic Approach. Journal of Political Economy 76, 169-217.

Becker, G.S., 1993. Nobel Lecture: The Economic Way at Looking at Behavior. Journal of Political Economy 101, 385-409.

Block, M., Heineke, J., 1975. A Labor Theoretic Analysis of Criminal Choice. American Economic Review 65, 314-325.

Blundell, R., and Bond, S., 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. Journal of Econometrics 87, 115-143.

Burdett, K., Lagos, R., and Wright, R., 2003. Inequality, Unemployment and Crime. Forthcoming in American Economic Review.

Chiu, W.H., Madden, P., 1998. Burglary and income inequality. Journal of Public Economics 69, 123-141.

Ehrlich, I., 1975. On the Relation between Education and Crime, in Juster, F.T. (Ed.), Education, Income and Human Behavior. New York: McGraw-Hill, pp. 313-337.

Ehrlich, I., 1973. Participation in Illegitimate Activities: A Theoretical and Empirical Investigation. Journal of Political Economy 81, 521-565.

Fajnzylber, P., Lederman, D., and Loayza, N., 2002a. What Causes Violent Crime?. European Economic Review 46, 1323-1357.

Fajnzylber, P., Lederman, D., and Loayza, N., 2002b. Inequality and Violent Crime. Journal of Law and Economics 45, 1-40.

Freeman, R.B., 1999. The Economics of Crime, in Ashenfelter, O.C., Card D. (Eds.), Handbook of Labor Economics. North-Holland, Amsterdam, pp. 3529-3571.

Freeman, R.B., 1991. Crime and the Employment of Disadvantaged Youths. NBER Working Paper no. 3875.

Glaeser, E.L., 1999. An Overview of Crime and Punishment. Mimeo.

Gould, E.D., Weinberg, B.A., and Mustard, D.B., 2002. Crime Rates Local Labor Opportunities in the United States: 1979-1997. Review of Economic Studies 84, 45-61.

Grogger, J., 1998. Market Wages and Youth Crime. Journal of Labor Economics 16, 756-91.

Levitt, S.D., Venkatesh, S.A., 1998. An economic analysis of a drug-selling gang's finance. Quarterly Journal of Economics 115, 755-789.

Lochner, L., 1999. Education, Work and Crime: Theory and Evidence. Working Paper No. 465, Center for Economic Research (RCER) - University of Rochester.

Lochner, L., 2004.Education, Work and Crime: A Human Capital Approach. International Economic Review 45, 811-843.

Lochner, L., and Moretti, E., 2004. The effect of education on crime: evidence from prison inmates, arrests, and self-reports. American Economic Review 94, 155-189.

Marselli, R., and Vannini, M., 1997. Estimating a Crime Equation in the Presence of Organized Crime: Evidence from Italy. International Review of Law and Economics 17, 89-113.

Raphael, S., Winter-Ebmer, R., 2001. Identifying the effect of unemployment on crime. Journal of Law and Economics 44, 259-283.

Scorcu, A.E., Cellini, R., 1998. Economic Activity and Crime in the Long Run: An Empirical Investigation on Aggregate Data from Italy, 1951–1994. International Review of Law and Economics 18, 279–292.

Usher, D., 1997. Education as Deterrent to Crime. Canadian Journal of Economics 30, 367-84.

Tauchen, H., Witte, A.D. and Griesinger, H. (1994). "Criminal deterrence: revisiting the issue with a birth cohort". Review of Economics and Statistics, 76, 399-412.

Wilson, J.Q., Herrnstein, R., 1985. Crime and Human Nature. Simon and Schuster, New York.

Witte, A.D., Tauchen, H., 1994. Work and Crime: An Exploration Using Panel Data. Public Finance 49, 155-167.

Appendix A (OK)

Generalyzed Method of Moments Methodology[7]

We use a dynamic model to test the relationship between unemployment and crime rates. Our model is:

$$y^*_{i,t} = \alpha y^*_{i,t-1} + \beta_1' UNEM_{i,t} + \beta_2' X_{i,t} + \eta_i + \vartheta_{i,t} \tag{8}$$

where y^* is the "true" crime (property, theft and total) rate, $UNEM$ is the unemployment variable, X is the set of explanatory variables other then unemployment, ϑ is the error term and η_i is a region-specific unobserved effect, that may be correlated to the explanatory variables. The subscripts i and t denote region and time period, respectively. As discussed in the paper, crime data are likely to be affected by measurement errors. In what follows we assume that measurement errors are determined not only by random errors but by specific and persistent characteristics of each region. Then, we define the measurement error as:

$$y_{i,t} = y^*_{i,t} + \upsilon_{i,t} + \lambda_i \text{ and } \upsilon_{i,t} \text{ is i.i.d.} \tag{9}$$

where y is crime rate recorded and observed and λ is a region-specific effect. Then we can rewrite equation (8) as

$$y_{i,t} = \alpha y_{i,t-1} + \beta_1' UNEM_{i,t} + \beta_2' X_{i,t} + \omega_i + \varepsilon_{i,t} \tag{10}$$

where

$$\omega_i = \eta_i + (1-\alpha)\lambda_i \text{ and } \varepsilon_{i,t} = \vartheta_{i,t} + \upsilon_{i,t} - \alpha\upsilon_{i,t-1} \tag{11}$$

The model that we test is represented by equation (10). There are several important estimation issues that may arise in the dynamic panel data specification employed in this paper. First, in the presence of region-specific unobserved effect, OLS coefficient will be biased when regressor and η_i are correlated. Second, OLS estimator will result in inconsistent parameter estimates since $y_{i,t-1}$ is correlated by construction with $\upsilon_{i,t-1}$. Then, consistent estimates required the use of an instrumental variables approach applied to dynamic panel data model. Therefore, we employ the GMM-system estimator

[7] This section is heavily based on the Appendix D of Fajnzylber, Lederman and Loayza (2002).

proposed by Arellano and Bover (1995) and Blundell and Bond (1998) that joins in a single system the regression equation in both differences and levels, using the proper set instrumental variables.

Following Fajnzylber, Lederman and Loayza (2002) we present each part of the system separately, although estimation is performed using the whole system jointly. First, taking the first differences of equation (10) allows us to eliminate the region-specific effects. Then we can write the regression equation as

$$y_{i,t}-y_{i,t-1} = \alpha(y_{i,t-1}-y_{i,t-2})+\beta_1'(UNEM_{i,t}-UNEM_{i,t-1})+\beta_2'(X_{i,t}-X_{i,t-1})+(\varepsilon_{i,t}-\varepsilon_{i,t-1}) \tag{12}$$

On the other hand, first difference transformation introduce correlation between the new error term, $(\varepsilon_{i,t} - \varepsilon_{i,t-1})$, and the differenced lagged dependent variable, $(y_{i,t-1} - y_{i,t-2})$, then we need to use instruments to deal with this problem. Moreover, the likely endogeneity of the explanatory variables, $UNEM$ and X, and the random measurement error of the lagged crime rate require the use of instruments. We assume weak exogeneity and no serial correlation in the error term, then the moment conditions are:

$$E\left[y_{i,t-s} \text{ x } (\varepsilon_{i,t} - \varepsilon_{i,t-1})\right] = 0 \quad \text{for} \quad s \geqslant 3 \tag{13}$$

and

$$E\left[X_{i,t-s} \text{ x } (\varepsilon_{i,t} - \varepsilon_{i,t-1})\right] = 0 \quad \text{for} \quad s \geqslant 2 \tag{14}$$

The GMM estimator based on the moment conditions represented by equations (13) and (14) is called the differences estimator. This estimator, although asymptotically consistent, has low asymptotic precision and large biases in small samples (Blundell and Bond (1998)), then to overcome this problem we need to complement it with the regression equation in levels. The regression in levels does not allow the elimination of the region-specific effect, then we need to control for using instrumental variables. The instruments used are the lagged differences of the corresponding variables. It is worth to notice that even if the levels of the right-hand-side variables and the country-specific effect may be correlated, the differences of these variables and the country-specific effect are uncorrelated. This assumption results from the following stationarity property:

$$E\left[y_{i,t+p} \text{ x } \eta_i\right] = E\left[y_{i,t+q} \text{ x } \eta_i\right] \text{ and } E\left[X_{i,t+p} \text{ x } \eta_i\right] = E\left[X_{i,t+q} \text{ x } \eta_i\right] \tag{15}$$

for all p and q.

Moreover, we need to consider the additional moment conditions for the regression in levels that are give by:

$$E\left[(y_{i,t-s} - y_{i,t-s-1}) \text{ x } (\eta_i + \varepsilon_{i,t})\right] = 0 \quad \text{for} \quad s = 2 \tag{16}$$

and

$$E\left[(X_{i,t-s} - X_{i,t-s-1}) \text{ x } (\eta_i + \varepsilon_{i,t})\right] = 0 \quad \text{for} \quad s = 1 \tag{17}$$

The consistency of the parameters obtained by means of the GMM estimator depends crucially on the validity of the instruments. We therefore consider two specification tests suggested by Arellano and Bond (1991) and Arellano and Bover (1995). The first test is the Sargan test of overidentifying restrictions, which tests the null hypothesis of overall validity of the instruments used. Failure to reject this null hypothesis gives support to the choice of the instruments. We also consider the test for serial correlation of the error term, which tests the null hypothesis that the differenced error term is first and second order serially correlated. Failure to reject the null hypothesis of no second-order serial correlation implies that the original error term is serially uncorrelated and the moment conditions are correctly specified.

Table 1
Descriptive Statistics

Period	North	Center	South	Italy
		High School Diploma		
1981-85	11.4%	13.7%	10.7%	11.6%
1986-90	15.1%	17.8%	13.8%	15.1%
1991-95	19.6%	21.8%	16.7%	18.9%
		University Degree		
1981-85	2.35%	3.34%	2.47%	2.58%
1986-90	3.05%	4.10%	3.05%	3.25%
1991-95	4.01%	5.20%	3.71%	4.13%
		Police Force		
1981-85	3.07	5.00	3.19	3.48
1986-90	3.66	6.15	3.86	4.21
1991-95	4.02	6.77	4.64	4.77
		Total Crimes		
1981-85	34.38	40.13	33.74	35.26
1986-90	36.89	51.09	30.75	37.36
1991-95	47.25	72.27	38.22	48.75
		Property Crimes		
1981-85	24.78	30.36	24.13	25.62
1986-90	26.20	35.48	21.44	26.23
1991-95	36.79	51.27	29.27	36.83
		Thefts		
1981-85	22.69	28.33	21.29	23.28
1986-90	23.20	32.27	17.76	22.94
1991-95	30.91	42.84	23.87	30.63

Table 2
Estimation Results: Fixed Effects

	PROPERTY CRIMES		THEFTS		TOTAL CRIMES	
	(a)	(b)	(a)	(b)	(a)	(b)
High School	-0.126		-0.088		-0.129	
	(-2.54)**		(-1.93)**		(-1.73)*	
University	0.090		-0.021		0.316	
	(0.293)		(-0.083)		(0.684)	
Avg School Years		-0.011		-0.010		-0.010
		(-2.15)**		(-2.09)**		(-1.63)*
Enrolment	-0.856	-0.725	-0.673	-0.539	-0.989	-0.948
	(-2.60)**	(-2.32)**	(-2.44)**	(-2.19)**	(-2.49)**	(-2.16)**
Employment	0.023	0.022	0.019	0.020	0.025	0.028
	(0.504)	(0.435)	(0.522)	(0.505)	(0.347)	(0.372)
Gdp per capita	0.002	0.002	0.002	0.002	0.003	0.002
	(2.86)***	(2.35)**	(3.01)***	(2.60)**	(2.92)***	(2.56)**
Growth rate	-0.001	0.001	-0.001	0.001	-0.002	-0.001
	(-0.110)	(0.137)	(-0.109)	(0.086)	(-0.149)	(-0.053)
Wage	-0.002	-0.002	-0.001	-0.001	-0.002	-0.003
	(-1.55)	(-1.88)*	(-1.42)	(-1.66)*	(-1.47)	(-1.88)*
Unknown	0.040	0.039	0.026	0.030	-0.008	-0.007
	(1.98)**	(1.72)*	(1.29)	(1.28)	(-0.392)	(-0.348)
Police	1.709	1.708	1.362	1.374	2.99	3.02
	(1.38)	(1.29)	(1.30)	(1.25)	(1.73)*	(1.63)*
Primo Grado	0.282	0.249	0.243	0.208	0.475	0.478
	(0.929)	(0.741)	(1.01)	(0.783)	(1.13)	(1.03)
Appello	-1.062	-1.024	-0.871	-0.856	-1.030	-0.994
	(-1.60)*	(-1.66)*	(-1.67)*	(-1.79)*	(-1.17)	(-1.22)
Hausman Test	28.46		35.67		35.02	
Wald (time)	193.6	1116	688.9	695.5	106.5	1315
R^2	0.859	0.859	0.861	0.862	0.846	0.845
Adjusted R^2	0.835		0.838		0.820	

Notes:
Standard errors are reported in parentheses. Standard errors are robust to heteroscedasticity and autocorrelation (Arellano, 1987). ***, ** and * indicate coefficient significant at the 1%, 5% and 10% levels, respectively.

Table 3
Estimation Results: GMM

	Property Crimes		Thefts		Total Crimes	
	(a)	(b)	(a)	(b)	(a)	(b)
$Crime_{-1}$	0.354	0.364	0.300	0.318	0.553	0.531
	(3.04)***	(3.04)***	(3.01)***	(3.30)***	(4.53)***	(3.89)***
High School	-0.082		-0.069		-0-045	
	(-2.58)**		(-2.38)**		(-1.11)	
University	-0.058		-0.076		0.009	
	(-0.239)		(-0.332)		(-0.032)	
Avg School years		-0.010		-0.010		-0.009
		(-2.72)***		(-2.07)**		(-2.05)**
Enrolment	-0.527	-0.309	-0.425	-0.277	-0.339	-0.207
	(-1.88)*	(-1.48)	(-1.74)*	(-1.56)	(-1.74)*	(-1.27)
Employment	0.003	-0.009	-0.005	-0.010	0.028	0.035
	(0.084)	(-0.286)	(-0.140)	(-0.225)	(0.633)	(0.794)
Gdp per capita	0.001	0.001	0.001	0.001	0.001	0.002
	(2.17)**	(2.00)**	(2.70)***	(2.14)**	(1.90)*	(1.88)*
Growth rate	0.001	0.001	-0.001	0.001	0.002	-0.002
	(0.070)	(0.130)	(-0.102)	(0.047)	(0.195)	(-0.190)
Wage	-0.001	-0.002	-0.001	-0.001	-0.002	-0.003
	(-1.38)	(-2.41)**	(-1.13)	(-1.65)*	(-2.15)**	(-2.89)***
Unknown	0.039	0.043	0.029	0.035	0.008	0.011
	(2.55)**	(2.47)**	(1.72)*	(1.60)*	(0.809)	(1.02)
Police	0.866	0.949	0.705	0.920	1.077	0.974
	(0.971)	(1.01)	(0.943)	(1.05)	(0.983)	(0.892)
Primo Grado	0.209	0.150	0.228	0.145	0.352	0.377
	(0.874)	(0.572)	(1.11)	(0.714)	(1.22)	(1.18)
Appello	-0.526	-0.761	-0.563	-0.495	-0.493	-0.498
	(-1.08)	(-1.32)	(-1.28)	(-1.14)	(-0.864)	(-0.810)
Specification tests						
Sargan Test	260.6	259.7	257.9	255.1	239.7	242.0
Serial Correlation						
First Order	-2.17*	-2.18*	-2.25*	-2.27*	-2.20*	-2.20*
Second Order	-0.63	-0.52	-0.18	0.02	-0.87	0.87

Notes: First Order and Second Order Test are test statistics for first and second order autocorrelations in residuals, respectively, distributed as standard normal N(0,1) under the null of no serial correlation. Sargan test is a test of overidentifying restrictions, distributed as chi-square under the null of instrument validity. Standard errors are reported in parentheses. Standard errors are robust to heteroscedasticity and autocorrelation (Arellano, 1987). ***, ** and * indicate coefficient significant at the 1%, 5% and 10% levels, respectively. All variables are instrumented using lag t-2.

Table 4
Non Linear Effects of Education (GMM estimation)

	Property Crimes		Thefts		Total Crimes	
	(a)	(b)	(a)	(b)	(a)	(b)
Crime$_{-1}$	0.257	0.303	0.199	0.280	0.413	0.463
	(2.81)***	(3.40)***	(2.38)**	(3.58)***	(3.78)***	(4.75)***
High School	-0.532		-0.541		-0.468	
	(-1.99)**		(-1.94)*		(-1.80)*	
High School^2	1.36		1.40		1.35	
	(1.85)*		(1.87)*		(1.84)*	
University	-0.465		-0.254		-1.07	
	(-1.23)		(-0.636)		(-2.55)**	
University^2	2.89		-0.075		10.31	
	(0.647)		(-0.014)		(2.67)***	
Avg School years		-0.039		-0.033		-0.056
		(-3.20)***		(-3.28)***		(-4.16)***
Avg School years^2		0.003		0.002		0.004
		(2.39)**		(2.37)**		(4.17)***
Enrolment	-0.203	-0.108	-0.145	-0.107	0.044	0.080
	(-0.819)	(-0.519)	(-0.627)	(-0.600)	(0.261)	(0.533)
Employment	0.003	-0.003	0.009	0.003	0.022	0.080
	(0.078)	(-0.088)	(0.194)	(0.073)	(0.555)	(2.46)**
Gdp per capita	0.001	-4.9E-06	0.001	0.0001	0.001	-0.001
	(2.09)**	(-0.004)	(2.45)**	(0.173)	(1.97)**	(-1.12)
Growth rate	0.002	0.015	-0.002	0.012	0.002	0.019
	(0.198)	(1.65)*	(-0.248)	(1.42)	(0.204)	(1.74)*
Wage	-0.001	-0.001	-0.001	-0.001	-0.002	-0.001
	(-0.900)	(-0.558)	(-0.392)	(-0.292)	(-1.52)	(0.195)
Unknown	0.043	0.047	0.036	0.041	0.014	0.017
	(3.16)***	(2.69)***	(1.85)*	(1.99)**	(1.72)*	(2.01)**
Police	0.844	0.906	0.611	0.586	1.35	1.85
	(0.970)	(1.05)	(0.982)	(0.876)	(1.18)	(1.34)
Primo Grado	0.115	0.304	0.149	0.192	0.192	0.498
	(0.588)	(1.16)	(0.853)	(0.942)	(0.754)	(1.67)*
Appello	-0.714	-1.20	-0.601	-0.827	-0.854	-0.589
	(-1.48)	(-2.11)**	(-1.38)	(-1.69)*	(-1.52)	(-1.13)
Specification tests						
Sargan Test	263.1	256.9	260.3	252.5	254.0	243.3
Serial Correlation						
First Order	-2.12*	-2.05*	-2.25*	-2.20*	-2.25*	-2.14*
Second Order	-0.82	-0.81	-0.44	-0.27	-0.93	-0.93

Notes: see Table 3

Table 5
Estimation Results: Fraud (White Collar Crimes)

	GMM
$Crime_{-1}$	0.455
	(3.25)***
High School	0.0004
	(0.155)
University	0.020
	(2.07)**
Employment	-0.0006
	(-0.255)
Gdp per capita	0.0001
	(0.238)
Growth rate	-0.0003
	(-0.497)
Wage	0.0001
	(0.304)
Unknown	0.0004
	(1.84)*
Police	0.024
	(0.992)
Primo Grado	0.010
	(1.42)
Appello	0.012
	(0.858)
Specification tests	
Sargan Test	254.1
Serial Correlation	
First Order	-2.88**
Second Order	-0.62

Notes: see Table 3