Essays on Economics of crime and Economic Analysis of Criminal Law

Mojtaba Ghasemi

Supervisor: Prof. Francesca Bettio

Thesis submitted for the degree of Doctor of Philosophy in Economics

Department of Economics and Statistics

University of Siena

November 2014

THIS THESIS IS DEDICATED

WITH RESPECT AND AFFECTION TO MY PARENTS

Mohammad and Ma’sumeh

Acknowledgment

I would like to express my special appreciation and thanks to my advisor Professor Dr. Francesca Bettio, who has been a tremendous mentor for me. I would like to thank you, Dr.Bettio, for encouraging my research and for allowing me to grow as a research scientist. Your advice on both research as well as on my career have been priceless. I would also especially like to thank all faculty members whom I learnt so much from these years, as well as my colleagues.

A special thanks to my family. Words cannot express how grateful I am to you, my mother and father, for all of the sacrifices that you've made on my behalf. Your prayer for me was what sustained me thus far. I would also like to thank all of my friends who supported me in writing, and encouraged me to strive towards my goal. Last but not least, I would like to thank all people who contributed to making my PhD career a wonderful and memorable life event in the amazing city of Siena. Beside education, I found the great opportunity to visit and learn many amazing Italian cultural and historical heritages too. I am deeply indebted to all of people who have been involved in both my academic and non-academic adventures in wonderful land of Italy. .

Thesis Abstract

This thesis focuses on certain issues concerning the economics of crime and the economic analysis of criminal law. The first chapter investigates the influence of visceral factors on criminal behavior and the policy implications thereof. To this purpose the chapter exploits concepts from the well-known Becker's model on the one hand and from behavioral economics on the other hand. Chapter 2 attempts an economic analysis of criminal law by applying Becker's social loss function from criminal activities. It addresses two interesting topics. Based on Becker's model, the first part of the chapter formalizes irreconcilabilities between retributive and utilitarian approaches to punishment as two major schools of thoughts in punishment. Although both Utilitarians and Retributivists support the institution of punishment they have their own distributive principles of punishment which make them irreconcilable. The chapter adapts Becker' formal model and diagrams to also shed light on actual irreconcilabilities between and criminal law-making in the reality. The second part of the chapter offers a formal explanation for diversity of criminal law (criminal codes and punishment) in different societies. Finally, chapter 3 applies a Dynamic Panel Data (DPD) model to provide state-of-the-art estimates of the economic model of crime by using panel of North Carolina counties from 1981-1987. This dataset was first used by Cornell and Trumble (1994) and later replicated by Baltagi (2006). The aim of this chapter is to apply GMM-System and GMM-Difference estimators to produce more reliable results.

Contents

Chapter 1

Visceral Factors, Criminal Behavior and Deterrence: Empirical Evidence and Policy Implications

Abstract: This chapter examines how visceral factors influence criminal behavior in the current literature of economics of crime and analyzes optimal and actual criminal law by means of Becker's model. By reviewing 15 empirical studies it investigates the comparative responsiveness of different kinds of crime to deterrence variables and verifies the hypothesis that visceral factors are more influential in violent crimes. The results of this survey confirmed that violent crimes are less responsive to deterrence variables than non-violent crimes. This point can be considered through lower elasticities of crime supply with respect to punishment and probability of apprehension in Becker's model. Optimality in this framework implies that these crimes should be punished leniently since for them, expected punishment does not work as a deterrent. Because visceral factors play a strong role in the perpetration of violent crimes, from a policy point of view, severe punishment may be ineffective and preventive policies addressing the roots of violent, visceral crimes may be a better alternative.

JEL*:* D03, K14
Keywords: visceral factors, deterrence hypothesis, law enforcement

1.1 Introduction

Since Becker (1968), economists have generated a large body of literature on crime. After this seminal paper, some economists tried to extend Becker's theoretical model and others tried to test the "deterrence hypothesis" in the empirical literature. Theoretical predictions of this hypothesis suggest that an increase in the probability of apprehension and severity of punishment has negative effects on crime level. Theoretical models of criminal behavior have been tested in many empirical studies. Specifically, the effects of the probability of apprehension, severity of punishment, as well as benefits and costs of legal and illegal activities on crime have been estimated. The influence of norms, tastes and abilities, corresponding to constitutional and acquired individual characteristics, has in some cases been studied indirectly by including variables like age, race, gender, etc. A variety of equations, specifications and estimation techniques has been used, and the studies have been based on levels of aggregation ranging from countries and states down to municipalities, campuses and individuals.

This chapter addresses a different set of questions. Considering the influence of visceral factors on behavior, violent crimes can be expected to be relatively less responsive to deterrence variables than property crimes. It is assumed that visceral factors have a more influential role in violent crimes than property crimes. This chapter tries to investigate the comparative responsiveness of different types of crimes to changes in the probability of apprehension and severity of punishment in a survey of 15 empirical studies with the following characteristics:

- they include different kinds of violent and property crimes.
- they consider effects of some deterrence variables on crime level.

The results of estimated coefficients or elasticities in the studies confirm that violent crimes (murder, rape …), which are presumably more influenced by visceral factors, are less responsive to deterrence variables than property crimes (burglary, car theft …).

After verifying the more influential role of visceral factors in violent crimes, we applied Becker's model to evaluate some of the current strategies for combating violent crimes.

Serious violent crimes, such as murder and rape, that occur when visceral factors are intensified, inflict high net social damage and respond poorly to deterrence variables. The optimality conditions of Becker's model suggest prescribing severe punishments for high net social damage and mild punishments because of their lower supply elasticity. In actual fact, most criminal law prescribes severe punishment, severity depending on the society's attitude to the social damage of these crimes. Indeed, these criminals, particularly murders and rapists, are punished severely because of the high net social damage they have inflicted on society, although severe prescribed punishments rarely deter potential offenders, because of the strong influence of visceral factors in these crimes.

In the case of violent crimes strongly associated with visceral factors, the message for policy makers is that prescribed punishment is not as deterrent as we imagine and it is better to focus on other crime control strategies. Policy makers should try to understand to more fundamental issues about these crimes, instead of invoking severe punishment to decrease them. In the case of rape, they should ask why there is a demand for rape. Is it because of sexual deprivation? May legalizing prostitution be useful for decreasing rape? Is it related to heavy drinking of alcohol?

The rest of the chapter is organized as follows: the next section briefly presents visceral factors and their influence on behavior. Section 3 concentrates on the empirical literature, ranging from time-series studies to cross-sectional and panel data studies, to investigate the comparative responsiveness of different kinds of crime to deterrence variables. Section 4 enters visceral factors in Becker's model to analyze different strategies and policies for controlling violent crimes. Final and concluding remarks are presented in the last section.

1.2 Influence of visceral factors on behavior and decision theory

Understanding discrepancies between self-interest and behavior has been a major theoretical challenge confronting decision theory since its origin. At sufficient levels of intensity, most visceral factors cause people to behave contrary to their own long-term self-interest, often with full awareness that they are doing so (Lowenstein, 2004). There is surely some truth to this. Consider a man who comes home, finds his wife in bed with another man, pulls out a gun, kills them both and spends the rest of his life in jail. The man might well regret his choice and say that he “lost his reason”, that “emotion took over” and the like. Indeed, this might qualify as a “crime of passion”. Undoubtedly, the man could have thought better. Instead of pulling the trigger, he would have been better off shrugging his shoulders and going to the bar in search of a new partner (Gilboa, 2010).

The defining characteristics of visceral factors are, first, a direct hedonic impact, and second, an influence on the relative desirability of different goods and actions. Hunger, for example, is a sensation that affects the desirability of eating. Anger is also typically unpleasant and increases one’s taste for various types of aggressive actions. Physical pain enhances the attractiveness of pain killers, food, and sex. Although from a purely formal standpoint one could regard visceral factors as inputs into tastes, such an approach would obscure several crucial qualitative differences between visceral factors and tastes:

1. Holding consumption constant, changes in visceral factors have direct hedonic consequences. In this case, visceral factors are similar to consumption, not tastes. The set of preferences that would make me better off is an abstract philosophical question, while whether I would be better off hungry or sated, angry or calm, in pain or pain-free, in each case holding consumption constant, is as obvious as whether I would prefer to consume more or less, holding tastes and visceral factors constant (Lowenstein, 2004).

2. External circumstances (stimulation, deprivation, and such) can predictably affect visceral factors but these transitory circumstances do not imply a permanent change in an individual’s behavioral disposition. On the contrary, changes in preferences are not only caused by slow experience and reflection but these changes also imply a permanent change in behavior (Lowenstein, 2004).

3. While tastes tend to be stable in the short term, they change in the long run, visceral factors typically changing more rapidly than tastes.

4. Finally, tastes and visceral factors have different neurophysiological mechanisms. Tastes, as mentioned above, are more stable in the short term and consist of information stored in memory concerning the relative desirability of different goods and activities [1](Lowenstein, 2004).

We can consider visceral factors in rational choice. It makes good sense to eat when we are hungry, to have sex when feeling amorous, and to take pain killers when in pain. However, it seems that many classic patterns of self-destructive behavior, such as overeating, sexual misconduct, substance abuse and crimes of passion, can be considered examples of an excessive influence of visceral factors on behavior. Intensity level of visceral factors can have different consequences. At low levels of intensity, people seem to be capable of dealing with visceral factors in a relatively optimal fashion. For example, someone who is feeling tired might decide to leave work early or to forgo an evening's entertainment to catch up on sleep. There is nothing obviously self-destructive about these decisions, even though they may not maximize ex post utility in every instance. Increases in the intensity of visceral factors, however, often produce clearly suboptimal patterns of behavior. For example, the momentary discomfort of rising early leads to "sleeping in", a behavioral syndrome with wide-ranging negative consequences. It is at intermediate levels of intensity that one observes classic cases of impulsive behavior and efforts at self-control, e.g. placing the alarm clock on the other side of the bedroom (Schelling 1984). Finally, at even greater levels of intensity, visceral factors can be so powerful as to virtually preclude decision making. No one decides to fall asleep at the wheel, but many people do (Lowenstein, 2004).

In a nutshell, visceral factors affect behavior of individuals as follows. As they intensify, they focus attention and motivation on activities and forms of consumption

[1] Although visceral factors are distinct from tastes in their underlying mechanisms and their effects on well-being and behavior, there are important relationships between them. Tastes are greatly shaped by visceral factors. For example, one's taste for barbecued chicken may well underlie one's visceral reaction to the combined smell of charcoal, fat and tomato sauce. At the same time, the visceral hunger produced by such smells, and the visceral pleasure produced by subsequent consumption, are likely to reinforce one's preexisting taste for barbecued chicken (Lowenstein, 2004)

that are associated with the visceral factor, e.g. hunger draws attention and motivation to food. Non-associated forms of consumption lose their value. At sufficient levels of intensity, individuals will sacrifice almost any quantity of goods not associated with the visceral factor for even a small amount of associated goods, a pattern most dramatically evident in the case of drug addicts. According to Gawin (1991), cocaine addicts report that "virtually all thoughts are focused on cocaine during binges; nourishment, sleep, money, loved ones, responsibility, and survival lose all significance." In economic jargon, the marginal rate of substitution between goods associated with the visceral factor and goods not so-associated becomes infinitesimal (Lowenstein, 2004).

Visceral factors also influence time, collapsing time perception into the present. For instance, a hungry person is likely to make short-sighted trade-offs between immediate and delayed food, even if tomorrow's hunger promises to be as intense as today's. This orientation, however, applies only to goods that are associated with the visceral factor, and only to trade-offs between the present and some other point in time (Lowenstein, 2004).

A *third* form of attention-narrowing involves the self versus others. Intense visceral factors tend to narrow one's focus inwardly, undermining altruism. People who are hungry, in pain, angry, or craving drugs tend to be selfish. This is evident in the behavior of addicts (Lowenstein, 2004).

The influence of visceral factors on behavior, particularly at highly intensified levels, suggests that they have relatively more influence in violent crimes than in property crimes. Violent crimes are therefore presumably less responsive to deterrence variables. Thus the "deterrence hypothesis" is more applicable to property crimes than violent crimes.

1.3 Influence of visceral factors and the criminal behavior: An empirical survey

It seems that visceral factors are more influential in violent than non-violent crimes. This section reviews relevant empirical studies and evaluates this hypothesis in the light of their results. Theoretical models of criminal behavior have been tested in many empirical studies, estimating the effect of the probability of apprehension, severity of prescribed punishment, and the benefits and costs of legal and illegal activities on crime. We only reviewed15 empirical studies which:

- included different kinds of violent and property crimes.
- considered effects of some deterrence variables on crime level.

All were run using aggregate data and different kinds of estimation techniques. The following sections review the studies separately by category: time series, cross sectional and panel data studies.

1.3.1 Time series analysis

These studies concentrate on a specific country, state or city and investigate the effects of deterrence and other covariates on crime level over time. They may consider different kinds of deterrence variables, depending on the availability of data. Some use several measures of apprehension and punishment variables.[2]

Corman and Mocan (2000) used monthly data on crime in New York from1970 to 1996 to study the deterrence hypothesis for five crime categories (murder, assault, robbery, burglary and motor-vehicle theft). They included two deterrence variables: arrests for a specific crime and number of police officers. The model includes police number as a determinant of crime because it may have an additional general deterrent effect in addition to arrests for specific crimes. Using high frequency (monthly) time series data enabled them to avoid most of the simultaneity issues of cross-section models. Indeed, because current arrests are likely to be related to current criminal activity, a simultaneity bias is created if simultaneous values of arrests are included in the crime equation. Exclusion of simultaneous values of arrests helps specify the crime equation and avoid simultaneity bias. It is plausible that increased arrests do not immediately affect criminal behavior. It takes time for criminals and potential criminals to perceive that an increase has occurred. If it takes at least a month for criminals to process this information and change their behavior, crime should depend on lagged arrests.

[2] Indeed, when only one type of sanction is considered, one would expect that the effect assigned to this variable really includes effects of punishment variables correlated with that type. However, a better alternative is to use several sanctions simultaneously (Eide, Rubin & Shepherd, 2006).

In time series models, the usual techniques of regression analysis can lead to misleading conclusions when the variables have stochastic trends. In particular, if the dependent variable and at least one independent variable contain stochastic trends, and if they are not co-integrated, the regression results are spurious. To correctly specify the crime equation, the variables must be checked for stochastic trends. In no case could Corman and Mocan (2000) reject the unit root hypothesis for employed variables. This means that the proper specification of the equation should involve regressing the first difference in crime variables on the first difference in police and arrests and should not include a time trend as regressor.

All five crime categories were influenced by the number of police officers with short lags. For example, changes in the simultaneous value and two past values of police-force growth (lags = 0-2) influenced the current rate of growth of murders; and the growth rate of assaults was affected by the growth rate of simultaneous and immediate past of police numbers, however the coefficients were not significant for assault even at10% significance level. It is interesting to note that arrests had different lag structures for violent and non-violent crimes. Arrests had short-lived impacts for assault ($\delta_1 = -0.056$) and murder ($\delta_1 = -0.127$): assaults were influenced by arrests up to four months previously, and murders were influenced by three month lags of murder arrests.[3]On the other hand, robberies, burglaries, and motor-vehicle thefts showed a longer-term dependence on arrests: robberies and motor-vehicle thefts were influenced by arrests that took place up to 12 and 14 months previously, respectively; burglaries showed the longest dependence on arrests with 21 month lags.[4]

The results of this study confirm that violent crimes, which are mostly affected by visceral factors (here murder and assault), are relatively less responsive to deterrence

[3] For murder, only the first lag of arrest was significant at the level of 10%. For assault none of the coefficients for arrest lags were significant.
[4] For robbery, burglary and vehicle theft, most coefficients for arrest lags were significant at 5% level.

variables (here number of police officers and arrests) than non-violent crimes (robbery, burglary and vehicle theft).[5]

Wolpin (1978) used annual data on crime in England and Wales for the period 1894-1967 (excluding the years of WWI and WWII) to test the deterrence hypothesis for a vast range of crimes. This study also included a wide range of deterrence variables (clearance rate, conviction rate and imprisonment rate as variables for probability of apprehension and average prison sentence, recognizance rate and fine rate as punishment variables).He also used a range of control variables. Exploiting time series data, Wolpin (1978) also checked for the conventional simultaneity problem between crime rate and deterrence variables. The magnitude and significance of estimated deterrence elasticity for different kinds of crime against property was relatively higher than estimated for crimes against persons. These results also confirm that comparatively more influential visceral factors in crimes against persons (violent crimes) decrease the effectiveness of deterrent mechanisms of the judicial system(for more detailed information about the magnitude of estimated elasticities, see Appendix , Table A.1).

Devine, Sheley and Smith (1988) used annual time-series US data for the period 1948-1985 to examine the influences of imprisonment rate and some macroeconomic variables (inflation and unemployment) on annual fluctuations in rates of homicide, robbery, and burglary. Considering the potential simultaneity problem related to crime rates and imprisonment rate and also existence of a unit root in applied variables, they specified first-difference equations and applied 2SLS to estimate coefficients. The signs of all the coefficients estimated for imprisonment rate, the only deterrence variable in their model, were negative and highly significant. The interesting point in line with our hypothesis was that the relative magnitude of the coefficients for burglary and robbery were higher than those for homicide. In some specifications, this difference was

---

[5] In some studies, robbery is considered a violent crime. Because the primary motive of robbery is pecuniary and violence is used as a tool, we assumed a relatively lower influence of visceral factors in robbery than in murder and assault.

considerable.[6] These results held even when the authors checked other covariates, such as age composition and criminal opportunities. Again, these results sustain our hypothesis that deterrence variables are less effective against crimes driven by visceral factors.

Schissel (1992) used annual time-series Canada data for the period 1962-1988 to study the influences of prison population size and some macroeconomic variables (inflation and unemployment) on annual fluctuations in rates of homicide, robbery, and theft. He ran his model applying first differences of variables. He also checked for conventional simultaneity and used lagged independent variables to deal with this problem. To avoid misleading results due to spurious regression he applied a first-difference model. However, unexpectedly, the coefficients estimated for the change in police numbers were positive for all crime groups, but only significant for robbery and not significant at all for the two other crime groups. This is may be partly due to the simultaneity problem. In contrast, all coefficients estimated for change in prison population size were negative and highly significant. The estimated deterrent effect for homicide, robbery and theft were -0.025, -0.487 and -10.884, respectively. The deterrent effect of imprisonment on theft was considerably higher than on the other two crimes. As expected, theft was more responsive to deterrence variables than homicide and robbery.

1.3.2 Cross- section studies

The bulk of econometric studies of crime consist of cross-section regression analyses based on aggregate data. Some are broad, including many types of regional areas, estimation techniques and crimes, whereas others concentrate on particular types of crime, such as property crimes or homicide. Most of the cross-section studies reviewed here allowed two-way causation to deal with the simultaneity problem by various specifications of the general model:

[6] For example in one specification, the coefficients estimated for homicide, robbery and burglary were -0.69, -1.86 and -13.87, respectively. All were significant at the level of 1% (see Table 1 in Devine, Sheley and Smith (1988), American Sociology Review).

$$
\begin{aligned}
C &= f(P, S, Z_j) \\
P &= g(C, R, Z_k) \\
R &= h(C, Z_l) \qquad (1.1)
\end{aligned}
$$

where C is the crime rate (number of crimes per head of population), P, the probability of punishment; S, severity of punishment; R, resources per capita devoted to the criminal justice system; and Z_j, Z_k, Z_l are vectors of socio-economic factors. Various socio-economic factors are included as explanatory variables in all three equations (Eide, Rubin & Shepherd, 2006).

The first major cross-section study appearing after Becker's theoretical article was by Ehrlich (1973). He studied seven types of crimes in US based on data from all states for 1940, 1950, and 1960. For lack of data on police expenditure in 1940 and 1950, the coefficients estimated by OLS in these years suffer from the simultaneity problem. We therefore report only the results for 1960, for which the coefficients were estimated by 2SLS and SUR using a simultaneous equation model.

Let us start with estimated elasticities for probability of apprehension. In the2SLS and SUR estimations they are negative and highly significant (columns 1 and 3 of Table A.2, Appendix). Indeed, except for robbery, estimated elasticities for other kinds of property crime are lower than those estimated for all types of crimes against persons. Murder responds poorly to imprisonment, whereas rape and assault are more responsive than some kinds of property crimes, such as car theft and robbery. Thus our hypothesis only holds for the violent crime of murder here. In contrast, both rape and assault were responsive to the deterrence measures, contrary to our hypothesis and the findings of other similar studies. Regarding the results for assault, Ehrlich writes: "To some extent crimes against the person may be complementary to crimes against property, since they may also occur as a by-product of the latter. This is particularly true in the case of assault, for it is generally agreed that some incidents of robbery are classified in practice as assault. This may be one reason why assault exhibits a greater similarity to crimes against property in its estimated functional form" (Ehrlich, 1973, p- 53).

In addition, Ehrlich's study has been thoroughly scrutinized by several authors, some of whom expressed harsh assessments. Revisions, replications, and extensions of Ehrlich's studies by Forst (1976), Vandaele (1978) and Nagin (1978) resulted in more moderate deterrent effects of probability of apprehension and severity of punishment. Forst(1976)also found that by introducing variables thought to be correlated with the punishment variables, such as population migration and population density, the punishment variables lost their statistical significance.

Kelly (2000) used data based on all metropolitan counties and the 200 largest counties of the US in 1991 to investigate the link between inequality, crimes against property and violent crimes. Expenditure per capita on police was the only deterrence variable included in his study. He first considered this deterrence variable exogenous and ran Poisson regressions with log explanatory variables, the estimated coefficients of which could be interpreted directly as elasticities. Although the elasticities estimated for violent crimes in all specifications were not significant even at 10% significance level, they were lower than the highly significant elasticities estimated for property crimes. He finally considered expenditure on police to be endogenous and estimated new police elasticities for violent and property crimes by instrumental variables and GMM. Again, the elasticity estimated for violent crime was not significant, but the elasticity estimated for property crime was significant and even higher than in the previous model (this result for property crime only held for the 200 largest counties).

Withers (1984) pooled cross-sectional and time series data for the eight states and territories of Australia on a fiscal year basis from 1963-64 to 1975-76 to examine the deterrent effects of court committals and imprisonment on a vast range of violent and property crimes. He checked for conventional simultaneity in the crime equation and applied simultaneous equation models to deal with it. His analysis found strong and robust results in favor of the deterrence hypothesis for various categories of property crime. Court committals and imprisonments were found to act as significant deterrents across a range of property crime categories and to provide significant explanation for the variations observed in recorded property crime rates over the study period. So-called "crimes of passion", such as homicide and rape, were found to be unresponsive to deterrence at the margin. The results of this study were in line with our hypothesis.

Furlong and Mehay (1981) used data based on 38 police districts in the metropolitan area of Montreal to design a simultaneous model (concerning the simultaneity problem) to examine deterrence and other socioeconomic variables in relation to certain crime categories. They focused on robbery, breaking and entering, theft, an index of property crime including these three crimes and a total crime index including some violent crimes and property crimes. They also emphasized the dynamic aspect of population in different districts and normalized the number of crimes for resident population and dynamic population.[7] The sign of the clearance rate (ratio of number of cleared crimes to total reported crimes) coefficient was negative in all crime categories and the associated t-values were generally high. The risk of police arrest appeared to produce a significant deterrent effect, even in the category of all major offences, which included violent crimes. The interesting point is that inclusion of violent crimes in the crime index decreased the deterrent effect of clearance rate in both crimes indices: crime normalized for resident population and dynamic population. Indeed, the highest estimated deterrence effect of clearance rate, -0.06, was related to breaking and entering, which seems to be less affected by visceral factors. In contrast, the coefficient estimated for the total crime index that included violent crimes was only -0.03, i.e. 50% lower than the deterrence effect on breaking and entering.

1.3.3 Panel Data Analysis

Economic models of crime using aggregate data that rely heavily on cross-section techniques do not control for unobserved heterogeneity. This is even true for studies using simultaneous equation models (Cornell and Trumbull, 1994). This section reviews studies that accounted for unobserved heterogeneity using panel data techniques for testing the deterrence hypothesis.

[7] Dynamic population includes people who move to a district for work or any other reason but are not resident in that district.

Bounanno and Montolio (2008) used a panel dataset of Spanish provinces from 1993 to 1999 to design a dynamic model of crime including dynamic features of crime and criminal behavior, due for example to recidivism. They applied the GMM estimator to study the deterrent effects of clearance rate and condemnation rate (ratio of condemned profiles to number of cleared crimes) on property crimes and crimes against persons. They also checked for certain demographic and socioeconomic variables in their model. Dynamic panel data model make it possible to check for province-specific effects and measurement errors in reported crimes. Both the Sargan test of over-identifying restrictions and test of serial correlation of error terms confirmed that the model was sufficiently well specified. However, the coefficients estimated for condemnation rate were not significant even at 10% level; whereas those for clearance rate for property crime was -0.0202 and highly significant at 1%. The coefficient estimated for crime against persons was only -0.001 and not significant. The results of this study confirmed that deterrent effects are more effective for property crimes than crimes against persons, which are presumably more sensitive to visceral factors.

Cherry and List (2001) used a panel data set of North Carolina counties for the period 1981-87 to investigate the deterrence hypothesis on a vast range of crimes. They emphasized aggregation bias due to pooling of crime types in a single decision model and ran a unique decision model for various kinds of crimes. They also considered great variation of sanctions and the probability of arrest across various types of crimes. Because clearance rates are much greater for violent crimes (0.78) than property crimes (0.22), they used specific arrest and clearance rate in their models. They applied fixed effects (FE or within estimator) to estimate deterrence effects of probability of arrest for different kinds of violent and property crimes. The estimated deterrent effect of probability of arrest was 45% greater for property crimes than for violent crimes, a difference that is significantly different from zero at the significance level of 5%. This differential was even more pronounced for disaggregated crime types as the estimated effect of probability of arrest was 55% greater for burglary and larceny than for murder and rape. All together, the results of this study, too, seem to be in line with our hypothesis.

Saridakis and Spengler (2012) used data based on a panel of Greek regions for the period 1991-98 to study the relationship between crime, deterrence and unemployment. They applied the GMM-system estimator to a dynamic model of panel data. The specification tests (Saragan test of over-identifying restrictions and serial correlation of error terms) indicated that the model was sufficiently well specified. The results showed that property crimes (breaking and entering and robbery) were significantly deterred by higher clearance rates. In this group, higher clearance rate had no significant effect on theft of motor cars. For violent crimes (murder, rape and serious assault), however, the effects of clearance rate were found to be consistently not significant.[8]

Gould, Weinberg and Mustard (2002) applied panel data on US counties for the period 1979-97 to examine the impact of wages and unemployment on crime; they also used instrumental variables to establish causality. To check the robustness of their results, they also included some deterrence, individual and family characteristics in their model. They applied the FE estimator. The only deterrence variable in their study was arrest rate. To avoid the endogeneity problem, they simply excluded per capita expenditure for police as a deterrence variable from their model. The arrest rate showed a significant negative effect for all types of crime. The coefficients estimated for property crimes were considerably larger than those estimated for violent crimes, in line with our main conjecture (for more details about estimated coefficients see Table A.4 in Appendix).

Mustard (2003) emphasized the bias of omitted variables due to conviction rates and time served along arrest rates, thus employing a more complete set of deterrence variables in his model. By analyzing comprehensive conviction and sentencing data, he provided new evidence about the relation between criminal behavior and sanctions and a more complete assessment of the penalties associated with illegal activity. Indeed, he observed that if arrest rates are positively correlated with omitted variables, ignoring them overstates the effect of arrest rates. The inverse is true when they are correlated

[8] The coefficients estimated for violent crimes, however, were not significant but all negative and lower than those estimated for property crimes.

negatively. Using panel data at US county level from 1977-92, he studied a more complete model of crime. He also applied the FE estimator. The elasticities estimated for sentence lengths were not significant for any crime type. Arrest rate significantly deterred all types of crime and the deterring effects of arrest rate were significantly higher for property crimes than violent crimes in a striking manner. This was also true for lagged arrest rates that had no deterrence effect at all for most violent crimes. The relative deterrent effect of conviction rate on various crimes was unlike that of arrest rate. Conviction rate did not deter burglary and robbery at all. Its deterrent effect was low for rape, but it considerably deterred murder, assault, car theft and larceny. An unexpected finding in relation to our hypothesis was the higher deterrent effect of conviction rate on murder and assault in comparison with car theft. This may partly be due to the low conviction rate for car theft in comparison to assault and murder.

Raphael and Winter-Ember (2001) used US state-level panel data for 1971-97 to study the deterrence effect of imprisonment rate on the property crimes and violent crimes. Their study mainly focused on the relationship between crime and unemployment and most of their results only considered coefficients estimated for unemployment. The deterrent effect of imprisonment rate was reported in only one case. For all crimes, they specified three models: models including state and year fixed effects; models including state and year fixed effects and state-specific linear trends; and models including state and year fixed effects and linear and quadratic trends. In all property crime models, the effect of imprisonment rate was negative and significant at 1% level. The magnitude of the estimated elacticities indicated that a 0.1% increase in imprisonment rate caused a 0.13-0.1% decline in the property crime rate. The results for violent crimes were mixed. In the first specification, the coefficient was small and insignificant. Adding linear time trends increased the point estimate of the imprisonment coefficient, but the variable remained not significant, even at 10% level. Finally, adding quadratic time trends to the model increased the point estimate further, and the coefficient became significant at 5% level. So only in the third specification did imprisonment show a deterrent effect on violent crime; estimated elasticity was -0.042, which is considerably lower than that estimated for property crimes, in line with our hypothesis.

Levitt (1998) used panel data for the 59 largest US cities over the period 1970-92 to discriminate between deterrence, incapacitation and measurement error in a study of the deterrent effect of arrest rates on crime level. He focused on the seven major felonies reported by the FBI (murder, rape, aggravated assault, robbery, burglary, larceny and motor vehicle theft). He ran a panel data model and checked for related socioeconomic covariates. He applied the FE estimator based on the Hausman test. He concluded that there was little evidence that measurement error was responsible for the observed relationship between arrest rates and crime rates in all seven crime groups. Then he tried to decompose deterrence and incapacitation effects for all crimes. He concluded that deterrence was empirically stronger than incapacitation in reducing crime, particularly property crimes. These conclusions, however, are subject to the important caveat that it is difficult to check for endogeneity of arrest rates. The deterrent effect of arrest for all kinds of property crimes was considerable and highly significant. In contrast, its effect on the violent crimes was unexpectedly positive but not significant (Table A.6 in Appendix). The estimated results are in line with our hypothesis. While violent crimes seem to be unresponsive to an increased arrest rate, various property crimes are highly responsive. This implicitly confirms that because of the influence of visceral factors in violent crimes, potential offenders do not care, or care relatively less, about the risk of apprehension and punishment.

Almost all the studies reviewed sustain the hypothesis that violent crimes are less responsive to deterrence variables than non-violent crimes because of the influence of visceral factors. Table 1.1 summarizes the types and results of the reviewed studies.

1.4 Influence of visceral factors and violent crimes

After verifying the comparatively lower responsiveness of violent crimes to deterrence variables by the empirical survey in the last section, we now draw on Lowenstein (2004) to present some propositions about visceral factors that may underpin the survey findings.[9] These propositions are applied to explain why violent crimes, such as rape,

[9] For more detailed and formal style of these propositions, see Lowenstein (2004).

murder and aggressive assault, are relatively less responsive to standard deterrence variables in the economics of crime literature.

Proposition 1. *The discrepancy between the actual and desired value*[10] *placed on a particular good or activity increases with the intensity of the immediate good-relevant visceral factor*. For instance, in the case of a rapist, an intensified visceral factor of sexual desire increases the discrepancy between rape as a method of satisfying sexual desire and sexual relations with one's own partner in a normal peaceful way. Another example is homicide when the murderer takes justice into his own hands. In both cases, intensified visceral factors increase the discrepancy between actual (rape and homicide) and desired (sexual courtship and court decision) values attributed by offenders. This is why most such offenders suffer remorse and confess that "they lost control" or "emotions took over".

Proposition 2. *Future visceral factors produce little discrepancy between the value we plan to place on goods in the future and the value we view as desirable*. The idea is that visceral factors mostly affect behavior and increase discrepancy between actual and desirable values when stimulated and intensified.

Proposition 3. *Increasing the level of an immediate and delayed visceral factor simultaneously enhances the actual valuation of immediate relative to delayed consumption of the associated good.* This proposition emphasizes the present-oriented influence of associated visceral factors. It can help explain why expected punishment is less deterrent for crimes with intense visceral factors (rape and homicide) because immediate visceral factors related to crime (lust and revenge) dominate the delayed visceral factors of fear of conviction and punishment.

Proposition 4. *Currently experienced visceral factors have a mild effect on decisions for the future, even when those factors will not be operative in the future*. This proposition again emphasizes the time horizon of the influence of visceral factors, which arise, are acted upon in the moment, and cease. In the other words, visceral

---

[10] "Actual value" means the value implied by the individual's behavior; "desired value" means the value that the individual views as being in his or her self-interest (Lowenstein, 2004).

factor influences is mostly short- rather than long-term. Combined with proposition 3 it emphasizes the relatively mild deterrent effect of expected punishment on potential offenders and even offenders who have been punished in the past. In other words, intensifying current fear of punishment by punishing convicted offenders may have little deterrent effect in the future for potential and convicted offenders. This proposition offers an explanation for recidivism of convicted offenders and even repeated victimization of potential victims.

Proposition 5. *People underestimate the impact of visceral factors on their own future behavior*. In a country where rape is punished severely (say, life imprisonment), if a subject is asked what he would do if given the opportunity for sexual intercourse by force with a desirable girl, he may answer that he would never take the opportunity because he does not wish to spend the rest of his life in the prison. However, his resolve may change in the real situation because of lust, the intensity which may depend on sexual deprivation of the offender or the provocative nature of the potential victim.

Proposition 6. *As time passes, people forget the degree of influence that visceral factors had on their own past behavior. As a result, past behavior that occurred under the influence of visceral factors will increasingly be forgotten, or will seem perplexing to the individual.* This proposition emphasizes the short-lived, permanent and independent nature of visceral factors. Visceral factors may be intensified by stimulus at any time, irrespective of previous experiences. This explains recidivism for crimes with intense visceral factors. For instance, a subject may be irascible and act aggressively.

Study	Case Study	Model & Data	Results
Corman & Mocan (2000)	NY	monthly time series (1970-96), OLS estimator	lower deterrent effects for violent crimes
Wolpin (1978)	UK & Wales	annual time series 1894-1967 (excluding years of WWI and WWII), OLS estimator	lower deterrent effects for violent crimes
Devine, Sheley and Smith (1988)	USA	annual time series 1948-85, 2SLS estimator	lower deterrent effects for violent crimes
Schissel (1992)	Canada	time series model (1962-88), OLS estimator	lower deterrent effects for violent crimes
Ehrlich (1973)	USA	cross section model using states in 1960, OLS estimator	Mixed results; some violent crimes are more responsive to deterrence variables than non-violent crimes
Kelly (2000)	USA	cross section model using all metropolitan counties and 200 largest counties in 1991, OLS and 2SLS estimator	lower deterrent effects for violent crimes
Withers (1984)	Australia	pooled cross-sectional model using states from 1963-64 to 1975-76, 2SLS estimator	lower deterrent effects for violent crimes
Furlong & Mehay (1981)	Montreal	cross section model using 38 police districts in metropolitan era, 2SLS estimator	lower deterrent effects for violent crimes
Bounanno & Montolio (2008)	Spain	Dynamic panel data model using set of provinces 1993-99, GMM-system estimator	lower deterrent effects for violent crimes
Cherry & list (2001)	North Carolina	panel data model using set of counties 1981-1988, FE estimator & FE2SLS	lower deterrent effects for violent crimes
Saridakis & Spengler (2012)	Greece	Dynamic panel data model (DPD) using set of regions 1991-1998, GMM-system estimator	lower deterrent effects for violent crimes
Gould, Weinberg & Mustard (2002)	USA	panel data model using set of counties 1979-1997, FE estimator	lower deterrent effects for violent crimes
Mustard (2003)	USA	panel data model using set of counties 1977-1992, FE estimator	lower deterrent effects of arrest rates for violent crimes but- mixed results of conviction rate and for some violent crimes indicated even higher deterrent effect
Raphael &Winter-Ember (2001)	USA	panel data model using set of states 1971-1997, FE estimator	lower deterrent effects for violent crimes
Levitt (1998)	USA	panel data model on 59 of the largest cities 1970-1992, FE estimator	lower deterrent effects for violent crimes

Table 1.1- Summarized results of empirical studies surveyed

Proposition 7. *The first six propositions apply to interpersonal as well as intrapersonal comparisons, where other people play the same role vis-a-vis the self as the delayed self plays relative to the current self:*

I. We tend to become less altruistic than we would like to be when visceral factors intensify. In all kinds of crime, whether property or violent crimes, offenders do not care about their victims, but the point about violent crimes is that visceral factors have more influence on this irresponsibility towards others. A clear example is homicide or rape. A murderer or rapist is the opposite of altruistic, i.e. selfish. He sacrifices the victim to satisfy visceral factors of revenge or lust.

II. When we experience a particular visceral factor, we tend to imagine others experiencing it as well, regardless of whether they actually are. This emphasizes the similar nature of human beings. High intensity visceral factors can happen to anybody.

III. People underestimate the impact of visceral factors on other people's behavior. This can be observed in others' judgments of those convicted of crimes with intense visceral factors, such as rape and murder. It seems strange to everyone that somebody might trade his life for short forcible sex by raping, however executions of convicted rapists continue in certain countries year by year. Another aspect of this feature can be observed in the behavior of victims. Victims usually underestimate the power of visceral factors over offenders' behavior and even stimulate the intensity of these factors. Indeed, in most cases, instead pouring water on the fire of visceral factors they pour on gasoline.

1.5 Visceral factors influences in Becker's model: Some Policy Implications

This section examines the influence of visceral factors on violent criminal behavior in Becker's model and analyzes optimal policies in this framework. In Becker's model, supply elasticity of crime with respect to punishment and probability of apprehension and conviction indicates the sensitivity of crime supply to these deterrence variables.[11] These elasticities are defined as follows:

$$\varepsilon_f = -\frac{\%\Delta f}{\%\Delta O} = -\frac{f}{O} O_f \qquad\qquad \varepsilon_p = -\frac{\%\Delta P}{\%\Delta O} = -\frac{p}{O} O_p \qquad (1.2)$$

Where p is the probability of apprehension and conviction, f is the amount of punishment and O is the number of offences. As verified in our empirical survey in the previous section, the more influential role of visceral factors in violent crimes suggest that these elasticities are considerably lower for violent crimes than property crimes. In a nutshell, the message we received from optimality conditions in a Beckerian framework is that when elasticity of crime supply with respect to apprehension and punishment is low, for example because visceral factors influence criminal behavior, whether these crimes incur high or low cost to society, apprehension and punishment do not deter potential offenders in the future. However, at the same time, optimality conditions imply that when crimes inflict high net damage on the society, as in the case of rape and murder, they should be punished severely. These two optimality implications move in opposite directions. As a conclusion, when apprehension and punishment do not work sufficiently well, as a guide to preventive policy making we should think more fundamentally about these crimes and try to limit them in other ways besides punishment. In this way, a preventive strategy could be to survey victims and get their advice on how to inform other potential victims and lower the likelihood of future victimization. At the same time, we should also try to answer to some more fundamental questions about the supply of these kinds of crime in order to limit them. For instance, for rape we could ask why in two countries with the same punishment for rape, the rate is high in one and low in the other, or whether legalization of prostitution could limit rape. Is this problem related to heavy drinking of alcohol and should we impose a higher tax on

[11] For more details about Becker's model, see Becker (1968).

the consumption of alcohol? For murder, we could consider whether gun and other policies may be related to high violence in a society.

1.6 Conclusion

Beginning with the seminal work of Becker (1968), extensive economic literature has analyzed criminal behavior and issues of criminal justice. The key aspect of economic models of crime is the idea of deterrence: rational agents, faced with higher probabilities of detection or heavier penalties, will commit fewer criminal acts.

Here we looked at this issue from a different view. We first studied the comparative responsiveness of different kinds of crime to deterrence variables, considering the diverse influence of visceral factors on them. We assumed that visceral factors are more influential in violent crimes than property crimes, so that the former are less responsive to deterrence variables than the latter.

To verify this hypothesis, we reviewed 15 empirical studies on different databases and with methodologies ranging from cross-sectional to panel data analysis. Their results were mostly in line with the hypothesis. Indeed, in most cases the coefficients or elasticities estimated for different kinds of violent crime (murder, rape, assault ...) were significantly less than those estimated for various kinds of property crime (burglary, larceny, car theft ...). On the basis of the accuracy of this hypothesis, it may be said that the rational choice theory of crime and its predictions are more applicable to property crimes than to violent crimes driven by strong emotions.

We then applied the verified hypothesis of the influence of visceral factors in violent crimes to Becker's model for an evaluation of optimal and currently employed policies for combating violent crimes. Inasmuch as violent crimes such as murder and rape inflict high social damage, optimality conditions suggest that they should be convicted and punished more severely. At the same time, because visceral factors play a role in these crimes, their elasticities with respect to apprehension and punishment are low and optimality conditions suggest that they should be convicted and punished leniently. In other words, apprehension and punishment inflict social loss but do not deter potential offenders. From a policy making point of view, it seems better to focus on other strategies for solving these problem, if they *can* be solved. In the other words, when apprehension and punishment do not work sufficiently well, in preventive policy making we should think more fundamentally about these crimes and try to decrease them in ways other than punishment. A survey of victims could be useful for this purpose. At the same time we should try to answer more fundamental questions about the supply of such crimes in order to limit them.

Appendix: Tables of summarizing results of empirical studies

This appendix shows the coefficients estimated for deterrence variables in the empirical papers surveyed in Section 3. Other covariates are not included.

	Larceny	burglary	car theft	Robbery	malicious wounding	felonious wounding
clearance rate[12]	-0.922	-0.743	-0.591	-1.35	-0.558	-0.529
	(3.53)	(4.45)	(2.56)	(3.82)	(1.36)	(1.98)
conviction rate[13]	-0.309	-0.617	-0.077	-1.019	-0.161	-0.284
	(0.45)	(1.6)	(0.44)	(0.99)	(0.56)	(2.01)
fine rate	-0.247	-0.048	…	…	-0.039	…
	(1.24)	(2.33)			(0.83)	
imprisonment rate	-0.573	-0.131	-0.159	-0.921	-0.152	-0.068
	(3.2)	(1.25)	(2.35)	(1.54)	(1.16)	(1.38)
recognizance rate[14]	-0.834	-0.591	-0.334	-0.611	-0.646	-0.671
	(4.66)	(4.77)	(1.22)	(1.30)	(2.15)	(1.6)
average prison sentence	0.393	-0.169	0.152	-0.722	-0.056	0.004
	(1.86)	(1.09)	(0.96)	(1.39)	(0.49)	(0.04)

Table A.1 - Estimated elasticity of deterrence variables for different kinds of crime
Source: Wolpin (1978)

- Numbers in brackets are t-values.
- For car theft, robbery and felonious wounding, data on fines was not available.

[12] Clearance rate is the rate at which crimes are cleared by the police and measured as the ratio of cleared crime to reported crimes.
[13] Conviction rate is the rate of conviction conditional on arrest and measured as the ratio of convictions to the number of arrests.
[14] Recognizance in this study is a kind of punishment.

	probability of apprehension	average time served in state prison	probability of apprehension	average time served in state prison
	2SLS		**SUR (seemingly unrelated regression)**	
Robbery	-1.303	-0.372	-1.112	-0.286
	(-7.011)	(-1.395)	(-6.532)	(-0.75)
Burglary	-0.724	-1.127	-0.624	-0.996
	(-6.003)	(-4.799)	(-5.376)	(-4.26)
Larceny	-0.371	-0.602	-0.358	-0.654
	(-2.482)	(-1.937)	(-2.445)	(-1.912)
car theft	-0.407	-0.246	-0.409	-0.239
	(-4.173)	(-1.682)	(-4.674)	(-1.747)
property crime	-0.796	-0.915		
	(-6.14)	(-4.297)		
Murder	-0.852	-0.087	-0.913	-0.018
	(-2.492)	(-0.655)	(-3.062)	(-1.71)
Rape	-0.896	-0.399	-0.93	-0.436
	(-6.08)	(-2.005)	(-6.64)	(-2.318)
Assault	-0.724	-0.979	-0.718	-0.78
	(-3.701)	(-2.301)	(-4.046)	(-2.036)
crimes against persons	-0.803	-0.495		
	(-6.603)	(-3.407)		

Table A.2 - Estimated deterrence elasticity for different types of crime
Source: Tables 4 and 5, Ehrlich (1973)

- Numbers in the brackets are t-values.
- The model is based on natural logarithm, so estimated coefficients are deterrence elasticity of crime.

	Violent	Property	Murder	rape	robbery	Assault	Burglary	larceny	car theft
probability of arrest	-0.284	-0.413	-0.327	-0.34	-0.167	-0.421	-0.557	-0.527	-0.313
	(-12.9)	(-25.81)	(-6.05)	(-7.5)	(-4.39)	(-9.8)	(-17.96)	(-18.17)	(-12.03)
probability of conviction	-0.194	-0.214	-0.028	-0.111	-0.131	-0.546	-0.265	-0.249	-0.169
	(-7.46)	(-11.26)	(-0.55)	(-2.22)	(-2.62)	(-10.7)	(-8.8)	(-8.5)	(-4.22)
probability of imprisonment	-0.115	-0.085	-0.1	-0.186	0.051	-0.229	-0.24	-0.132	0.044
	(-2.4)	(-2.36)	(-1.06)	(-2)	(-0.55)	(-2.49)	(-4.44)	(-2.44)	(0.59)
length of sentence	0.104	-0.007	0.119	0.119	0.096	0.0807	-0.036	0.016	0.001
	(2.54)	(-0.22)	(1.48)	(1.5)	(1.21)	(1.02)	(-0.78)	(0.35)	(0.016)
Police	-0.2	-0.367	-0.157	-0.23	-0.281	-0.119	-0.393	-0.395	-0.25
	(-5.12)	(-12.23)	(-2.03)	(-3.02)	(-3.7)	(-1.58)	(-8.73)	(-8.9)	(-4.16)

Table A.3 - Estimated deterrence elasticities for various kinds of crime

Source: Cherry and List (2001)

- Numbers in brackets are t-values.

	property crime index	car theft	Burglary	larceny	violent crime index	aggravated assault	murder	robbery	rape
county arrest rate	-0.01	-0.01	-0.01	-0.01	-0.004	-0.003	-0.002	-0.006	-0.004
	(0.002)	(0.001)	(0.003)	(0.001)	(0.0003)	(0.0003)	(0.0002)	(0.0005)	(0.001)
log state per capita	0.3	0.51	0.34	0.26	0.36	0.39	0.23	0.55	0.04
expenditure on police	(0.1)	(0.176)	(0.11)	(0.1)	(0.11)	(0.14)	(0.16)	(0.13)	(0.10)
log state per capita	-0.02	-0.26	-0.04	0.07	-0.06	0.01	-0.28	0.14	0.45
employment of police	(0.13)	(0.2)	(0.16)	(0.12)	(0.14)	(0.16)	(0.18)	(0.18)	(0.12)

Table A.4 - Estimated coefficients for deterrence variables

Source: Gould, Weinberg and Mustard (2002)

- Numbers in brackets are standard errors.

	arrest rate	lagged arrest rate	conviction rate	length of sentence
Ln (murder rate)	-0.0035*	0	-0.0028*	0.00002
Ln (rape)	-0.0026***	-0.0031	-0.0009***	0.0005
Ln (robbery)	-0.0016*	-0.0035	-0.0025	0.00064
Ln (assault)	-0.0019**	-0.0038*	-0.0061***	0.0002
Ln (burglary)	-0.0123**	-0.0102**	-0.0006	0.00116
Ln (larceny)	-0.0072**	-0.0046***	-0.0076**	0.00036
Ln (car theft)	-0.0052**	0.0003	-0.0023***	0.00007

Table A.5 - Logarithm of crime rates on regressors, including conviction and sentencing data.

Source: Mustard (2003)

- *, **, *** are significant at 1%, 5% and 10% levels, respectively.

	Total effect	Deterrence	Incapacitation
Murder	0.015	0.045	-0.03
	(0.056)	(0.076)	(0.046)
Rape	-0.277	0.241	-0.518
	(0.054)	(0.162)	(0.141)
aggravated assault	-1.253	-0.565	-0.688
	(0.143)	(0.291)	(0.199)
Robbery	-0.399	-0.365	-0.034
	(0.072)	(0.122)	(0.094)
Burglary	-2.438	-2.342	-0.096
	(0.219)	(0.355)	(0.2)
Larceny	-1.565	-1.448	-0.117
	(0.125)	(0.162)	(0.093)
car theft	-0.583	-0.457	-0.126
	(0.181)	(0.225)	(0.103)

Table A.6 – Deterrence and incapacitation effects per arrest

Source: Levitt (1998)

- Numbers in brackets are standard errors.
- Column 1 is the change in the number of crimes in a given category per arrest and reflects both deterrence and incapacitation effects. Columns 2 and 3 show the estimated number of crimes of a given category eliminated by deterrence or incapacitation per arrest.

Chapter 2

Economic Analysis of Criminal Law

Abstract: This chapter seeks to answer two inter-related type of questions. The first question compares the two opposing approaches that have traditionally inspired criminal law, namely the utilitarian and the retributive approach asking in what ways there might be irreconcilabilities between the two. The second question concerns criminal law in practice, which is actually a hybrid of the two approaches, and asks why different societies treat certain crimes rather differently.

Both traditional justifications for punishment, utilitarian and retributive perspectives lead to the same conclusion: support for an institution of punishment. However, there have been irreconcilabilities between these two rival approaches in the distribution of punishment. The utilitarian perspective sees deterrence as the main distributive principle of punishment and cares about efficacy of severity and certainty of punishment in reducing crime as well as about the costs of imposing punishment. This has led to a deviation from what retributivists believes to be a deserved punishment.

In practice, criminal law reflects both utilitarians' and retributivists' principles of punishment (deterrence, incapacitation, rehabilitation and "just deserts", respectively). However, depending on the type of crime and the specific characteristics of some offenders, each of these distributive principles of punishment can have priority over others. The scope of criminal law depends on activities considered harmless or harmful. The divergence in the associated punishment for a certain crime stems from the relative importance of factors that different societies consider in optimizing social loss from criminal activities. These factors are: the degree of harmfulness of the crime, retributive or regretful emotions towards offenders or what is called "humanity of punishment" and the deterrent effect of

certain punishments. Different attitudes towards these aspects lead to differences in criminal law.

JEL: K14, K42

Keywords: Law enforcement, Retributive approach, Utilitarian approach, Comparative criminal law, Efficiency

2.1 Introduction

This chapter includes two different sections. We apply Becker's social loss function from criminal activities to deal with two different set of questions; first we try to formalize the apparent irreconcilabilities between Utilitarians and Retributivists about justification of punishment. Building on Becker's work we use economic modelling to shed light on the irreconcilabilities between these two old rivals in philosophy of punishment. Moreover, we advance a more formalized explanation of why comparative criminal law across different societies, for example why they impose different punishment for a certain kind of crime, say murder. The value added of our analysis in this chapter is precisely the attempt to relate fundamental questions in formalized way, thus instituting a dialogue between bare analytical models on the one hand and merely discursive analyses on the other hand. .

The chapter is organized as follows: the next section introduces Becker's social loss function from criminal activities. Using the social loss function section 3 formalizes the comparison between the arguments of the utilitarian and the retributive justifications for punishment, while section 4 proposes a formalized reading of comparative criminal law from an economic perspective. Section 5 concludes.

2.2 Crime, punishment and social loss

Let's start with examining the social loss function from criminal activities which first was introduced by Becker in 1968 in his seminal paper which we will follow closely to develop our arguments throughout this chapter. This social loss function (*L*) is defined as follows:

$$L = D(O(p, f)) + C(p, O(p, f)) + bpfO(p, f) \qquad (2.1)$$

Before defining the components, it is worth mentioning its basis in rational criminal choice theory, as indicated by the supply function of offences $O = O(p, f)$, where O is the number of offences, p and f indicate the probability of apprehension and conviction, and the amount of punishment, respectively. Consistently with the expected utility of illegal behavior we have:

$$O_p = \frac{\partial O}{\partial p} < 0 \qquad (2.2)$$

$$O_f = \frac{\partial O}{\partial f} < 0 \qquad (2.3)$$

These equations show the lynchpin of the economic model of crime: rational agents, faced with higher probabilities of detection or heavier sanctions, commit fewer criminal acts.

Now let us introduce the components of our social loss function from criminal activities. $D(O) = H(O) - G(O)$ is net social damage from offences, calculated by subtracting the gains obtained by offenders (*G*) from the harm incurred by victims (*H*). In this perspective, gains of offenders as members of society are counted in the social welfare function or equivalently are subtracted from the social loss function. Since more crimes inflict more harm on victims, they bring more gains to offenders. The assumptions on *H* and *G* are:

$$H_o = H' = \frac{\partial H}{\partial O} > 0 \qquad (2.4)$$

$$G_o = G' = \frac{\partial G}{\partial O} > 0 \qquad (2.5)$$

Since $H' > 0$ and $G' > 0$, the sign of $D' = \frac{\partial D}{\partial O}$ depends on their relative magnitudes. However, it is also, and plausibly assumed that there are plausibly diminishing marginal gains to offenders ($G'' < 0$) and increasing marginal harm to victims ($H'' > 0$) and thus $D'' = H'' - G'' > 0$. This is a key condition of optimality analysis in the following sections.

C indicates administrative costs of apprehension and conviction. Simply, we need policemen, judges, counsel and juries and other imperative infrastructures for apprehension and conviction of offenders, which all incur costs for society. Intuitively, these costs are approximated by the level of activity of the sections of the criminal justice system involved in apprehension and conviction. As an empirical measure, this "activity" can be approximated by the number of offences ending in conviction (it can simply be written as $A \cong pO$). Increased "activity" of apprehension and conviction (either higher levels of p or an increase in offences) would be more costly, as summarized by the relations:

$$C_p = \frac{\partial C(pO)}{\partial p} = C'O > 0 \qquad (2.6)$$

$$C_o = \frac{\partial C(pO)}{\partial O} = C'p > 0 \qquad (2.7)$$

If marginal costs of increasing activity are rising, as it is presumed in this model, then we have these further implications:

$$C_{pp} = C''O^2 > 0 \qquad (2.8)$$

$$C_{oo} = C''p^2 > 0 \qquad (2.9)$$

$$C_{op} = C_{po} = C''pO + C' > 0 \qquad (2.10)$$

To avoid a corner solution it is sufficient to restrict the second derivatives of this cost function as follows:

$$C_{pp} \geq 0\,,\ C_{oo} \geq 0 \text{ and } C_{po} \cong 0 \qquad (2.11)$$

Finally, the last component of Becker's social loss function indicates the costs that punishing convicted offenders impose on society as a whole. Punishment is the last link in the law enforcement chain. Convicted offenders (*pO*) receive an amount *f* of punishment in the law enforcement chain. Punishments not only affect offenders but also other members of society. Aside from collection costs, fines paid by offenders are received as revenue by others and it is presumably a socially costless transfer of money. Most punishments, however, hurt other members as well as offenders: for example, imprisonment requires expenditure on guards, supervisory personnel, buildings, food, etc. We channel the costs of punishment to society as a whole (including convicted offenders) through coefficient *b*. In the last term, *b* is the coefficient that transfers punishment of convicted offenders to the society as a loss and its magnitude depends on the type of punishment. The size of *b* varies greatly between different kinds of punishments. For instance, fines produce a gain to the latter that equals the cost to offenders, apart from collection costs, and so the social cost of fines is about zero, whereas for torture, probation, parole, imprisonment and most other punishments we can assume $b > 1$. These punishments not only incur costs for convicted offenders but resources must also be allocated to implement them, so it makes sense that for them $b > 1$.

2.3 Distributive principles of punishment: Utilitarianism versus Retributivism in an Economic Perspective

Reflection on the subject of punishment raises a number of questions, the most fundamental of which is why we punish offenders. Some answers to this question seem to conflict with others. Criminal punishment can be justified on two broad grounds. The first is "utilitarian" or "consequentialist": punishment for a past offence is justified by the future benefits it provides. The future benefits are to avoid or at least reduce future crimes by deterrence, incapacitation and rehabilitation. The other justification is based solely on the past behavior of an offender: imposing punishment on convicted offenders is a valuable end in itself and needs no further justification. This is typically referred to as a "retributive" or "just deserts" view. The arguments of these two approaches are generally considered irreconcilable: the utilitarian view is justified on the basis of utility, and "just deserts" on the basis of fulfilling a deontological moral mandate (Robinson & Darley, 1997). However, both traditional justifications lead to the same conclusion: support for an institution of punishment. A system of punishment can mete out justice and avoid future crime by deterring potential offenders, rehabilitating those convicted and incapacitating dangerous offenders. The debate between utilitrians and retributivists matters practically[15].

Debate on the justification for punishing criminals has been rife with confusion in the course of history. The evolution of criminal procedures and customs has been accompanied by rigorous academic debate over the goals and rationales of criminal punishment. Debate has occupied many great minds from Aristotle to Bentham, Kant and many others. In the late 18th century, Jeremy Bentham argued that "general prevention or deterrence ought to be the chief end of punishment, as it is its real justification". From this, Bentham developed the classic formulation of the deterrence rationale for punishment, namely that the determination of punishment must be based on deterrence not deservedness. Bentham mentioned that "if the apparent magnitude, or rather value of the pain is greater than the apparent magnitude or value of the pleasure or good he expects to be the consequence of the act, he will be absolutely prevented from performing it". The extended and formalized version of this theory, known as "deterrence theory", was revived by Becker in 1968. The lynchpin of Becker's economic model of crime is likewise the concept of deterrence: rational

[15] For a more detailed discussion on justification of punishment see Tunick (1992).

agents, faced with high probabilities of detection and severer sanctions, will commit fewer criminal acts.

Unlike Bentham, his German contemporary Immanuel Kant was a leading proponent of the "just deserts" rationale. Kant believed that "punishment can never be administered merely as a means for promoting another good. Punishment ought to be pronounced over all criminals *proportionate* to their internal wickedness". The retributive view is "backward-looking" in the sense that it focuses on the conduct of offenders in order to determine an appropriate sanction. In contrast, the utilitarian view is "forward-looking", being concerned about the effect of punishment on the future conduct of potential offenders.

Debate over retributive justification and the various utilitarian justifications (deterrence, incapacitation and rehabilitation) has continued to divide scholars of criminal law to this day. Policy and law-making have been divided accordingly. As Robinson and Darley (1997) observes regarding the experience of the United States; "In this country we have seen repeated confusions in what we might call the public philosophy of punishment, the reasons claimed for the justification of criminal sanctions by policy-makers and legislators". Indeed, there have been shifts in penal theory, and hence in the purposes of punishment, for a complex set of reasons including politics, public policy and social movements. In a cyclic process, an early focus on deterrence as the rationale for punishment gave way to a focus on reform and rehabilitation. This, in turn, led to a return to punishment based on the notion of retribution and just deserts. Over the last quarter of the twentieth century and into the early part of the twenty-first century, retributivism has reestablished itself as the dominant theory behind criminal justice (Haist, 2009). As Robinson (2008) observed, "Deserved punishment, referred to variously as deserts, just punishment, retributive punishment or simply doing justice, has moved to center stage in the UK and is on its way in the United States, both in academic debate and in real-world institutions."[16]

The question that seems to be more important than the philosophical one is how criminal liability and punishment should be distributed within a punishment system. Who should be punished and how much? These are questions that every designer of criminal justice systems must answer, whether giving instructions to criminal code drafters or sentencing guideline drafters. Individual judges exercising discretion in interpretation of the code or in sentencing offenders must also ask these questions. When used as a distributive principle, the different purposes of punishment would give quite different distributions of punishment. Because each would distribute punishment differently, we must decide which of the competing distributive principles should prevail when they conflict. One might initially suspect that the issue of the distribution of criminal liability and punishment is as academic a question as the justification of the institution of punishment, because the two questions have commonly been combined as one. However, the truth is that the criminal justice system's distributive principle is of enormous practical importance. Indeed, it is the single most important decision to be made in the construction of any criminal justice system.[17]

This section aims to give a more detailed and formalized explanation of the confusing debate over distributive principles of punishment. To do so, we apply the social-loss function introduced in previous section. As discussed in the following sections, although the utilitarian perspective considers the harmfulness of the crime in determining optimal punishment, concerns about deterrence lead utilitarians to deviate from retributive punishment based on this criterion. It seems that criminal

[16] Indeed, the American Law Institute recently revised the Model Penal Code (the first since its promulgation in 1962) setting desert as the official dominant principle for sentencing. Courts have identified desert as the guiding principle in a variety of contexts, as with the Supreme Court's enthroning retributivism as the "primary justification for the death penalty". For more details, see Robinson (2008).

[17] For more details about this, see Robinson (2008), chapters 1 and 2.

law-making in the real world is a hybrid of utilitarian and retributive distributive principles of punishment. However, for some kinds of crimes or for particular types of criminals, some of these distributive principles have priority over others. For instance, regarding the role of highly intensive visceral factors in committing serious violent felonies and so less deterrent power of punishment, it seems that the retributive principle of "just deserts" has priority over other utilitarian principles, such as deterrence. Or in the case of imposing harsher punishment on habitual offenders, the incapacitation principle is top priority in distribution of punishment rather than other utilitarian and retributive principles of punishment.

2.3.1 Utilitarian justification for punishment

While in a retributive approach punishment is something valuable, from a utilitarian perspective punishment is considered mischief and only justified if it brings society future benefits. Normatively, these two views on punishment are so different that they can involve various social loss functions from criminal activities. These different views on punishment can be presented in a more formalized way in terms of social loss function from offences. These social loss functions are then used to explain distributive principles of punishment and their optimal magnitude for these two counterparts.

To utilitarian philosophers like Bentham, punishment can be justified if the harm that it prevents is greater than the harm inflicted on the offender through punishing him or her. According to Bentham, "Punishment is itself a mischief, or evil, since it inflicts pain, and on the principle of utility it ought only to be admitted in as far as it promises to exclude some greater evil". In this view, therefore, punishment simply adds to the totality of human suffering unless it deters further crimes in the future. In other words, utilitarians justify punishment by referring to its beneficial effects or consequences. Bentham specifically considers "*deterrence*" as the main aim of punishment:

> "Pain and pleasure are the great springs of human action. When a man perceives or supposes pain to be the consequence of an act he is acted on in such manner as tends with a certain force to withdraw him as it were from the commission of that act. If the apparent magnitude be greater than the magnitude of the pleasure expected, he will be absolutely prevented from performing it."(1789:ChIII, 61)

Indeed, Bentham justifies punishment by showing that it promotes something good in the future. Likewise for Beccaria, punishment is an evil and should only be used when it increases social utility in terms of reducing criminal activities. Beccaria gives what is essentially a deterrence theory of punishment:

> "The degree of the punishment, and the consequence of a crime, ought to be so contrived as to have the greatest possible effect on others, with the least possible pain to the delinquent — for mankind, by their union, originally intended to subject themselves to the least evils possible. The intent of punishment is not to torment or to undo past crime, but to deter future injury to society, and punishment ought to be chosen to maximize its deterrent effect."(1809:73)

Utilitarians understand punishment only as a tool for achieving an end and not as an end valuable *per se*. They perceive punishment in terms of its ability to reduce crime and do not focus on the punishment for fulfilling a deontological mandate.

The feature that distinguishes a utilitarian view on punishment from its counterpart retributive approach is that spending resources on apprehension and conviction, and specifically punishing convicted offenders, incurs loss to society in the hope of benefits in the form of a future reduction in crime. Becker (1968) tried to revive the utilitarian approach to punishment. In his viewpoint, crime and punishing offenders both impose costs on society as a whole. Indeed, social loss function from criminal activities (equation (2.1)) introduced in previous section can be considered as utilitarian social loss function.

In a nutshell, crime committed by offenders and apprehension, conviction and punishment by public authorities all incur costs for society as discussed in section 2 in detail. Now the stage is set for a discussion of social policy. The objective of public authority is to minimize the social loss function. What are its tools for doing so? The state has authority to determine the budget *C* for combating offenses, the punishment *f* per offense for those convicted, and the form of punishment, *b*. Following Becker and for analytical convenience, let *p* rather than *C* be a policy variable. Also, assume that coefficient *b* is a constant greater than zero. Then public authority decides values for policy variables *p* and *f* by optimizing the social loss function. Their optimum values are found by differentiating *L* (social loss function) with respect to *p* and *f*. The first-order conditions are as follows:

$$\begin{aligned} \frac{\partial L}{\partial f} &= D'O_f + C'O_f + bpfO_f + bpO = 0 \\ \frac{\partial L}{\partial p} &= D'O_p + C'O_p + C_p + bpfO_p + bfO = 0 \end{aligned} \tag{2.12}$$

Dividing above equations by O_f and O_p, respectively, and recombining terms gives us the following more interesting optimality conditions:

$$D' + C' = -bpf\,(1 - \frac{1}{\varepsilon_f}) \tag{2.13}$$

$$D' + C' + C_p \frac{1}{O_p} = -bpf\,(1 - \frac{1}{\varepsilon_p}) \tag{2.14}$$

Where

$$\varepsilon_f = -\frac{f}{O} O_f \tag{2.15}$$

$$\varepsilon_p = -\frac{p}{O} O_p \tag{2.16}$$

These are the familiar optimality conditions of equality of marginal revenues and marginal costs in the economic literature. To be more precise, the term on the left side of each of the optimality equations gives the marginal cost of increasing the number of offenses, *O*: in equation (2.13) through a reduction in *f* and in (2.14) through a reduction in *p*. Average revenue, given by $-bpf$, is negative but marginal revenue, given by the right-hand side of equations (2.13) and (2.14), is not necessarily negative and is positive if the elasticities ε_p and ε_f are less than one. To put it briefly, spending resources on apprehension, conviction and punishment of offenders incurs costs for the society and can also bring some benefits in the form of crime reduction. The public authority tries to balance these costs and benefits by the above optimization process. The number of optimal offences and optimal policy variables *p* and *f* can be extracted from the above equations. Figure 2.1 shows optimality conditions and their intersections that determine the optimal number of offences.

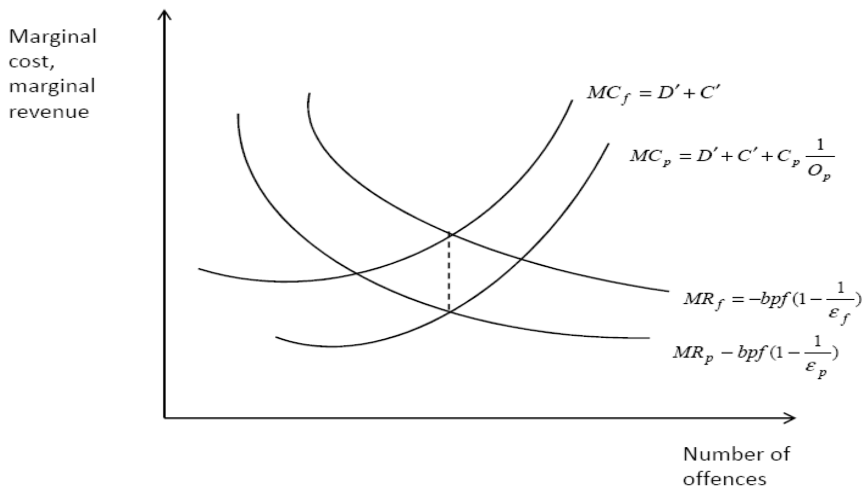

Fig 2.1- optimal number of offences

Source: Becker (1968:189)

Now consider our agenda to determine the distributive principles of punishment from a utilitarian point of view. From a technical perspective, the question becomes to determine the factors that establish the optimal amount of punishment from utilitarian perspective, hereafter f_u^*. For comparative purposes our main focus will be only on the f_u^* rather than optimal probability of apprehension and conviction (*p*).

From a utilitarian perspective, comprehensive cost-benefit analysis leads to a set of factors that determines the optimal level of punishment (f_u^*). Although utilitarians consider harmfulness of crime and judicial administrative costs in the process of optimization, their main justifications of punishment are deterrence, incapacitation and rehabilitation. Among these justifications, deterrence, the main concern of utilitarians, can be identified by ε_p (supply elasticity of crime with respect to probability of apprehension and conviction) and ε_f (supply elasticity of crime with respect to severity of punishment) in our model. These elasticities determine the slope of the Marginal Revenue curves in Figure 2.1.

Following Becker (1968) we use comparative static analysis to see how the optimal level of punishment (f_u^*) changes in response to effective factors. Let us start with degree of harmfulness of crimes. Intuitively, when a certain crime is more harmful, say murder, it incurs more costs for society. So it makes sense to spend more money on apprehension and conviction of murderers and imposing harsh enough punishment on them in the hope of deterrence (benefit) and prevention of future murders. Suppose that D' in optimality condition in equation (2.13) is positively related to an exogenous variable α where α stands for the harmfulness of a crime. Then the effect of a change in α on the optimal level of punishment can be found by differentiating the optimality condition in equation (2.13) with respect to the policy variable *f*. It is easy to show that (refer to equations A1 to A8 in the mathematical appendix):

$$\frac{\partial f_u^*}{\partial \alpha} > 0 \qquad (2.17)$$

So optimality requires an increase in punishment when the marginal cost of crime (crime's harmfulness) increases. The conclusion is that severe offences that inflict more damage on society should be punished more severely, in line with Kant's "internal wickedness". As it will be discussed in the next section, considering the harmfulness of a crime is the only similarity between retributivists and utilitarians. However, it involves a subtle difference that will also be discussed in some detail.

The first divergence between the utilitarian and retributive views is utilitarians' consideration of administrative costs of the apprehension and conviction which is given by the second term in equation (2.1). Similarly, since an increase in the marginal cost of apprehension and conviction for a given number of offences C' has the same effects as increase in marginal damage D', it reduces the optimal number of offences and increases the optimal value of punishment. In a similar way, if C' is positively related to an exogenous variable β, differentiating optimality equation (2.13) with respect to β gives the effect of an increase in β on optimal punishment as follows (for analogy see equations A1 to A8 in the Appendix):

$$\frac{\partial f_u^*}{\partial \beta} > 0 \qquad (2.18)$$

This simply shows how considering the administrative costs of apprehension and conviction of offenders can affect the optimal level of punishment. When an increase in the number of offences leads to higher judicial administrative costs, then optimality implies an increase in punishment in the hope that crime decreases, thus reducing these costs.

A more interesting point that is missing in the retributive approach is consideration of the interaction between probability of apprehension and conviction (p) and severity of punishment (*f*) that has important policy implications from a utilitarian point of view. Assume that probability of apprehension and conviction (*p*) is positively related to an exogenous variable *r*. For instance, technological progress, such as fingerprinting, which can increase the probability of conviction and apprehension of offenders, can be considered a positive shock to *p*. The effect of this positive shock to *p* on the optimal punishment can be found by differentiating optimality condition in equation (2.13) with respect to *r*. It is easy to see that (refer to equations A9 to A11):

$$\frac{\partial f_u^*}{\partial r} < 0 \qquad (2.19)$$

This simply identifies how probability of apprehension and conviction and optimal amount of punishment can interact with each other in utilitarian perspective of policy making for combating crime. Although the interaction between severity of punishment and probability of apprehension and conviction has important effects on the level of optimal punishment in the utilitarian perspective, this interaction is missing in the retributive approach.

The ultimate aim of punishment as a tool is to control crime and reduce it by deterrence, rehabilitation and incapacitation. The allocation of scarce resources for apprehension/conviction and severity of punishment crucially depend on their efficacy in crime reduction, indicated by crime supply elasticity with respect to *p* and *f* in our model. From a utilitarian perspective, punishment incurs huge costs for society expressed by the third term in equation (2.1). These costs besides the administrative costs of apprehension and conviction, must be compensated by the benefits of deterrence (general and individual), incapacitation and rehabilitation to justify punishment, otherwise they just add to the total suffering of the society. In this framework, both ε_p and ε_f as indicators of deterrence have a crucial role. Rearrange ε_f as $E_f = \frac{1}{\varepsilon_f} = \frac{-\partial f}{\partial O}\frac{O}{f}$ and suppose E_f is positively related to exogenous variable γ. Then the effect of an increase in γ, i.e. a decrease in crime supply elasticity with respect to punishment can be easily found by differentiating optimality condition in equation (2.13) with respect to γ. It is easy to see that (refer to equation A12 to A13 in Appendix):

$$\frac{\partial f_u^*}{\partial \gamma} < 0 \qquad (2.20)$$

Equation (2.20) indicates that when the supply of crime is inelastic with respect to the severity of punishment, i.e. severity of punishment has little deterrent effect on the behavior of potential offenders, optimality suggests reducing the severity of punishment.[18]

In a nutshell, various factors can affect the optimal amount of punishment in the utilitarian perspective. If we assume that *p* is fixed (for simplicity and comparative purposes)[19] then these factors are the degree of harmfulness of crime (D'), the marginal cost of apprehension and conviction for a given number of offences (C'), the level of probability of apprehension and conviction (*p*) and crime supply elasticity

[18] Another feature of interaction between policy variables of *p* and *f* in the utilitarian perspective on combating crime can be specified when we relax the simplifying assumption of a constant *p* that we made in deriving of our main equations in our comparative static analysis. In this case, the optimal levels of *p* and *f* can also interact through their own crime supply elasticities. Relax the assumption of *p* being fixed and let policy variables *p* and *f* be determined simultaneously. If ε_f and ε_p are negatively related to γ and μ, respectively, then $\frac{\partial f_u^*}{\partial \mu} > 0$ and $\frac{\partial p_u^*}{\partial \gamma} > 0$ which again implies the interaction between policy variables *p* and *f* in utilitarian approach to criminal policy-making that is completely missing in the retributive perspective.

[19] Relaxing this assumption, the level of *P* (as exogenous variable) is replaced by crime supply elasticity with respect to probability of apprehension and conviction (ε_p) and $C_p = \frac{\partial C}{\partial p}$ in the list of determinant factors of optimal amount of punishment in equation (2.21). In this case we have $\frac{\partial f_u^*}{\partial \varepsilon_p} < 0$ and $\frac{\partial f_u^*}{\partial C_p} > 0$. For policy variable *p* we also have $\frac{\partial p_u^*}{\partial \varepsilon_p} > 0$, $\frac{\partial p_u^*}{\partial \varepsilon_f} < 0$, $\frac{\partial p_u^*}{\partial C_p} < 0$. All other derivatives in equation (2.22) remain unchanged. These are also other features of interaction between policy variables of *p* and *f* in utilitarian approach that is completely missing in retributive one.

with respect to severity of punishment (ε_f). We can summarize them in the following function:

$$f_u^* = f(D', C', p, \varepsilon_f) \tag{2.21}$$

Where

$$\frac{\partial f_u^*}{\partial D'} > 0 \qquad \frac{\partial f_u^*}{\partial C'} > 0 \qquad \frac{\partial f_u^*}{\partial p} < 0 \qquad \frac{\partial f_u^*}{\partial \varepsilon_f} > 0 \tag{2.22}$$

Clearly, it is necessary to have information about harmfulness, judicial system cost function and especially efficacy of punishment in preventing future crimes in order to determine the optimal amount of punishment in this framework. The most challenging of this information is ε_f (and ε_p when f and p are determined simultaneously). By and large, critics of the utilitarian approach mostly challenge the deterrent power of punishment and the amount of this elasticity.

2.3.2 Retributive justification for punishment

In the retributive approach, punishment is justified because it is deserved regardless of whether or not it deters future crime. The perpetrator deserves to be punished for past harm done. Punishment is a valuable end in itself and needs no further justification. The major goal of this approach is to make the offender take responsibility for the suffering, pain or loss inflicted on victims by paying for injustice to society. Proponents argue that a wrong must be made right, and that the offender must pay his debt to society. So the main distributive principle of punishment is blameworthiness and deservedness of offenders.

There are two subtle points that distinguish the retributive and utilitarian approaches. First of all, retributivists seem reluctant to include offenders' gain from criminal activities in the social welfare function, only being concerned about harm inflicted on victims, which in their view makes society imbalanced.[20] Secondly, regarding punishment as a valuable end *per se* implies that the process of punishing convicted offenders does not incur losses for society. In other words, retributivists disregard the last two terms in the social loss function (equation 2.1). If gains from criminal activities are wrong and must be excluded from the social welfare function, while enforcement of punishment on offenders is good and does not incur loss for society, the retributive approach social loss function from criminal activities is simply:

$$L = H(O(p, f)) \qquad (2.23)$$

Where H is harm inflicted by offenders on victims. Now the key question is how the imposed punishment is determined in this perspective for different kinds of crimes. At one hand, if the gains from harmful conduct were excluded from social welfare, the main consequence for our analysis would be that, for this type of conduct, society would want to achieve greater, possibly complete, deterrence. That, in turn, would tend to make a higher sanction and a higher probability of detection desirable. At the other hand, according to Kant, punishment ought to be "pronounced over *all* criminals *proportionate* to their *internal wickedness*". In other words, there is a well-known "proportionate principle" in this perspective[21]. Systems of retribution for crime have long existed, the best known being the "lex talionis"[22] of Biblical times, calling for "an eye for an eye, a tooth for a tooth, and a life for a life" (Hudson, 1996). Retributivists claim a moral link between punishment and guilt, and see punishment as a question of responsibility or accountability (Bean, 1981). According

[20] For a more detailed discussion of valuation of criminal gains in welfare economics and cost-benefit analysis, see Lewin & Trumbull (1990) and Stigler (1970).
[21] For example Rawls offers the following characterization of "proportionate principle":
"What we may call the retributive view is that punishment is justified on the grounds that wrongdoing merits punishment. It is morally fitting that a person who does wrong should suffer in proportion to his wrongdoing. That a criminal should be punished follows from his guilt, and the severity of the appropriate punishment depends on the depravity of his act. The state of affairs where a wrongdoer suffers punishment is morally better than the state of affairs where he does not; and it is better irrespective of any of the consequences of punishing him". (Rawls 1955)
[22] The basic principle of "lex talionis" is that punishment should inflict the same on the offender as the offender inflicted on his or her victim. Surely this principle, though requiring likeness of punishment, does not require exact likeness in all respects.

to Robinson (2008), the modern concept of "just deserts" includes three different conceptions; vengeful desert, deontological desert and empirical desert. The above "proportionate principle" and optimal punishment have different interpretations in these three conceptions. In other words, the optimal punishment may be different under the different conceptions of desert.[23]

Let us start with minimization of the retributive social loss function in equation (2.23) with respect to policy variables p and f to determine the retributive prescription for combating crime. In this case, the two optimality conditions (2.13) and (2.14) reduce to the same simple condition:

$$H'(O) = 0 \qquad (2.24)$$

This equation is equivalent to a punishment that sets marginal harm at zero. In other words, when there is no concern about the costs of apprehension, conviction and punishing offenders and when offences harmed society ("created society imbalances" in retributive terminology), the social loss function from offences is minimized by setting both probability of apprehension and conviction and severity of punishment high enough to eliminate all offences. This is equivalent to saying that all offenders should be caught and punished to make them pay back their debt to society. However we also have to consider the proportionality of harm inflicted on the victim and punishment. The optimal amount of punishment observes the "proportionate principle", with severe offences punished harshly, according to Kant's "internal wickedness". Regarding the probability of apprehension and conviction, the retributive approach seems to suffer from idealism in considering that all offenders can be caught and punished, i.e. $p=1$. This is called "positive retributive principle[24]" which emphasizes on the punishing of all offenders. In contrast to the utilitarian approach, there is no interaction at all between the probability of

[23] Here, I just consider the difference between diverse notions of desert briefly and from a technical point of view in order to clarify the different optimal punishment under each of them. For more details about these notions, see Robinson (2008).
[24] For more detail see Tunick (1992).

apprehension/conviction and the severity of punishment. As severity of punishment depends on harm inflicted on society, the probability of apprehension/conviction seems to be utopian rather than pragmatic.

The retributive approach is not concerned about the efficacy of punishment for future crime reduction. However, deterrence effects of punishment through its censoring function are acknowledged but deterrence is not a distributive principle of punishment as it is in utilitarian perspective. Retributivists concentrate on the harmfulness of crime and on imposing a proportionate punishment. So any change in punishment from the well-known "proportionate principle" motivated by the principle of deterrence can be considered as deviation from the retributive principle of punishment. Furthermore, there are differences between the different notions of desert that affect constraints set on optimal punishment. The major difference between vengeful desert and the other two conceptions of desert is the importance of the absolute amount of punishment in the former. Vengeful desert centers on this absolute amount: it must be equal in amount, if not also in the means, to the suffering caused by the offence, whereas for deontological and empirical deserts, the absolute amount of punishment is of limited interest. Their central concern is the relative amount of punishment among cases with differing degrees of moral blameworthiness. These latter conceptions of justice focus primarily on ensuring that the offender is given that amount of punishment that puts him in his proper ordinal rank among in relation to blameworthiness (Robinson, 2008).

Once a society has committed itself, as all societies must do, to a particular endpoint in its punishment continuum - be it the death penalty, life imprisonment, fifteen years imprisonment, or something less - the ordinal rank of any given case necessarily converts to a specific amount of punishment: that amount of punishment that assigns the offender his appropriate ordinal rank. But for deontological and empirical deserts, the absolute amount of punishment has no other significance. If the endpoint of the punishment continuum changes, the amount of punishment that an offender deserves under these two conceptions of justice also changes to the amount necessary to keep the offender in his proper ordinal rank (Robinson, 2008). Mathematically, for all conceptions of desert, the authority minimizes the loss function subject to policy variables and a constraint on the endpoint of punishment. Indeed, this would be a kind of constrained optimization. Indeed, deontological and empirical deserts are a somewhat civilized version of vengeful deserts or "lex talionis". The subtle difference between deontological and empirical deserts is the source of determination of the endpoint of punishment. While in deontological desert, moral philosophers have critical role in determining the endpoint, empirical desert gives more weight to the community's intuition of justice or simply the harshness or leniency of deserved punishment. (Robinson, 2008).

Although these different notions of desert can lead to different optimal punishment, they have something in common that completely distinguishes them from the utilitarian approach; they are more concerned about blameworthiness and being justified in imposing punishment on offenders, rather that caring much about the final consequences of using punishment as a means of crime control and about the costs of the criminal justice system, from apprehension/conviction to punishment of convicted offenders.

Furthermore, determination of optimal punishment from a retributive view requires less information than the rival utilitarian approach. The retributive conception of punishment includes the notion that the purpose of punishment is to lay moral blame on the offender for the crime, while his or her future conduct does not concern the punishment decision. This is why retributivists do not care whether the shift in the marginal cost curve of Figure 2.1 changes the number of offences.

2.3.3 Retributivism versus Utilitarianism

The debate over distributive principles of punishment of "just deserts" and the various utilitarian principles such as deterrence, incapacitation and rehabilitation has continued to divide criminal law scholars to this day. This section aims to shed more light on potential irreconcilabilities between retributivists and utilitarians using the model presented in previous sections. Probably all practices have or once had a use or function, but whether the practice is justified by some principle of utility is another matter. In other words, although among the above-mentioned utilitarian principles of punishment, deterrence and incapacitation can also be considered by-products of the retributive principle but the optimal punishment may be different when they are regarded as distributive principles for punishment by utilitarians. For instance, imprisonment as a form of punishment, whether motivated by retribution or utility, prevents offenders from committing more crimes (at least outside the prison). However, the length of imprisonment as optimal punishment may be different in these approaches. Considering deservedness or deterrence as distributive principle for punishment can lead to different periods of imprisonment in this case. In other words, although deterrence, for example, can be a by-product of punishment in the retributive approach, it is not a distributive principle of punishment at all. It is what retributivists call *censure* which is also an important component of retributive thinking.[25]

Having different distributive principles for punishment leads to irreconcilabilities between these two approaches. What are the sources of conflicts between retributivism and utilitarianism? The central distributive principle of punishment from a retributive approach is deservedness on the basis of the "proportionate principle" or "internal wickedness". Some utilitarian principles distribute punishment in a way that deviates from prescribed punishment based on the "proportionate principle". Caring about the efficacy of punishment in reducing future crime

[25] For example, Andrew von Hirsch, a leading theorist of just deserts sentencing believes: "Desert and punishment can rest on a much simpler idea: the idea of censure. Punishment connotes censure. Penalties should comport with seriousness of crimes so that the reprobation on the offender through his penalty fairly reflects the blameworthiness of his conduct".

(deterrence) and considering the costs of punishing offenders are two major sources of conflict between retributivists and utilitarians.

For a more formal version of these irreconcilabilities, let f_r^* be optimal punishment based on the “proportionate principle” from a retributive perspective. Now let us consider the probability of apprehension and conviction of offenders. As mentioned in previous sections, when the aim is simply to make "the punishment fit the crime," *p* could be close to 1. Such a policy ignores the social costs of increases in *p* and *f*. In contrast, the utilitarian approach takes the administrative costs of the judicial system into account and tries to optimize them by creating an interaction between *p* and *f* so that the expected punishment (*pf*) to be severe enough to deter potential offenders. As discussed before, this interaction is missing in the retributive approach and is a major source of conflict between them and utilitarians.

For convenience of analysis, let us assume that retributive optimal punishment (f_r^*) based on the “proportionate principle” is simply equal to harm (*h*) inflicted on the victim.[26] In contrast, from a utilitarian perspective considering both inclusion of offenders’ gain in social welfare function and interaction between *p* and *f*, we will have $f_u^* p = d < h$ which leads to an optimal level of punishment of $f_u^* = \frac{d}{p} < h$ from a utilitarian perspective.[27] Now let *I* be the ratio of retributive optimal punishment to utilitarian optimal punishment:

$$I = \frac{f_r^*}{f_u^*} = \frac{h}{d/p} = p(\frac{h}{d}) \qquad (2.25)$$

Or

$$f_u^* = f_r^* \frac{1}{p}(\frac{d}{h}) \qquad (2.26)$$

[26] For convenience ignore the mentioned difference between diverse conceptions of deserts.
[27] For more details on such level of optimal punishment which is less than harm inflicted, see Polinsky & Shavell (2000).

The diagram in Fig 2.2 depicts the utilitarian's iso-expected punishment along which expected punishment is constant and equal to *d*.[28]Now, we can have better image of deviations of utilitarian optimal punishment from retributive one by means of this diagram. This deviation stems from setting deterrence as the main distributive principle of punishment in the utilitarian approach and available interaction between severity of punishment and probability of apprehension and conviction.

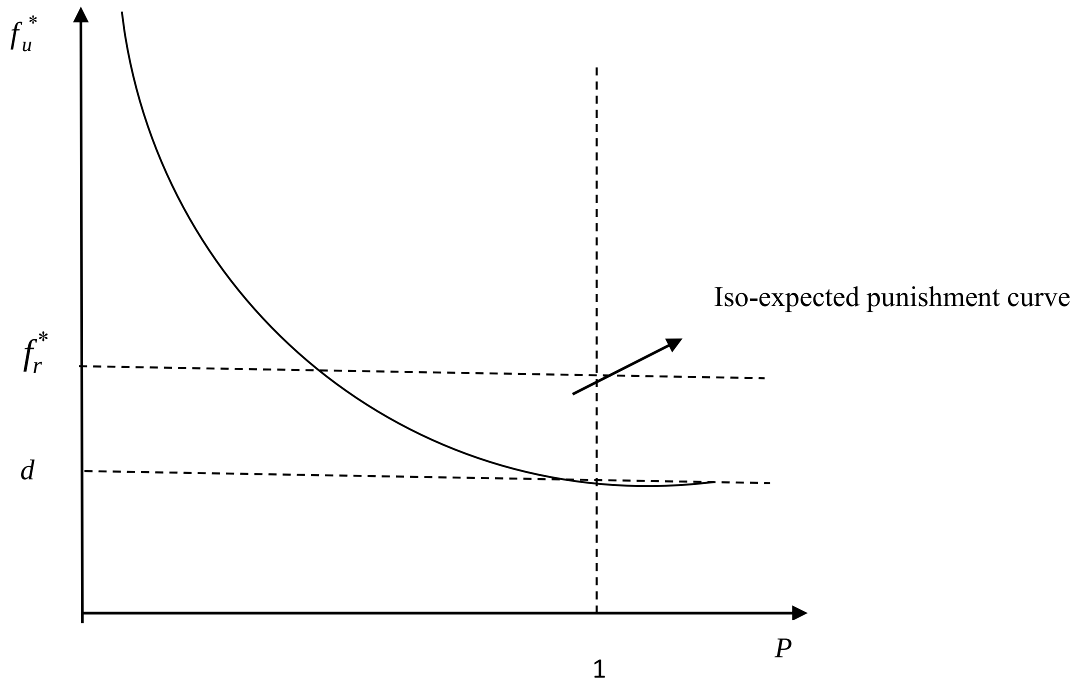

Figure 2.2 - Interaction between policy variables *p* and *f* and utilitarins' iso-expected punishment

As already mentioned, what distinguishes the utilitarian approach to punishment from the retributive one is the interaction between policy variables *p* and *f*. If we assume that *p* is fixed in utilitarians' optimization then it is easy to see from equation (2.25) that:

[28] An increase in net harm of crime that presumably holds for more serious felonies would shift the iso-expected punishment curve up and rightward, which implies higher expected punishment for more serious felonies. The horizontal line f_r^* would also shift to up for more harmful crimes to fulfill the well-known "proportionate principle" in a retributive perspective.

$$\frac{\partial I}{\partial p} = \frac{h}{d} > 0 \qquad (2.35)$$

This indicates that due to interaction between *p* and *f* in the utilitarian approach (in the case of fixed *p,* this interaction is specified by derivative $\frac{\partial f_u^*}{\partial p} < 0$, see equations (2.19) and (2.22)), an increase in probability of apprehension and conviction leads to decrease in utilitarian optimal punishment and consequently increase in *I*, a measure of deviation of f_u^* from f_r^*. Indeed, whatever affects f_u^* in equation (2.22), affects *I* too.

If we assume that *p* and *f* are determined simultaneously during optimization of the social loss function, then the interaction between *p* and *f* will have two features: the relative efficacy of severity of punishment and apprehension/conviction in reducing crime (the relative amounts of ε_p and ε_f) and the relative costs of severity of punishment and of implementing apprehension and conviction (see equations (2.21), (2.22) and footnote 19). These features together determine which point along Iso-expected punishment curve is chosen by the public authority.

Now let us see how setting deterrence as the main distributive principle and caring about relative costs of severity of punishment and implementation of apprehension/conviction by interaction of policy variables *p* and *f* lead to deviation of utilitarian from retributive optimal punishment, which is simply based on the "proportionate principle".

Let us start with crime supply elasticity ε_p and ε_f with respect to policy variables *p* and *f*, respectively. When the supply of a certain kind of crime is elastic with respect to severity of punishment (higher amounts of ε_f) then optimality implies harsher punishment in order to deter potential offenders and spend less money on apprehension and conviction ($\frac{\partial f_u^*}{\partial \varepsilon_f} > 0$, $\frac{\partial p_u^*}{\partial \varepsilon_f} < 0$). All other things being equal, if we start from point A in Figure 2.3 which indicates retributive optimal punishment, an increase in efficacy of severity of punishment in controlling future crime increases the deviation from retributive optimal punishment towards harsher punishment (from A to B in Figure 2.3). In this case, from a utilitarian point of view, it is optimal to allocate more resources to severity of punishment than to apprehension and conviction of offenders. The inverse holds when supply of crime is inelastic with respect to severity of punishment.

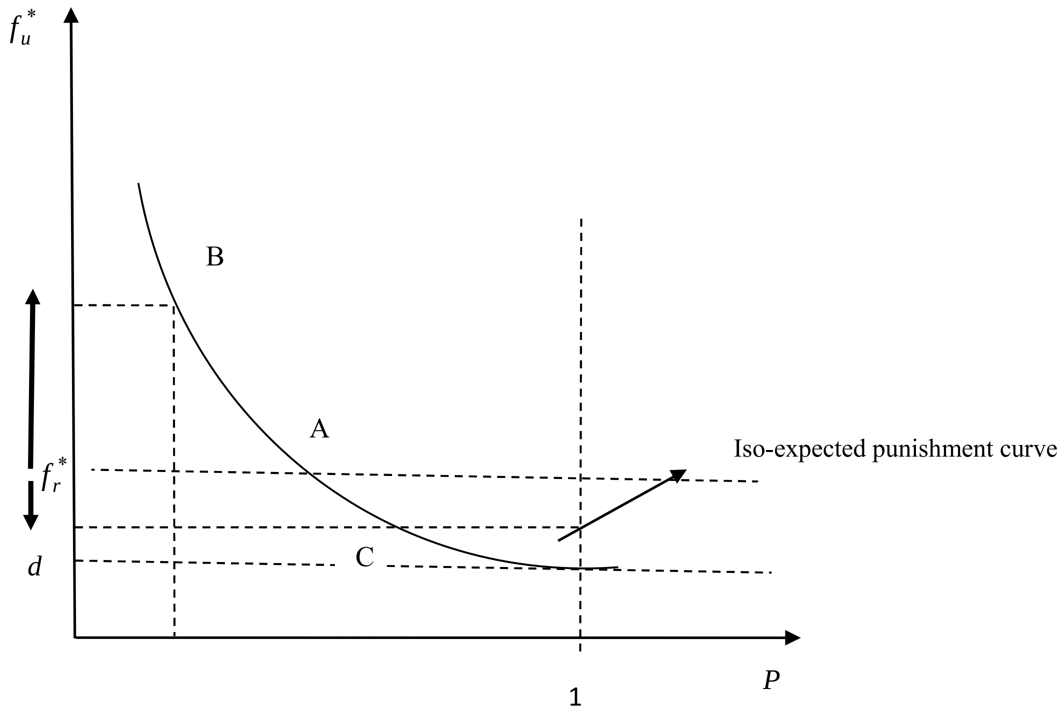

Figure 2.3 - An increase in efficacy of severity of punishment in crime control makes utilitarian optimal punishment harsher than retributive punishment (movement along iso-expected punishment from point A to point B).

An increase in efficacy of apprehension and conviction in crime control makes utilitarian optimal punishment more lenient that retributive punishment (movement along iso-expected punishment from point A to point C).

In the case of higher efficacy of apprehension and conviction in controlling crime, all other things being equal, optimality implies investing more resources in police and courts to increase the apprehension and conviction level and decrease severity of punishment ($\frac{\partial f_u^*}{\partial \varepsilon_p} < 0$, $\frac{\partial p_u^*}{\partial \varepsilon_p} > 0$). All other things being equal, starting from point A in Figure 2.3, an increase in ε_p deviates utilitarian optimal level of punishment from deserved punishment (f_r^*) towards more lenient punishment (from point A to C in Figure 2.3). The inverse holds when the crime supply is inelastic with respect to apprehension and conviction.

Finally, besides the efficacy of expected punishment in crime control that leads to deviation from retributive optimal punishment, caring about the running costs of expected punishment (second and third terms in the social loss function expressed by equation (2.1)) can also lead to deviation. Regarding the assumption that *b* is a constant greater than zero, if the marginal cost of administrative costs of apprehension and conviction ($C_p = \frac{\partial C}{\partial p}$) increases, i.e. higher levels of apprehension and conviction are more expensive, then optimality requires an increase in punishment and its deviation from retributive punishment ($\frac{\partial f_u^*}{\partial c_p} > 0$, $\frac{\partial p_u^*}{\partial c_p} < 0$). The point is that being concerned about the relative costs of severity of punishment and administrative costs of apprehension and conviction can be a potential source of irreconcilability between utilitarians and retributivists.

So, in contrast to the retributive approach in which punishment for a certain kind of a crime is simply determined by the harm inflicted on victims ("proportionate principle"), the utilitarian approach considers many factors and uses deterrence as the main distributive principle of punishment. Consideration of the relative efficacy of severity of punishment and levels of apprehension/conviction in reducing future crime and the associated running costs of a certain level of expected punishment can cause utilitarian optimal punishment to deviate from what retributivists believe to be deserved punishment.

2.3.4 Retributivism and Certainty of Punishment

As mentioned in previous sections, according to the positive retributive principle, the punishment must be run over all deserved offenders. Those holding to the positive retributive principle find selective enforcement unacceptable, not merely because it is unfair, but because it is premised on the mistaken view that it is not inherently wrong to allow some crimes to go unpunished. Pragmatically, it might not be feasible to run punishment over all criminals. The certainty of punishment crucially depends on the government resources and is unlikely to be ever 1. Indeed, the normative positive retributive principle has been criticized by utilitarianism which mention it is inefficient to make P=1 and always punish[29]. In this section we try to see how presumably in retributive approach caring about administrative costs of justice system would change optimal punishment prescribed by retributivists. It will be showed how appropriate is the retributive social loss function and its optimal punishment introduced in previous section. We will show that caring about administrative costs of punishment (second term in utilitarian social loss function in equation 2.1) would lead to a deviation from well-known retributivism's "proportionate principle". Let's redefine retributive social loss function from criminal activities as below that includes administrative costs of punishment too:

$$L = H(O(p, f)) + C(p, O(p, f)) \qquad (2.36)$$

First, assume *p* is a given constant (less or even equal to one) and a retributive authority tries to minimize the social loss function subject to only punishment (*f*). Then the optimality condition is:

[29] For example see Kaplow & Shavell (2006), chapter 6.

$$\frac{\partial L}{\partial f} = H' + C' = 0 \qquad (2.37)$$

Which interestingly implies the optimal punishment should be higher than inflicted harm (internal wickedness). So in this case, $f_r^* > h$ which can be interpreted as a deviation from what seems to be deserved punishment in retributive approach.

Now assume the retributive authority can also determine the certainty of punishment and so minimize his social loss function subject to both *p* and *f*. Then another optimality condition will be added to the above optimality condition in equation (2.37):

$$\frac{\partial L}{\partial p} = H'O_p + C'O_p + C_p = 0 \qquad (2.38)$$

In this case again the optimal punishment would be higher that inflicted harm (*h*) and even higher than the one prescribed when *p* was assumed to be fixed. Both cases imply that caring about administrative costs of justice system would lead to a deviation of optimal punishment from what is believed to be based on "proportionate principle". So it seems that social loss function and derived optimal punishment for retributivism presented in previous sections to be much more similar to what retributivists concern.

2.3.5 Conclusion: Hybrid Distributive Principles

Both the utilitarian and retributive perspectives on punishment justify the existence of the institution of punishment. Although they achieve some common aims, such as deterrence and incapacitation, choosing "just deserts" or "deterrence"[30] as the main

[30] In the model presented here, the main focus is on deterrence as the main utilitarian justification for punishment. Setting other utilitarian justifications for punishment, incapacitation or rehabilitation, as the main distributive principles for punishment would also lead to a different distribution of punishment. For example, when incapacitation is set as the main distributive principle of punishment,

distributive principle of punishment can lead to different distributions of punishment on convicted offenders. Consequently, prescribed punishment by utilitarians (regarding deterrence as the main distributive principle of punishment) can deviate from prescribed punishment by retributivists based on deservedness.

Criminal law-making in the real world seems to be a hybrid of utilitarian and retributive principles of punishment distribution. However, it seems that the priority given to each of these distributive principles, either utilitarian (deterrence, incapacitation and rehabilitation) or retributive, has changed remarkably in the course of time. As Haist (2009) mentions the early focus on deterrence as the rationale for punishment gave way to a focus on reform and rehabilitation. He argues that over the last quarter of the twentieth century and into the early part of the twenty-first century, retributivism has reestablished itself as the dominant theory behind criminal justice. It seems that depending on the kind of crime or characteristics of offenders, any distributive principle can be the primary one and have priority over the others. Let us illustrate this with some examples.

Consider murder or rape as the most serious felonies. Visceral factors generally seem to be more influential in violent crimes than in non-violent crimes. When visceral factors such as anger or sexual desire are intense, punishment is less deterrent.[31] Thus, setting deterrence as the main distributive principle for punishment implies a more lenient punishment for murderers than if punishment were decided on the basis of "just deserts". Bentham argues that we should not punish in cases where this would be ineffective, unprofitable or too expensive. Some retributivists would reply that justice demands that we punish even in such cases. However, in reflecting on such cases, Bentham is not saying that we ought not to punish, but rather that punishing would be of no use, "though it should be inflicted." This might indicate that he recognizes implicitly that there is some ethical demand for punishing offenses,

---

the dangerousness of convicted offenders is the main factor in sentence determination rather than deservedness or deterrence.

[31] For more a comprehensive debate on how visceral factors influence criminal behavior see Chapter1.

regardless of the bearing on utility of actually inflicting punishment in these cases. Perhaps Bentham's response to the retributivists' objection that it is unjust not to punish someone who commits an offence would be to say that we should always punish but we should adjust its level (more lenient punishment), thereby satisfying both the demands of retribution and utility (Tunick, 1992). In reality, murderers and rapists are punished severely in most societies, say life sentence or even capital punishment for murder. For serious violent crimes, "just deserts" seems to have priority over other distributive principles of punishment. The retributive approach dominates not only because of the moral credibility of the punishment but also because punishment can partly achieve some utilitarian aims as well. Regarding deterrence, the censure expressed through criminal law has the role of giving third parties reasons for not committing acts defined as criminal. In other words, censure can have a deterrent effect. This deterrent effect can be affected by the moral credibility of the justice system, as proponents of the censure function of punishment admire it. For instance, Robinson & Darly (1997) write:

> "The criminal law's moral credibility is essential to effective crime control, and is enhanced if the distribution of criminal liability is perceived as ''doing justice,'' that is, if it assigns liability and punishment in ways that the community perceives as consistent with the community's principles of appropriate liability and punishment. Conversely, the system's moral credibility, and therefore its crime control effectiveness, is undermined by a distribution of liability that deviates from community perceptions of just desert".

With regard to non-violent crimes, potential offenders seem to have more capacity for calculation and cost-benefit analysis in deciding to commit a crime or not commit. There are many examples, from tax evasion to not paying on public transport. In these cases, it seems that punishment has more deterrent power and deterrence has priority over other distributive principles and can be considered as the main distributive principle for punishment.[32]

Apart from the nature of crimes, in some cases it is the characteristics of offenders that can give priority to a certain distributive principle of punishment, such as incapacitation or rehabilitation. Consider “repeated offenders”: in many cases conviction records are considered an escalating factor that increases the length of a sentence. In this case, incapacitation seems to be considered the main distributive principle of punishment. Three Strikes Law in the US is a specific example. This law significantly increases the prison sentences of offenders previously convicted of two or more violent crimes or felonies. Depending on the seriousness of the current and prior crimes committed by the offender, the sentence can range from a minimum of 25 years to a maximum of life imprisonment. Indeed, the use of incapacitation as the main distributive principle of punishment is based on forecasting the public threat represented by offenders. A bad past record is a signal of a habitual offender and justifies incapacitation as an instrument for crime control. Another feature of offenders that can affect the distribution of punishment seems to be demographic. For young offenders, rehabilitation seems to have some priority over other distributive principles of punishment.

2.4 Comparative Criminal Law: An Economic Perspective

Criminal law provides a code of conduct that all members of society are expected to follow. It is enforced by the state, and violations are considered acts against the state and against individual victims. In other words, criminal law is a formal means of social control, designed to regulate human behavior and interaction, enforced by representatives of the state: police, courts and correction agencies.

[32] For a comparative analysis of efficacy of expected punishment in reducing different kinds of violent and non-violent crimes see Chapter1.

According to Dubber (2006), criminal law has traditionally been the most parochial of legal disciplines. Dubber believes that if a state's criminal law were subject to critique, that critique would have to be entirely internal, i.e. framed in terms of principles, concepts, and practices constituent of the state itself. He adds that any reference to external law is at best beside the point, at worst offensive. He claims that there are no choice-of-law problems in criminal law, because the notion of a sovereign applying the criminal law of another is preposterous.

As a matter of fact, different jurisdictions define criminal acts and penalties differently. Today, criminal law is codified. One consequence of codification is that every country goes its own way. Every country has its own conception of punishable behavior, its own definitions of offences, its own principles for determining questions of self-defense, necessity, insanity, negligence, and complicity. Criminal law has become state law, parochial law. However, there are also many similarities across jurisdictional boundaries. Despite the parochialism of criminal law, the aspiration of harmonizing criminal law is stronger today.

Traditionally, comparative law is thought to be comparison of one country's law with that of another. Many distinguished comparative lawyers have insisted on the virtues of comparative law as a means of expanding knowledge generally and of achieving a better understanding of law. On a purely cognitive level, comparative law should therefore contribute to a better understanding of one's own national law through the contrasts and greater range of information it provides. As a practical consequence, it should be a primary instrument of domestic law reform through legislation, and indeed, the need for comparative research is a constant theme on the agenda of law reform agencies and ministries of justice. It has been said, however, that comparative law is a means of evaluating the efficiency of different models as well as recognizing the resistance of particular national traditions (the language of 'path dependency' is often used) in the 'evolution' towards efficiency. A larger pragmatic objective seems to be the regional or international harmonization of law, of great importance today within Europe, but also in the worldwide process of development of international and transnational law. Despite these ambitions for harmonization, development of a "common law of humanity" mentioned by some commentators is considered utopian today (Gellen 2006).

Besides traditional comparative law, another academic movement called "Comparative law and economics" explains the convergence of legal systems as a movement towards efficiency. Indeed, legal systems or components of them produce different legal doctrines or techniques for the solution of a given problem. All these different inputs enter what we may call the market of legal culture. In this market, suppliers meet the needs of consumers. This process of competition may determine the survival of the most efficient legal doctrine. However, "comparative law and economics" has always made it clear that divergences between legal systems do not necessarily imply inefficiencies. From the very beginning it was admitted that "different legal traditions may develop alternative solutions for the same legal problem that are neutral from the standpoint of efficiency" (Mattei, 1994). Nevertheless, especially in its initial stage, comparative law and economics assumed that "the more efficient legal theories and solutions would spread around in a world with zero transaction costs": in such a world, efficient legal solutions would survive, while inefficient legal solutions would disappear" (Ajani and Mattei, 1995). In actual fact, however, legal systems have very high transaction costs. Such costs, opposing the spread of efficient laws, are created by the so-called legal traditions or by legal parochialism. For instance, an efficient English solution will find its way into France with difficulty because the two systems are built on profoundly different traditions. Transaction costs may be created by "prestige", precluding transplants from "non-prestigious" legal systems. In its normative dimension, comparative law and economics may be of help, acting as a prestigious support to legal systems that have already reached an efficient solution without having the internal strength to export it (Caterina, 2006). Some years ago, Thomas Ulen (1997) emphasized the need for law and economics scholars to turn their attention to comparative law, in order to develop a "unifying theory" that would make a consistent and coherent whole of the different national legal systems with their convergences and divergences.

Criminal law is regarded as deeply rooted in a country's social mores and cultural preferences, defying transnational assimilation and harmonization. The purpose of studying foreign systems of criminal justice has therefore been more educational than practical: looking at other nations' criminal laws was thought to help understand one's own laws and put them into context; comparison could also demonstrate the relativity of legal solutions to social problems (Wiegend, 2006).

Nonetheless, there are observers, such as Jescheck (1974) and Eser (1998), who see the main function of comparative law in informing and promoting law reform (Weigend, 2006). As Alan Watson (1981) observed, "law develops mainly by borrowing". Looking beyond national borders broadens the supply of possible solutions for social or legal problems, which, in a world of global assimilation, tend to show the same or similar traits. Despite the methodological drawbacks involved in borrowing ready-made legal concepts from foreign systems, there still exist successful transplants stories. One of them is introduction of the "day fine" system to various other European countries which was originally developed in Scandinavia (Hillsman, 1990).

Since the 1980s and 1990s, the idea of internationalization of criminal law has created a large new market for comparative criminal law. Several criminal law and criminal procedure developments are responsible for its increased relevance. At the same time, protection of and respect for basic human rights, including those concerning the criminal process, have become major themes in international discourse and politics, initiating a strong movement towards the recognition and implementation of common minimum standards of criminal justice. However, as the Austrian pioneer of comparative criminal law Franz von Liszt predicted more than a century ago, the aim of comparison is not the establishment of universal common law in criminal matters (Wiegend, 2006). Indeed, attempts to explain legal transplants from one system to another have relied largely on what Mattei (1994) calls the empty idea of "prestige." This shortcoming is due to the fact that comparativists who have been working on legal transplants are less interested in theoretical explanations of why legal borrowing happens than in observing its occurrence. An aim of comparative law and economics is to fill this theoretical gap.

The field of penal law has three aspects: substantive criminal law or criminal law proper (which concerns itself with the general principles of criminal liability, in its general part, and specific offense definitions, in its special part), criminal procedure (which deals with the imposition of the general and specific norms of substantive criminal law in particular cases), and the law of punishment execution and sanctions (which covers the quantity and quality of sanctions for violation of criminal norms, as well as their conditions of application to convicted offenders). In the case of punishment, common law and civil law systems both generally operate with the same palette of rationales: retribution, general and specific deterrence, incapacitation and rehabilitation.

Among above aspects of penal law, punishment appears to have received much less attention in comparative research. This may be a result of the fact that punishment is the final stage in the system of criminal justice and is therefore seen as determined by prior stages in the criminal justice process. Such a conception, however, seems to underestimate the impact of the dynamics of punishment. Punishment offers many opportunities for comparative reflection, and it is on this that this section is centered, as well as on specific offense definitions.

As a matter of fact, there is some diversity in the scope of criminal law in different societies. While a range of activities, say same-sex marriage and abortion, are considered legal in some societies, they are illegal in other societies. Different societies impose different punishment, even for a crime such as murder. National legal systems, moreover, have undergone important transformations over the course of history. A controversial example is the death penalty in the criminal justice system.

While most countries have abolished the death penalty for any crime, there are still some countries that impose death penalty for certain crimes. For example Japan and the United States, as examples of modern democracies, impose the death penalty for murder. In US, there is capital punishment in 36 states, as well as in federal law. The domain of the death penalty also differs among these countries. In some it is only imposed for murder. In others it can also be imposed for other violent crimes such as rape or even non-violent crimes such as drug offences and adultery. What is the reason for these different perspectives in the application of capital punishment (for instance) in different societies? Why do different societies have different criminal law?

Short of endeavoring to work out a unitary theory of the causes of legal transplants or even a systematic discussion of patterns of legal change or explanation of divergence in criminal justice systems worldwide, this section tries to push legal scholarship a step forward in this field. This section develops a formal model to shed light on the diversity of criminal law in different societies and its evolution in the course of history. Its emphasis is on the reasons for this current diversity in the body of criminal law and on its evolution in a more formalized way.

2.4.1 Criminal law-making: an economic perspective

From an economic perspective, criminal law-making can be considered a process in which the state or public authority tries to maximize social welfare or equivalently minimize social loss from criminal activities. Here we use the principle of social loss minimization as benchmark for criminal law-making. We use Becker's social loss function from criminal activities, introduced in section 2.2.

Criminal law-making has two main stages: i) defining the domain of criminal activities and ii) determination of optimal expected punishment by minimizing the above-mentioned loss function. Some features of this loss function seem to be normative enough to change the scope of criminal activities and to prescribe different optimal levels of punishment for a given crime. What are these seemingly normative features? We will try to specify them formally through the above social loss function and show how having different normative perspectives can lead to different scope of criminal law and even different punishment for certain crimes. Interestingly, this model will also shed light on the historical evolution of criminal law, and specifically punishment from an efficiency point of view.

2.4.2 The scope of criminal law

Let us start with the scope of criminal activities. Different societies make different choices about the appropriate scope of criminal law. Acts which are legal in one society may be illegal in another. Traditional crimes such as murder, rape and theft are the subject of the common body of criminal law in all societies. Although this common body is a large part of criminal activities, there are still activities, the legality/illegality of which is more controversial. Comparing these traditional crimes with acts that have been excluded from the scope of criminal law over time in some societies allows us to reflect on the conditions under which an act can be considered a crime.

In our model, the scope of criminal activities is expressed by D. As mentioned, D is net social damage from crime. Whether an activity is harmful or harmless is the main question influencing the attitudes of societies towards its legality/illegality. Selling and using drugs is an example. Other examples are adultery and extramarital sex. Further examples show that societies differ in how they determine the scope of criminal activities: drinking alcohol is a normal part of lifestyle in many countries and drink is freely available in supermarkets, whereas possession and drinking alcohol publicly is a crime in some Muslim countries and can be punished by fine or even flogging. In other words, such activities (using drugs, drinking alcohol, etc) are viewed as inflicting loss in some societies, whereas in others they are viewed as normal and not codified as a crime at all. We can specify this in a more formalized way as follows. If society A considers an activity harmless, then there is no potential victim from that and *H(.)* is zero in equation *D(.)*. Thus this certain activity does not impose a loss on society but has utility. There is no need for it to be considered a crime. On the other hand, in society B this certain activity may be regarded as undermining social values and may be punished harshly. This reasoning can be generalized to other controversial activities considered harmless in one society and harmful in another and accordingly classified as a crime.

The reason for considering some activities harmful for society and classifying them as crimes may be derived from religious or secular faith promoted or upheld by democratic or authoritarian states. Whatever the source and type of their constitutions, the point is that these societies have different perspectives on these activities and apply criminal law to them in an attempt to preserve what they consider their social identity. This is the first step towards divergence in criminal law which is deeply rooted in society's beliefs.

2.4.3 Diversity of punishment for certain crimes

If we even disregard differences in the scope of criminal activities and their reflection in the criminal law of different societies, there are still remarkable divergences in the associated punishments[33]. As an example, consider murder or rape. Both are crimes in all criminal justice systems worldwide, but there is a wide range of punishment for them. For instance, a rapist might be imprisoned for years in European countries but executed in China and certain Muslim countries. The range of punishments for murder goes from the death penalty or life imprisonment to only 20 years imprisonment. What is the reason for this remarkable divergence in punishment for the same crime in these societies? We now apply the social loss function to obtain insights into this issue in a more formalized way.

Consider an activity, murder or rape, which is a crime under all criminal laws worldwide. The aim of the public authority is to minimize the loss function from

[33] Indeed, there can also differences in procedural aspects that are not the concern of this study.

these activities according to equation (2.1) by applying its judicial system (apprehension, conviction and punishment). Given that our aim is to explain divergences in criminal law, we only focus on the optimal amount of punishment (f^*) that minimizes the social loss function. In order to so, we assume that variables p and b are fixed. Then, the first-order optimality condition optimal value of punishment is the same as equation (2.13) derived by differentiating L (social loss function in equation (2.1)) with respect to f:

$$D' + C' = -bpf(1 - \frac{1}{\varepsilon_f}) \qquad (2.36)$$

Where

$$\varepsilon_f = -\frac{f}{O} O_f \qquad (2.37)$$

Where (2.37) is crime supply elasticity with respect to severity of punishment. This is the familiar optimality conditions of equality of marginal revenues and marginal costs in the economic literature (for more details see section 2).

Now let us return to our main agenda on diversity of punishment in different societies in a more formalized way. Why might different societies have different optimal values for punishment (f^*) for a given crime, say rape? This phenomenon also seems to have roots in social beliefs and attitudes towards crime and punishment that we will try to specify through the present model. We again use comparative static analysis to see how the optimal level of punishment (f^*) changes in response to effective factors.

2.4.3.1 Degree of harmfulness of a crime

Just as the harmfulness of a certain activity, say abortion, can be open to dispute between societies, so can the degree of harmfulness of a certain crime. Indeed, different societies can have various attitudes towards the degree of harmfulness of different crimes, say rape. As a matter of fact, both retributive and utilitarian approaches to punishment confirm that punishment inflicted on the convicted offender must be in proportion to harm inflicted on the victim[34]. The subtle point is that there can be no agreement on the degree of harmfulness of a given crime, say rape, between different societies. This can lead to different levels of optimal punishment for the same crime. We can use comparative static analysis to see how a change in crime harmfulness can affect optimal punishment. Assume that D' is positively related to an exogenous variable α. Differentiating first order optimality condition in equation (2.36) with respect to α determines the effect of a change in crime harmfulness on the optimal punishment (f^*). Again it's easy to see that (refer to equations A1 to A8 in the Appendix):

$$\frac{\partial f^*}{\partial \alpha} > 0 \qquad (2.38)$$

So optimality requires an increase in punishment when the marginal cost of crime (crime's harmfulness) increases. Indeed, α can be interpreted as social attitudes to degree of harmfulness of a certain crime, say rape. In other words, it can be interpreted as a mirror of social sensitivity to a certain crime[35]. So if people in society A have different attitudes towards rape from people in society B and believe that rape is an extremely harmful crime and rapists should be punished severely, it is optimal for society A to inflict a more severe punishment on rapists than that inflicted by society B.

[34] This "proportionate principle" is called "internal wickedness" in Kant's terminology. It is worth mentioning that considering this principle is just a part of more comprehensive cost-benefit analysis of imposing punishment in a utilitarian approach that emphasizes "deterrence" as the main justification for punishing convicted offenders. For more details see section 3.

[35] This reasoning can even be applicable to particular cases of certain crimes in a single society too. A real example is the 21 year-old girl in India who was raped and killed by five men in a bus. This rape case led to public outrage in India and demonstrators asked for more severe punishment (higher f^*) for the convicted offenders.

As we discuss in the following paragraphs these attitudes are not the only effective factor in determining optimal punishment. For example, another crime with a range of punishments across societies is murder. However, it can be claimed that different public attitudes to the harmfulness of murder across societies can only partly explain the current wide range of punishments. Logically, murder imposes the same harm on all societies, if we assume people's lives should have the same value in all societies. Why then do some societies prefer to execute murderers and others prefer to impose life sentences or long imprisonment? Is the victim's life in the former societies considered more valuable than in the latter? The next sections endeavor to explain these questions on the basis of humanity, civilization of punishment, dignity of murderers as human and some utilitarian justifications for punishment.

2.4.3.2 Humanity or civilization of punishment

The death penalty has been abolished in many countries for any kind of crime. Other countries maintain capital punishment in their criminal justice systems. The model we adopted endeavors to explain this fact on the basis of different attitudes of societies towards the justification for a certain kind of punishment, say the death penalty. Consider the last term, $bpfo$, in the social loss function from criminal activities in equation (2.1). The coefficient b transforms the punishment imposed on convicted offenders into loss for the society. On one hand, as mentioned by Becker (1968), the magnitude of b depends technically on the kind of punishment. For instance, for fines, $b=0$ that means a fine does not impose any loss on society and is just a transfer of money from offenders to victims or the rest of society. In contrast, as discussed in section 1 for punishments such as imprisonment or the death penalty, $b>1$. The widespread use of fines as punishment today is justified from an efficiency perspective.

At the other hand, from a normative point of view, the magnitude of b may also capture social attitudes to a certain kind of punishment. Since different societies can have different attitudes to a certain kind of punishment, say the death penalty, the social loss from executing murderers can vary between societies. People in society A may be retributive towards murderers, considering the death penalty to be deserved, whereas people in society B may regret application of the death penalty, even in the case of murder, considering it "cruel and unusual". This exemplifies the greater magnitude of coefficient *b* for the death penalty in society B than in society A. Thus, not only does the magnitude of *b* depend technically on the kind of punishment but also normatively on social attitudes to a certain kind of punishment. We can consider coefficient *b* to be a function of social attitudes to a certain kind of punishment. If *b* is positively related to μ (regret for a certain kind of punishment, say the death penalty) then differentiation optimality condition with respect to μ yields (refer to equations A14 to A16 in Appendix):

$$\frac{\partial f^*}{\partial \mu} < 0 \tag{2.39}$$

The relation in the above equation looks like what is interpreted by commentators, such as John Pratt, as higher tolerance to offenders in what he calls civilized society. It has a clear message for policy makers; abolish punishments that society believes are "cruel and unusual". If we relax the assumption of fixed *p* in optimization process and let *f* and *p* determine simultaneously, then by comparative static analysis we will have $\frac{\partial p^*}{\partial \mu} > 0$. This indicates that more resources should be allocated to police and courts in order to increase the probability of apprehension and conviction (p^*). In this case, the prescription is that the deterrent effect of an increase in *p* partly compensates the deterrent effects lost by abolishing these cruel punishments, if they had any deterrent effect at all. In other words, the message is a shift from imposing a punishment that society believes to be "cruel and unusual" to more lenient and but more certain punishment. The interesting point inferred from the optimality requirements is that complying with social attitudes would lead to a dramatic change in resource allocation for fighting crime.

In a nutshell, if people in society A think it justified to impose the death penalty for murder, for example, then for them the magnitude of coefficient *b* would be lower than for people in society B who think that imposing the death penalty is never justified. Consequently, optimality implies that society B abolishes the death penalty while society A maintains it. Indeed, the magnitude of coefficient *b* for what a society calls "cruel and unusual" punishment is so large that it is optimal to abolish it and replace it with punishments having a lower magnitude of *b*, say imprisonment. Indeed, the remarkable proliferation of prisons seen in the last hundred years is inevitable from this perspective and confirms how changing social attitudes towards punishment affect the allocation of scarce resources and create or strengthen an institution, in this case prisons.

2.4.3.3 Deterrent effects of punishment

It has usually been held by policy makers and people that punishing offenders helps maintain law and order in society by deterrence. However, in academic discussions this claim is not as certain as it is in non-academic circles. An example is the controversial deterrent effect of capital punishment among scholars in recent decades. While there are dozens of empirical studies that confirm significant deterrent effects of capital punishment, there are also other empirical studies that reject any deterrent effect. This topic remains one of the most controversial among scholars in the field[36].

With regard to the emphasis on the deterrent effect of punishment, which is mainly a utilitarian justification for punishment, some commentators believe that ordinary citizens are insufficiently interested or competent to make sensible judgments about the efficacy of a certain kind of punishment, such as the death penalty, in crime

[36] For a more detailed review of empirical studies about deterrent effects of capital punishment, see Zimmerman (2006) "The economics of capital punishment and deterrence".

control. For instance, David Garland suggests that ordinary citizens may be much more concerned about whether punishment "rights wrongs" than about its function as an effective device for controlling crime (Rowan, 2011). It seems that discussion about maintenance or abolition of punishment such as the death penalty due its deterrent effects has been dominated by the academic elite, who can also affect social attitudes.

As we discussed in previous sections, in general there are two main rival approaches justifying punishment. As mentioned above, in the utilitarian approach, deterrence has a critical role in justification of punishment that can help to abolish or preserve a punishment in relation to its deterrent effects. Indeed, proponents of this approach may either be for or against a certain punishment, such as the death penalty for murder, depending critically on the believed deterrent effects of capital punishment. As a common point for all utilitarians, imposing punishment is simply a means to an end (in this case deterrence and a decrease in the murder rate) rather than a valuable end per se. So if there are empirical studies confirming this controversial deterrent effect, imposing capital punishment on murderers is easily justified by utilitarian proponents because it will save lives of innocent potential victims in the future. On the contrary, if empirical studies show that there is no deterrent effect of the death penalty, those against it can claim that the criminal justice system inflicts a huge cost on society by imposing the death penalty on convicted murderers without getting any benefit in terms of deterrence. In the absence of conclusive empirical studies on deterrence effects of punishment (say, capital punishment), people's attitudes to the deterrent effects of a certain kind of punishment, particularly one with huge costs such as the death penalty, are greatly affected by elite class opinions. The elite's polarization on this issue can polarize social attitudes as well.

The importance of deterrent effects of punishment in maintenance or abolition of a certain punishments, say the death penalty, can be formalized by the crime's supply elasticity of punishment (ε_f). Suppose $E_f = \frac{1}{\varepsilon_f} = \frac{-\partial f}{\partial O}\frac{O}{f}$ is positively related to exogenous variable γ. Then the effect of an increase in γ, i.e. a decrease in crime supply elasticity with respect to punishment, can be easily found by differentiating optimality equation (2.36) with respect to γ. It is easy to see that (refer to equation A12 to A13 in the Appendix):

$$\frac{\partial f^*}{\partial \gamma} < 0 \tag{2.40}$$

Equation (2.40) indicates that when the supply of crime is inelastic with respect to the severity of punishment, i.e. severity of punishment has little deterrent effect on the behavior of potential offenders, optimality suggests reducing the severity of punishment. Furthermore, if we relax the assumption of fixed *p* in the optimization process, then $\frac{\partial p^*}{\partial \gamma} > 0$. Interestingly, this shows another possible reallocation of scarce resources in the criminal justice context. If a society believes that a certain punishment has no deterrent effect, it will prefer to allocate more resources to certainty of a more lenient punishment and abolish the severe one. Empirical evidence from the last hundred years confirms that many societies have abolished the death penalty and allocated more resources to policing and other institutions of the criminal justice system, significantly increasing the certainty of a more lenient punishment with respect to the first decades of the 1900s. Interestingly, quantifying the deterrent effects of a certain punishment is an empirical question that has been studied by academic elites. Depending to these empirical findings, if academic elites of society A believe that the death penalty, for example, has no deterrent effect, they will recommend policy makers to abolish it and invest more resources in the certainty of a more lenient punishment, such as long imprisonment. For example, empirical findings on deterrence effects of capital punishment have always had influential effects on debates over maintenance or abolishment of capital punishment (see Nagin & Pepper, 2012).

Table 2.1 depicts the sources of divergence that we have discussed and how they affect criminal law across societies. The table not only lists the sources of divergence in criminal policy but can also be used to explain the process, and even the historical evolution, of criminal legislation in a given society, as discussed in next section. From a cross-society point of view, different societies give different weights to the mentioned factors in their cost-benefit analysis of the law-making process, leading to divergence between criminal justice systems. From a single society point of view, a society gives different weights to the factors in Table 2.1 in their cost-benefit analysis of the criminal law-making process. A society decides to abolish public execution even if they know it has a remarkable deterrent effect. In this case, humanity of punishment indicated by coefficient b in our model outweighs the deterrent effect identified by supply elasticity of punishment.

2.4.4 Historical evolution of punishment

In point of fact, throughout history many methods of punishment have been abolished in almost all societies. The abolition list includes the most brutal kinds of punishment, ranging from crucifixion to death by stoning. Even in countries with capital punishment, execution methods have evolved with elimination of hanging and its replacement by lethal injection or gas chamber in a private room, instead of publicly. Have these abolitions and evolutions happened accidentally or does it have something to do with efficiency? It will be shown in a formalized way that the evolution of punishments or what is called “evolving standards of decency” or “more civilized punishment” has led to higher efficiency in terms of lower social loss.

Sources of Divergence	Mechanism of divergence
Dispute over harmfulness or harmlessness of activity; Determination of scope of criminal activities.	$D(.) = H(.) - G(.)$ If $D(.) > 0$, then the activity is harmful and considered a crime
Dispute over degree of harmfulness of a crime; α is social attitude regarding harmfulness of a crime or even fear of crime.	If $D' = D'(\alpha)$ and $D''(\alpha) > 0$ then: $\frac{\partial f^*}{\partial \alpha} > 0$
Dispute over humanity of punishment; regret versus retributive emotions.	If $b = b(\mu)$ and $b'(.) > 0$, then $\frac{\partial f^*}{\partial \mu} < 0$, $\frac{\partial p^*}{\partial \mu} > 0$
Dispute over deterrent effects of a certain punishment, say death penalty.	If $\varepsilon_f = \varepsilon(\gamma)$ and $\varepsilon'(\gamma) < 0$ then $\frac{\partial f^*}{\partial \gamma} < 0$, $\frac{\partial p^*}{\partial \gamma} > 0$

Table 2.1- Sources of divergence in criminal law across societies

Various factors have lead to abolishment of "cruel and unusual" punishment over time. One is what is called "humanity of punishment" that has already been clarified in previous section. In addition to evolving standards of decency and higher tolerance to offenders, some other technical factors have also lead to abolishment of "cruel and unusual" punishments and encouraged societies to invest more resources in the probability of apprehension and conviction (p^*). Let us start with technological advances in policing and forensic aspects which have occurred in time. Technically, these improvements have increased the probability of apprehension and conviction. Assume that probability of apprehension and conviction (p) is positively related to an exogenous variable r. The effect of a positive shock to p on the optimal punishment can be found by differentiating of equation (2.36) with respect to r. it is easy to see that (refer to equation A17 to A19 in Appendix):

$$\frac{\partial f^*}{\partial r} < 0 \qquad (2.41)$$

This simply identifies how probability of apprehension and conviction and optimal amount of punishment can interact with each other in a utilitarian perspective of policy making for combating crime. Optimality implies the abolishment of severe punishments when certainty of a more lenient punishment increases following an exogenous technological shock (*r*).

Furthermore, if there is accumulated evidence and knowledge that higher levels of probability and conviction has greater deterrent effects than severity of punishment, optimality implies abolishing "cruel and unusual" punishment and investing more resources in the police, courts and prisons. In the model, if the assumption of fixed *p* is relaxed and *p* and *f* are let to be determined simultaneously and if ε_p (crime supply elasticity with respect to probability of apprehension and conviction) is assumed to be positively related to exogenous variable φ then we will have:

$$\frac{\partial f^*}{\partial \varphi} < 0 \quad , \quad \frac{\partial p^*}{\partial \varphi} > 0 \qquad (2.42)$$

Historical evolution of punishment as well as police and courts confirm these findings[37]. On one hand, we see abolishment of "cruel and unusual" punishments and on the other the institutions of police and courts have become more organized and more professional over time. This can be regarded as a paradigm shift in policy-making for fighting crime, a shift from severity of punishment to certainty of punishment. Interestingly, this shows how a mix of changes in attitudes towards humanity of punishment, technological advances and criminology has lead to resource allocation and consequently a remarkable change in criminal justice systems.

[37] For a comprehensive review of evolution of punishment in English-speaking countries, see John Pratt (2002).

2.4.5 Conclusion: Comparative Criminal law

Criminal law provides a code of conduct that everyone in a society is expected to follow. Different societies have different conduct codes. Although a remarkable part of the body of criminal law is common to them all, there are still different parochial features. The scope of criminal activities and associated punishments for a given crime is different across societies. Why is this so?

Divergence in criminal law between societies is rooted in social beliefs. To regard an activity as a crime, it must be harmful. Societies can have utterly different attitudes towards the harmfulness of a certain activity. While some societies view same-sex marriage as a symbol of freedom of expression of love and make it legal, other societies consider it evil and a threat to social cohesion and accordingly criminalize it.

Apart from divergent views on the scope of criminal activities, societies can also have different attitudes towards a variety of factors affecting optimal punishment for even same criminal activities. The first factor that can be a source of divergence in optimal punishment between societies is the degree of harmfulness ascribed to a certain crime. If people in society A think that a certain crime, say rape, is more harmful than do people in society B, then optimality implies that the former society will impose more severe punishment on convicted rapists. Furthermore, societies concerned with humanity of punishment and the human rights of offenders promote more lenient punishment. A society that considers a punishment "cruel and unusual" believes that cruel punishments inflict high social loss, making it more efficient to abolish it. Indeed, societies can also have different views on punishment, based on retributive or regretful emotions that lead to divergence in optimal punishment. Finally, different views on the deterrent effects of a certain punishment, say the death penalty, can lead to its abolition or maintenance. Different views on these various factors lead to different criminal law systems. The evolution of criminal law in a society depends on the relative importance of these factors for criminal law decision-making.

2.5 Concluding summary

This chapter applies an economic model to try to give a more formalized answer to two interesting questions, respectively; what are the sources of irreconcilabilities between utilitarian perspective and retributive perspective to punishment and why different societies might have different criminal law.

Although both utilitarian and retributive perspectives support the existence of institution of punishment, emphasis on different distributive principles of punishment has led to different optimal level of punishment. The well-known “proportionate principle” which is the only punishment-deriving factor in the retributive perspective is also the only part that Retributivists and Utilitarians share. However, it seems that other principles of deterrence, incapacitation and rehabilitation which are justifications for punishment have more critical role in punishment determination in utilitarian perspective. One of the major sources of divergence in optimal punishment in these two rival schools of punishment is the interaction between severity and certainty of punishment in order to reach to a certain level of expected punishment which is completely missing in retributive approach.

In actuality, criminal law making is a mix of above-mentioned principles by both Utilitarians and Retributivists. Generally speaking, criminal law addresses criminal activities and their associated punishments. Comparative criminal law can be illustrated by reference to some key factors such as different attitudes towards harmfulness of a certain activity, say abortion, the degree of harmfulness of a certain crime, say rape, the concerns about humanity and standards of decency of a certain kind of punishment, say death penalty and finally the academic disputes over deterrence power of a certain punishment. Differences in criminal law results from the fact that societies may have different attitudes towards these factors and give different weights to them in the process of criminal law making.

Mathematical Appendix

Suppose that D' is positively related to an exogenous variable α. Then the effect of a change in α (harmfulness of a crime) on the optimal level of punishment can be found by differentiating the optimality condition with respect to the policy variable *f*. If *p* and *b* are assumed fixed, the value of *f* that minimizes *L* can be found by the following necessary condition:

$$\frac{\partial L}{\partial f} = (D' + C')O_f + bpf(1 - \frac{1}{\varepsilon_f}) = 0 \qquad \text{(A1)}$$

Rearranging equation (A1):

$$\frac{\partial L}{\partial f} = (D' + C')O_f + bpf(1 - E_f) = 0 \qquad \text{(A2)}$$

Where

$$E_f = \frac{1}{\varepsilon_f} = \frac{-\partial f}{\partial O}\frac{O}{f} \qquad \text{(A3)}$$

The sufficient condition is:

$$\frac{\partial^2 L}{\partial f^2} = (D'' + C'')O^2{}_f + bp(1 - E_f)O_f = 0 \qquad \text{(A4)}$$

The second order condition can be rearranged as:

$$\phi = (D'' + C'') + bp(1 - E_f)\frac{1}{O_f} > 0 \qquad \text{(A5)}$$

Assume that D' is positively related to an exogenous variable α. Differentiating equation (A2) with respect to α determines the effect of a change in crime harmfulness on the optimal punishment (f_u^*) :

$$D'_\alpha + (D'' + C'')O_f \frac{\partial f_u^*}{\partial \alpha} + bp(1 - E_f)\frac{\partial f_u^*}{\partial \alpha} = 0 \qquad \text{(A6)}$$

Or

$$\frac{\partial f_u^*}{\partial \alpha} = \frac{-D'_\alpha (1/O_f)}{\phi} \qquad \text{(A7)}$$

Since $\phi > 0$, $O_f < 0$ and by assumption $D'_\alpha > 0$ then:

$$\frac{\partial f_u^*}{\partial \alpha} = \frac{-D'_\alpha (1/O_f)}{\phi} = \frac{+}{+} > 0 \qquad \text{(A8)}$$

Assume that probability of apprehension and conviction (*p*) is positively related to an exogenous variable *r*. For instance, technological progress, such as fingerprinting, which can increase the probability of conviction and apprehension of offenders, can be considered a positive shock to *p*. The effect of this positive shock to *p* on the optimal punishment can be found by differentiating of optimal condition in equation (A2) with respect to *r:*

$$(D'' + C'')O_f \frac{\partial f_u^*}{\partial r} + (D'' + C'')O_p p_r + C_{po} p_r + bp(1 - E_f)\frac{\partial f_u^*}{\partial r} + bf(1 - E_f)p_r = 0$$

(A9)

Since by assumption $C_{po} = 0$, by arranging equation (A9) we have:

$$\frac{\partial f_u^*}{\partial r} = \frac{-(D'' + C'')O_p (1/O_f)p_r - bf(1 - E_f)p_r (1/O_f)}{\phi}$$

(A10)

Since $(D'' + C'') > 0$, $O_f < 0$, $O_p < 0$ and $p_r > 0$, then we have:

$$\frac{\partial f_u^*}{\partial r} = \frac{(-)+(-)}{+} = \frac{-}{+} < 0 \tag{A11}$$

Suppose E_f is positively related to γ. Then the effect of an increase in γ, i.e. a decrease in crime supply elasticity with respect to punishment can be easily found by differentiating equation (2.13) with respect to γ:

$$(D'' + C'')O_f \frac{\partial f_u^*}{\partial \gamma} + bp(1 - E_f)\frac{\partial f_u^*}{\partial \gamma} - bpfE_{f\gamma} = 0 \tag{A12}$$

Rearranging (A12) we get:

$$\frac{\partial f_u^*}{\partial \gamma} = \frac{bpfE_{f\gamma}(1/O_f)}{\phi} = \frac{-}{+} < 0 \tag{A13}$$

If b is positively related to μ (regret for a certain kind of punishment, say the death penalty) then differentiation optimality condition with respect to μ we will have:

$$(D'' + C'')O_f \frac{\partial f^*}{\partial \mu} + bf(1 - E_f)\frac{\partial f^*}{\partial \mu} + pf(1 - E_f)b_\mu = 0 \tag{A14}$$

Or

$$\frac{\partial f^*}{\partial \mu} = \frac{-b_\mu pf(1 - E_f)(1/O_f)}{\phi} \tag{A15}$$

Since $1 - E_f < 0$, and by assumption $b_\mu > 0$ then:

$$\frac{\partial f^*}{\partial \mu} = \frac{-}{+} < 0 \tag{A16}$$

Chapter 3

Estimating a Dynamic Economic Model of Crime Using Panel Data from North Carolina

Abstract: This chapter estimates a dynamic economic model of crime using panel data on 90 counties in North Carolina over the period 1981-87. Cornwell and Trumbull (1994) estimated a Panel model using this data set and Baltagi (2006) replicated their results and suggested some corrections. Both CT (1994) and Baltagi (2006) also address the conventional simultaneity problem for two explanatory variables (Probability of arrest and Police per capita) and apply instrumental variables (IV) to deal with it. We start to correct for one of deterrent variables; probability of conviction (P_c) proxied by the ratio of number of convictions to the number of arrest. Regarding that any arrested suspect either convicts or acquits it's expected that $0 \le P_c \le 1$. However, there are 71 observations (almost 11%) for which $P_c > 1$. It might likely be due to measurement mistakes in original dataset. We correct the dataset by simply removing observation which $P_c > 1$. Furthermore, we checked for legitimacy of endogeneity problem and continue to see how well their IV-model has been identified to deal with it. The IV-model applied by two previous studies hasn't identified well enough to produce reliable estimates. Finally, considering appropriate panel dimensions of data set which fits "small T, large N" and existence of inertia in criminal activities, we estimates a dynamic panel data model. Our estimates for deterrence variables are highly significant and much lower than ones estimated by CT (1994) and Baltagi (2006). Upward bias in deterrence is partly due to mistakes in original dataset and partly due to weak

instruments applied for dealing endogenity problems in above-mentioned covariates.

JEL: K14, C23

Keywords: Deterrence, Dynamic Panel Data model, Economics of crime

3.1 Introduction

Over the last three decades, a growing amount of research effort, largely inspired by Becker (1968) and Ehrlich (1973), has been devoted to study the socio-economic determinants of criminal behavior, partly motivated by the remarkable increase in criminal activities in many developed countries. Before 1968, offenders were regarded as deviant individuals with atypical motivations. The theory of crime was largely composed of recommendations made by sociologists, psychologists, criminologists and law professors that were not based on rigorous empirical investigations but on beliefs about concepts like depravity, insanity, and abnormality. Modern studies in the field of "Economics of Crime" have been stimulated by the dramatic increase of crime rates in developed countries, on the one hand, and by recent social and economic problems on the other. The focus of these contributions has changed from the pure testing of the deterrence hypothesis, as the main concern of deterrence theory, to the analysis of socioeconomic and demographic crime factors.

Cornwell and Trumbull (1994), hereafter (CT), estimated an economic model of crime using panel data on 90 counties in North Carolina over the period 1981–1987. The empirical model follows Becker (1968) and Ehrlich (1973) among others, and relates the crime rate to a set of explanatory variables which include deterrent variables, demographic variables as well as variables measuring returns to legal opportunities. They also address simultaneity problem of two regressors (probability of arrest and police per capita) and try to deal with it by using instrumental variables. Their Hausman type test for choosing between "Fixed Effects (FE) estimator" (presumably more efficient but inconsistent under the null hypothesis) and FE2SLS (presumably consistent under the null hypothesis) rejects the later in favor of the former. Based on FE estimator results, they conclude that both some labor market incentives and law enforcement incentives matter in crime combatting. They also conclude that neglecting county heterogeneity biases upward deterrent effects estimates.

Baltagi (2006) replicated CT (1994) estimation results and corrected some estimation mistakes and specifications test. Baltagi (2006) argues that the usual Hausman test, based on the difference between Fixed Effects and Random Effects, may lead to misleading inference when some regressors are endogenous. Consequently, he suggests running a random effects 2SLS estimator which allows for endogeneity of some explanatory variables too. This estimator is denoted by Error Components 2SLS or EC2SLS in Baltagi (2006). A Hausman test based on the difference between FE2SLS and EC2SLS, suggested by Baltagi (2004), cannot reject the consistency of EC2SLS. According to Baltagi (2006), this estimator yields plausible and significant estimates of the crime model.

This chapter makes a fresh attempt at estimating Becker's type model using the North Caroline data set with a double purpose. First, starting to delve into dataset applied in these studies, there is likely a serious problem with measuring of one of the deterrent variables; probability of conviction (P_c) which is proxied by the ratio of number of convictions to arrests. For having a better image of law enforcement process and deterrent variables applied in above-mentioned studies see Figure 1.

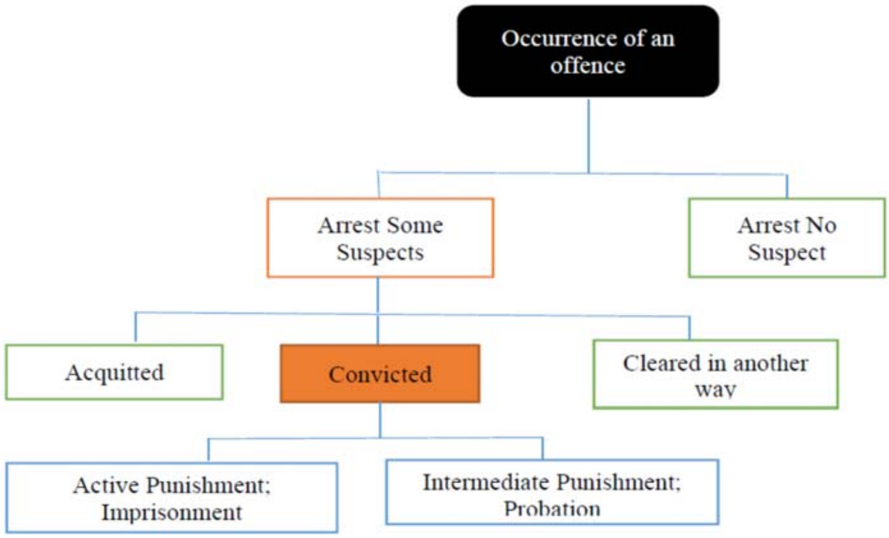

Figure 1- The Process of Law Enforcement
*There are many reasons as to why an incident may be cleared in another way. Examples include the death of a suspect or the dropping of a charge to a less serious offence which is not felony.

As a matter of fact, an arrested suspect is either convicted or acquitted (or cleared in another way). So it's expected that $0 \leq P_c \leq 1$. When this ratio is larger than one, it might be likely a mistake in measurement[38]. In this dataset there are 71 observations (almost 11%) which for them this ratio is larger than 1[39]. It might be said that an arrested suspect can convict for more than one crime which can oversize the probability of conviction or even increase this ratio to more than one. If so then it might be better to measure this ratio as number of convictions to the number of reported crimes. But then the

[38] The number of arrests for each county in the denominator of probability of conviction is sum of arrests reported by available police agencies in that county. In some years for some counties, number of police agencies actively involved in the crime reporting including arrests has declined dramatically. Regarding that number of convictions are reported by North Carolina Sentencing and Policy Advisory Commission, this has undersized number of arrest and consequently oversized the probability of convictions.

[39] There is even an observation which probability of conviction is 37.

probability of conviction will be undersized[40]. Furthermore, one might wonder that these convictions might be related to crimes occurred in previous years and just have been convicted with delay due to time-consuming process of trial. This usually only happens for serious felonies of murder and rape which are in average less than 1 percentage of all crimes reported. In sum, it seems that the above-mentioned problem concerning probability of conviction to be likely because of measurement mistakes. Our strategy for dealing this problem is to remove data for which $P_c > 1$. Then we can check how estimated results by CT (1994) and Baltagi (2006) will change. We'll do it in the next section after presentation of data.

Even after doing correction of the dataset, there is still another more interesting agenda. Both CT (1994) and Baltagi (2006) have dealt with conventional enodogeneity problem in two explanatory variables (Probability of arrest and Police per capita) and have used Instrumental Variables (IV) to deal with this problem for getting consistent and reliable estimates. It is well known that when some regressors are endogenous and the IV-models are applied, legitimacy of instruments has crucial role in consistency of estimation results. We delve more into their models with endogeneity test and after that some identification tests to see how well their models have been identified to produce reliable estimates. The endogeneity test confirmed the joint endogeneity of the suspect explanatory variables[41]. However, different identification tests all confirmed that the presumably endogenous regressors have been either under-identified or weakly identified by the instruments used by the two set of authors, which implies their estimation results (even after correction of the dataset for P_c) may not be reliable enough.

[40] However, there are still 12 observations (almost 2%) which the number of convictions are even more than number of reported crimes. See footnote 38 for an explanation for this.
[41] Interestingly, doing the same test using the original data (the uncorrected version) cannot reject the exogeneity of these regressors.

We therefore proceed with alternative estimators that allow for endogeneity but can tackle the problem of weak instruments such as those available in the North Caroline dataset. Considering the dynamic feature of crime as well as feasibility of Dynamic Panel Data (DPD) models for our dataset which fits "Small T, large N" (T=7, N=90), we apply a DPD model to deal not only with conventional endogeneity problem but also with measurement error problem which has been fairly documented in empirical literature of economics of crime. We will estimate both "Difference-GMM" and "System-GMM" to have a better image for economic model of crime.

In presence of weak instruments which may bias the results, difference-GMM first proposed by Holtz-Eakin et al. (1988) flank the proposed instruments with lagged levels of the endogenous regressors thus removing the correlation with the error term. However, the lagged levels of the regressors maybe are poor instruments for the first-differenced regressors. The augmented version of the difference GMM – the system GMM- may then be considered where a two-equation systems is estimated (one equation in levels, the other in difference) with in levels endogenous variables are instrumented by the corresponding lagged in-difference variables[42].

Our estimated results using Dynamic Panel Data (DPD) model confirms that crime has inertia implying counties with higher crime rates in the past are expected to experience higher crime levels in the future. Furthermore, our estimated elasticities for deterrent variables are much lower than the estimated coefficients for EC2SLS estimator suggested by Balatgi even after correction of the dataset. So it seems that both CT (1994) and Baltagi (2006)'s estimation results for deterrent variables have upward bias. This bias partly stems from their models' weak identification due to using weak instruments and partly due

[42] For an overview of system and difference GMM and their implementation in STATA see Roodman (2009b).

to using probability of conviction in a range which does not make sense i.e. P_c>1.

The rest of the chapter organizes as follows. We start with presentation of dataset and discussing the factors which might affect the crime level in Section 2. The endogeneity and some identification tests are presented in Section 3. Section 4 delves into measurement error in economic model of crime. A dynamic panel data model will be introduced in section 5. We present estimation results in section 6. And concluding remarks will be presented in section 7.

3.2 The Data and socioeconomic determinants of crime

The data used are a panel on 90 counties in North Carolina with observations running from 1981-1987. Data on crime are taken from the FBI crime index reports and are available annually on county level basis. Summary statistics, expressed in values per capita where applicable, are presented in Table 3.1.

The explanatory variables are separated into three groups: deterrence variables, demographic variables and economic variables. Deterrence variables which are at the center of "deterrence hypothesis" in Becker's model determine the expected returns from crimes. This study includes Probability of arrest P_a (which is measured by the ratio of arrests to number of reported offences), probability of conviction given arrest P_c (which is measured by the ratio of convictions to number of arrests), probability of imprisonment P_p (which is measured by the proportion of total convictions resulting in prison sentences) and number of police officers as deterrence variables.

The other set of control variables concerns for demographic features of counties. Young men are said to be more prone to engage in criminal activities than the rest of the population (Freeman 1996; Grogger 1998). Thus regarding crime as a business of young male people, percentage of young male between

15-24 years old has been included as control variable. Another demographic feature considered here is the proportion of non-white or minorities. The other two more demographic variables are population density and Urbanization.

This dataset includes a set of economic variables which are average weekly wages for different industries. Similar to analysis made by Ehrlich (1973) for median income, average weekly wages can be considered as proxy for economic prosperity and thus as an indicator for illegal income opportunities. However, both CT (1994) and Baltagi (2006) use these weekly wages to measure legal income opportunities. They argue that higher wages are indicators for more rewarding legal jobs. An answer to this question can only be obtained from empirical evidence, i.e., from the sign of the estimated coefficient of the wages variable in a multivariate analysis. A positive (negative) coefficient would support the interpretation of higher wages as indicators of illegal (legal) income opportunities. However, we can expects an increase in wages in industries that are more simple labor-intensive such as Wholesale and retail sale or Construction (with lowest average weekly wage) likely decrease the crimes. In contrast, an increase in wages of white-collar industries such as Federal Government (with relatively higher average wage) can potentially increase crime by increasing more prosperous targets as mentioned by Ehrlich (1973).

As mentioned in previous section, there is likely a measurement mistake in probability of conviction which we deal with it in this section to see how CT (1994) and Baltagi (2006) estimates will change after our correction of the dataset. As Table 3.2 displays, our estimated results after correction to P_c and CT (1994)'s original FE-estimator are almost the same. The only considerable difference is estimated coefficient for Police which is still positive but almost half. Likewise in CT (1994), here again a Hausman-type test between FE and EF2SLS cannot reject the consistency of the FE estimator. In contrast, the estimator suggested by Baltagi (2006) is more vulnerable to performed correction to dataset. We cannot reject the consistency of EC2SLS estimator even after doing correction of the dataset. However, as the last two columns in Table 3.2 shows, probability of arrest is no longer significant at all. In addition, the probability of conviction (P_c) and probability of prison (P_p) are still significant but their amounts are almost half of the Baltagi's original estimated coefficients. In sum, Baltagi (2006)'s estimates for deterrent variables have upward bias which can be due to mistakes in the dataset. Finally, the estimated coefficient for Police is still unexpectedly positive but almost 20% lower than before.

3.3 Endogeneity test, First-Stage regressions and identification of endogenous regressors

Before presenting the Dynamic Panel Data (DPD) model which is the core of this chapter, let's see how well the fixed effects –IV model has been identified. Both CT (1994) and Baltagi (2006) have estimated FE2SLS estimator and both of them have rejected it in favor of "FE estimator" and "EC2SLS" estimator, respectively. As it has been mentioned by Baltagi (2006), these results depend upon the legitimacy of instruments chosen by CT (1994). This section tries to check the legitimacy of these instruments by applying different identification tests. As mentioned in previous section, even after correction of the dataset, we still cannot reject the consistency of FE and EC2SLS estimators. All tests and estimations in next sections are running on the corrected version of the dataset.

3.3.1 Test of Endogeneity

Two explanatory variables of probability of arrest and police per capita have been considered as endogenous in both CT (1994) and Baltagi (2006). They claim that, at one hand, counties experiencing rising crime rates, holding police resources constant, would see probabilities of arrest fall. On the other hand, increases in crime may motivate a county to increase policing resources which, in turn, would increase the probability of arrest. So depending on crime level, both probability of arrest and police staff could be endogenous.

In order to see how well endogeneity test confirm this claim, we can use Davidson and Mackinnon (1993)'s test to check for exogeneity of suspected endogenous explanatory variables. Under the null hypothesis of exogeneity, OLS estimator rather than IV estimator, would yield consistent estimates. In other words, any suspicion of endogeneity would not have deleterious effects on the OLS estimates. This test, which is similar to the Hausman test in this context, will always yield a computable test statistic, while the Hausman test, depending on the difference of estimated covariance matrices being a positive definite matrix, often cannot be computed by standard matrix inverse methods. The test statistic[43], under the null hypothesis, is distributed as *F (m, N-k)*, where *m* is the number of regressors specified as endogenous in the original instrumental variables regression. A rejection indicates that the *FE-IV* estimator should be employed. As Davidson and MacKinnon (1993, p. 241-242) discuss, the test may be applied to a subset of the endogenous variables too (See Davidson and MacKinnon (1993, p. 237-240) and Hayashi (2000, p. 233-34)).

In our case the null hypothesis assumes exogeneity of the probability of arrest and police per capita and the test statistic is $F(2,447) = 3.26$ with *p-value=0.039* which indicates we can reject the null hypothesis at 5% level of significance ; nor probability of arrest neither police per capita are exogenous[44]. So we can reject the null hypothesis of exogeneity of these regressors. It seems endogeneity problem to be more relevant for police. (See footnote 43).

---

[43] This test involves an ordinary least squares regression (using command *xtreg, fe* in Stata) of the original dependent variable on the original regressors, augmented by the residuals from each of the first stage instrumental variables regressions. This regression gives rise to an F test for the joint hypothesis that each of the coefficients on the residual series are zero.

[44] However, when we just test endogeneity of either probability of arrest or police per capita the associated test statistics are $\chi_1^2 = 0.001, p-value = 0.97$ and $\chi_1^2 = 1.35, p-value = 0.25$, respectively.

Table 3.1-Summary Statistics

	Mean	Standard Deviation
Crime rate	0.0316	0.0181
Deterrence Variables		
Pa(Probability of Arrest)	0.307	0.171
Pc (Probability of Conviction)	0.689	1.690
Pp (Probability of Prison)	0.426	0.087
S (Average Sentence (in days))	8.955	2.658
Police Per Capita	0.00192	0.00273
Demographic variables		
Density	1.386	1.440
Pym: Percentage ages 15-24	0.089	0.024
Pmin: Percentage of minorities	0.257	0.169
Average Weekly Wage		
wcon: Construction	245.67	121.98
Wtuc: Transportation, Utilities, Communication	406.10	266.51
Wtrd: Whole sale & retail trade	192.82	88.41
Wfir: Finance, Insurance, Real estate	272.06	55.77
Wser: Services	224.67	104.87
Wmfg: Manufacturing	285.17	82.37
Wfed: Federal government	403.90	63.07
Wsta: State government	296.91	53.43
Wloc: Local government	257.98	41.36
Geographical variables		
Urban	0.089	0.285
Central	0.378	0.485
West	0.233	0.423

Notes: Crime rate is the ratio of FBI index of number of crimes divided by the county population. The probability of arrest is measured by the ratio of arrests to offences. Probability of conviction given arrest is the ratio of convictions to arrests and probability of a prison sentence given a conviction is measured by the proportion of total convictions resulting in prison sentences. Average prison sentence in days is a proxy for sanction severity. The population density, which is the county population divided by county land area. Dummy variable of Urban indicates whether the county is in the SMSA with population larger than 50,000. Percentage of minorities is the proportion of the county's population that is minority or non-white.

Table 3.2- Estimated results after correction to Probability of conviction

	(1) FE CT	(2) FE Corrected	(3) FE2SLS CT	(4) FE2SLS Corrected	(5) EC2SLS- Baltagi	(6) EC2SLS Corrected
P_A	-0.355***	-0.340***	-0.576	-0.320	-0.413***	-0.0857
	(0.0322)	(0.0339)	(0.802)	(0.580)	(0.0974)	(0.113)
P_C	-0.282***	-0.229***	-0.423	-0.260	-0.323***	-0.148**
	(0.0211)	(0.0221)	(0.502)	(0.247)	(0.0536)	(0.0452)
P_P	-0.173***	-0.179***	-0.250	-0.185	-0.186***	-0.121**
	(0.0323)	(0.0310)	(0.279)	(0.158)	(0.0419)	(0.0413)
S	-0.00245	0.0143	0.00910	0.0532	-0.0102	0.00760
	(0.0261)	(0.0238)	(0.0490)	(0.0574)	(0.0270)	(0.0261)
Police	0.413***	0.222***	0.658	-1.017	0.435***	0.349***
	(0.0266)	(0.0544)	(0.847)	(1.576)	(0.0897)	(0.0994)
Density	0.414	0.514*	0.139	0.854	0.429***	0.463***
	(0.283)	(0.249)	(1.021)	(0.467)	(0.0548)	(0.0534)
Wcon	-0.0378	-0.0540	-0.0287	-0.0672	-0.00748	-0.0566
	(0.0391)	(0.0333)	(0.0535)	(0.0568)	(0.0396)	(0.0365)
Wtuc	0.0455*	0.0522**	0.0391	0.0671*	0.0454*	0.0553**
	(0.0190)	(0.0168)	(0.0309)	(0.0302)	(0.0198)	(0.0184)
Wtrd	-0.0205	-0.0274	-0.0178	-0.0102	-0.00814	-0.0248
	(0.0405)	(0.0331)	(0.0453)	(0.0564)	(0.0414)	(0.0363)
Wfir	-0.00390	-0.0254	-0.00934	-0.0359	-0.00364	-0.00708
	(0.0283)	(0.0467)	(0.0366)	(0.0711)	(0.0289)	(0.0501)
Wser	0.00888	-0.00689	0.0186	-0.0170	0.00561	-0.00496
	(0.0191)	(0.0165)	(0.0388)	(0.0253)	(0.0201)	(0.0181)
Wmfg	-0.360**	-0.303**	-0.243	-0.369*	-0.204*	-0.0696
	(0.112)	(0.113)	(0.420)	(0.185)	(0.0804)	(0.0859)
Wfed	-0.309	-0.275	-0.451	0.0986	-0.164	-0.102
	(0.176)	(0.152)	(0.527)	(0.544)	(0.159)	(0.143)
Wsta	0.0529	0.0186	-0.0187	-0.0000116	-0.0540	-0.0384
	(0.114)	(0.0978)	(0.281)	(0.166)	(0.106)	(0.0976)
Wloc	0.182	0.0826	0.263	-0.258	0.163	0.122
	(0.118)	(0.112)	(0.312)	(0.453)	(0.120)	(0.122)
PYM	0.627	0.907**	0.351	1.897	-0.108	0.0618
	(0.364)	(0.325)	(1.011)	(1.522)	(0.140)	(0.124)
PMIN					0.189***	0.122**
					(0.0415)	(0.0426)
West					-0.227*	-0.292**
					(0.0996)	(0.100)
Central					-0.194**	-0.205***
					(0.0598)	(0.0600)
Urban					-0.225	-0.125
					(0.116)	(0.107)
_cons	2.393	2.336	2.943	-3.463	-0.954	-0.913
	(1.678)	(1.546)	(2.694)	(9.013)	(1.284)	(1.402)
N	630	559	630	559	630	559

Robust Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Lack of or weak endogeneity of suspect endogenous variables might be due to time lag effects. With regard to police per capita, the process of training and recruiting police staff is time consuming, hence it is plausible that crime level in year *t* affects police recruitment at least with a one year delay[45].As for the probability of arrest it seems to be more complicated. At one hand, considering probability of arrest definition (the ratio of arrest to number of offences), an increase in crimes number will increase the denominator and so decreases the probability of arrest as far as arrests remain fixed or increase proportionally less than crimes. At the other hand, an increase in crime supposedly will increase either police staff or the current police time-working which will increase number of arrests. In the end, it's not clear what will happen to the probability of arrest. It can decrease, increase or even remain unchanged.

3.3.2 Under-identification and weak identification tests

Regarding both of these regressors (Probability of arrest and police per capita) as endogenous, we find that none of them has been identified well enough in previous studies by CT (1994) and Baltagi (2006). We can test both under-identification and weak identification, reported for each endogenous regressor separately, using the method described by Angrist and Pischke (2008). The Angrist-Pischke (AP) first-stage χ^2 and F statistics are tests of under-identification and weak identification, respectively, of individual endogenous regressors. They are constructed by "partialling-out" linear projections of the remaining endogenous regressors. The AP's Wald statistic is distributed as $\chi^2(L_1 - K_1 + 1)$ under the null that the particular endogenous regressor in question is unidentified where L_1 is the number of excluded instruments (or simply instrumental variables) and K_1 is the number of endogenous regressors.

[45] It seems that police per capita to be a predetermined variable which is affected by past realization of the error term and not with the current and future realization.

The AP test will fail to reject if a particular endogenous regressor is unidentified.

The AP first-stage *F* statistic is the *F* form of the same test statistic. It can be used as a diagnostic for whether a particular endogenous regressor is "weakly identified". Critical values for the AP first-stage *F* as a test of weak identification are not available, but the test statistic can be compared to the Stock-Yogo (2005) critical values for the Cragg-Donald *F* statistic with $K_1 = 1$. The AP test statistics are presented in Table 3.3.

Table 3.3- Angrist and Pischke (2008) under-identification and weak identification test

	Under-identification		Weak identification
Endogenous Variable	AP $\chi^2(1)$	p-value	AP F (1,518)
Probability of Arrest	3.59	0.058	3.42
Police per Capita	1.22	0.27	1.17
Stock-Yogo (2005) weak Id test for single endogenous regressor			
	10% maximal IV size		16.38
	15% maximal IV size		8.96
	20% maximal IV size		6.66
	25% maximal IV size		5.53

According to the AP's test, the presumably endogenous regressors are both under-identified and weakly identified and consequently applied instruments are not legitimate enough to make estimation results for FE2SLS estimator reliable.

In addition to AP test statistics which checks under-identification and weak identification for each of the endogenous regressors separately, there are some other tests which check the identification of endogenous regressors together rather than separately. "Weak identification" arises when the excluded instruments are correlated with the endogenous regressors, but only weakly.

Estimators can perform poorly when instruments are weak. When errors are assumed to be independently and identically distributed (i.i.d), we can check the weak identification by an *F* version of the Cragg-Donald (1993) Wald statistic. Under the null hypothesis, the equation has been identified weakly. Stock and Yogo (2005) have compiled critical values for the Cragg-Donald *F* statistic for several different estimators, several different configurations and several definitions of "perform poorly". When the Cragg-Donald *F* statistic is lower than critical values, it indicates that we have "weak identification" problem. The first row of Table 3.4 depicts the Cragg-Donald Wald *F* statistic along Stock and Yogo (2005) critical values when $L_1 = 2$ and $K_1 = 2$. If we drop out the i.i.d assumption for errors, the Cragg-Donald *F* statistic is no longer valid. In this case, correspondingly-robust Klienbergen-Paap Wald *F* statistic is valid. The critical values for this *F* statistic are the same Stock and Yogo's critical values for the Cragg-Donald i.i.d case. The second row of Table 3.4 presents this non-i.i.d robust statistic. In both cases, we cannot reject the null hypothesis implying that we have "weak identification" problem. Again, we conclude that FE2SLS's estimates may not be reliable enough[46].

Table 3.4- Weak identification test

Null Hypothesis: the equation is weakly identified		
Cragg-Donald Wald F statistic		0.45
Kleibergen-Paap Wald rk F statistic		0.47
Stock-Yogo (2005) weak Id test critical values for $L_1 = 2$, $K_1 = 2$:		
	10% maximal IV size	7.03
	15% maximal IV size	4.58
	20% maximal IV size	3.95
	25% maximal IV size	3.63

[46] There are some other under-identification tests which all get the same results as do the tests mentioned in the text. The results of these tests are available upon requests.

So, based on above-discussed test, we can conclude that i) we cannot reject the endogeneity of suspected endogenous regressors when they are considered jointly and ii) the instruments applied are not enough legitimate to make reliable estimates[47].

3.4 Errors-in-variables and the apparent effect of arrest rates on crime

This section discusses the relevance of measurement errors in estimation of crime model prior to introducing the estimation methodology that addresses this (and other issues; endogeneity). Measurement error in the applied literature on the economics of crime are more likely to arise from the underreporting crimes, with the consequent under-estimation of key variables such as the crime rate. Consider the case where the dependent variable, C, the crime rate, is measured with error where C is essentially attempting to measure the true crime rate C^*. Let the error be $C - C^* = e$. We are interested in estimation of what follows:

$$C_t^* = \beta_0 + X\beta + \varepsilon_t \qquad (3.1)$$

But we can only observe *C* which restricts us to estimate a model with a composite error term:

$$C_t = \beta_0 + X\beta + \varepsilon_t + e_t \qquad (3.2)$$

If we assume that *e* is not systematic in that it is not correlated with the independent variables, then the measurement error only weakens the fitness of the model but does not introduce bias in either point or interval estimates. If *e* is correlated with one or more regressors then OLS estimates will be biased and inconsistent. In our context, *e* might be correlated with the number of police staffs as counties with more policing services might report more crimes. If so, the explanatory variable of police per capita violates the zero-conditional-mean

[47] Considering that number of excluded instruments equals to the number of endogenous regressors, it's not possible to run over-identification test.

assumption which leads to bias and inconsistency in OLS estimates. Clearly biasness and inconsistency arising from measurement errors are similar to those entailed by the conventional endogeneity problem of police staff (which assumes that counties with higher crime rates motivate to employ more police) leads to.

The probability of arrest, one of the law enforcement covariates, can also be affected by measurement errors in crime rates. The probability of arrest is defined as the ratio of number of arrests to the total crimes reported. If there is measurement error in reported crimes, then the effect of arrest rates on crime may lead to potentially biased estimates. What follows tries to shed some light on this issue. Suppose the true relationship between crime and arrests is as follows:

$$\ln(c_t) = \lambda \ln(\frac{a_t}{c_t}) + \varepsilon_t \qquad (3.3)$$

Where t indexes time periods, c is the actual number of crimes, a is the number of arrests, ε is a random noise term. From the economic model of crime we expect $\lambda < 0$. We cannot observe the actual number of crimes in equation (3.3), but rather the number of crimes that are reported to the police. Denoting the number of reported crimes as c^r, let this quantity varies according to:

$$c_t^r = \gamma_t c_t \qquad (3.4)$$

Where γ_t indicates the percentage of crimes reported to the police in period t. Combining equations (3.3) and (3.4), the actual specification estimated by the econometric model is:

$$\ln(\gamma_t c_t) = \lambda \ln(\frac{a_t}{\gamma_t c_t}) + \varepsilon_t \qquad (3.5)$$

Because the arrest rate is defined as the percentage of reported crimes, bias arises. Indeed, any stochastic error in the reporting rate (γ_t) will also appear in the denominator of the arrest rate. Ceteris paribus, when the reporting rate is high, the arrest rate is low, inducing a negative bias in λ . A downward bias in λ exaggerates the impact of arrests on the crime rate (Levitt, 1998).

Griliches and Hausman (1986) present a method that identifies and corrects for the presence of measurement error in panel data. According to them, the bias associated with measurement error is sensitive to the choice of estimator. For instance, it is acknowledged that estimating with first differences (FD estimator) tends to make worse the bias due to measurement error. It is possible to determine the degree of the measurement error problem and correct for it by comparing alternative estimators that are differentially affected by the presence of measurement error. Assume that we have the simple panel data model as follows:

$$y_{it} = \alpha_i + \beta z_{it} + \eta_{it} \qquad (3.6)$$

Where α_i are unobserved individual fixed effects, and η is an i.i.d error term with mean zero and variance σ_η^2 . However, we cannot observe the true z_{it} (say actual probability of arrest having number of crimes in its denominator), rather only an imperfect signal, x_{it} (probability arrest measured by reported crimes), is available:

$$x_{it} = z_{it} + v_{it} \qquad (3.7)$$

Where v_{it} is an i.i.d measurement error with variance σ_v^2 .

Finally, assume that the realization of z_{it} are positively serially correlated, as is common in economic time series:

$$\sigma_{z_t, z_{t-k}} = \rho^k \qquad (3.8)$$

Where $0 < \rho < 1$[48].

According to Griliches and Hausman (1986), we can compare the asymptotic bias in estimates of β by moving from a first difference estimator to second differences $[y_{it} - y_{it-2} = \beta(x_{it} - x_{it-2}) + (\eta_{it} - \eta_{it-2})]$ and other higher order differences. The intuition of Griliches and Hausman can carry over measurement error that affects both the right-hand and left-hand side variables presented in equation (3.5). Algebraic manipulation of this equation yields:

$$p\lim \hat{\lambda}_j - \lambda = -\{\sigma_\gamma^2(1+\lambda)/[\sigma_{a/c}^2(1-\rho^j) + \sigma_\gamma^2] \qquad (3.9)$$

Where λ_j is the estimator based on the *j*-th difference, σ_γ^2 is the variance in the reporting rate, $\sigma_{a/c}^2$ is the variance in the actual arrest rate and ρ is the serial correlation in the actual arrest rate. In the arrest-rate case, the measurement error coefficient is -1 (i.e., in the regression $\ln(\gamma) = \beta \ln(1/\gamma), \beta = -1$). As a consequence, the OLS estimate of reported crimes on the arrest rate with reported crimes in the denominator is a weighted average of the true coefficient and -1. Thus, errors-in-variables in the number of reported crimes lead to a bias that overstates rather than lessens the estimated coefficients, as long as $0 > \lambda > -1$ (the likely case in economic model of crime when examining elasticities of crime with respect to the arrest rate). As long as $0 < \rho < 1$, the magnitude of asymptotic bias declines as longer differences are used (Levitt, 1998).

[48] Possible sources of such correlation include non-stationarity of the underlying process, serial correlation in the shocks that determine z_{it}, partial adjustment in z_{it} due to informational or behavioral lags, or convex adjustment costs in z_{it} are among possible sources of such correlation (Levitt, 1998).

We can now apply the illustrated methodology to test whether measurement error from using reported instead of actual crimes imparts bias on the estimated impact of arrest rate. For this exercise we use the model specification in CT (1994). Table 3.5 presents crime estimates with the arrest rate when estimated using first differences through fourth differences of our model. Since the crime rate and the arrest rates are both in logs the coefficients represent elasticities. We do not present here the estimated parameters for the other covariates, which are of only secondary interest. As predicted by the economic model of crime, all coefficients are negative: an increase in the likelihood of arrest decreases the amount of crime. And all the estimates are statistically different from zero.

Table 3.5- Difference estimators and measurement error in arrest rate

	(1) First Difference	(2) Second Difference	(3) Third Difference	(4) Fourth Difference
Probability of Arrest	-0.287***	-0.269***	-0.28***	-0.213***
	(0.03)	(0.0317)	(0.0371)	(0.051)
N	477	393	313	234

If measurement error resulting from the use of reported crime is a problem, one would expect the coefficients to be monotonically declining in magnitude moving from column 1 to column 4 in table 3.5 since longer differences should be less sensitive to measurement error. The pattern in table 3.5 does not conform to the expectation, and even the estimates do not differ at the 5% significance level. In summary, there is little evidence that the use of reported crime rates includes a substantial bias in the estimated effects of arrest rates since estimators that are differently affected by such bias yield very similar results. Nonetheless, we still assume probability of arrest to be endogenous regarding explanations in previous section.

3.5 A Dynamic Panel Data model of crime

Starting from the theoretical framework, based on Becker (1968) and Ehrlich (1973), we apply a dynamic panel data (DPD) econometric model for dealing with likely bias due to weak identification in both CT (1994) and Baltagi (2006) estimation results presented in previous sections. The most complete econometric specification of our empirical model is as follows:

$$C_{i,t} = \alpha_0 + \alpha_1 C_{i,t-1} + X'_{i,t}\beta + \eta_i + \eta_t + \varepsilon_{i,t} \qquad i=1, 2. \ldots 90 \ , t=1, 2. \ldots 7 \tag{3.10}$$

Where C is crime rate, X is a matrix of a set of explanatory variables summarized in Table 3.1, η_i is unobserved county fixed effects, η_t is time fixed-effects and $\varepsilon_{i,t}$ is typical disturbance term, assumed to be i.i.d with zero mean and constant variance. The subscripts *i* and *t* represent county and time period, respectively.

Having crime inertia ($C_{i,t-1}$) as an explanatory variable in the model would lead "FE estimator" to be biased and inconsistent. Indeed, within transformation would lead to correlation between $C_{i,t-1}$ and the disturbance term, particularly in the case of "small T, large N". In this case the within estimator (FE estimator) will be biased[49] and its consistency depends upon T (sample periods) being large (Nickell, 1981). Including more regressors does not remove this bias. This bias is not caused by an autocorrelation in the error process (ε_{it}) and arises even if the error process is i.i.d. If the error process is auto-correlated the problem is even more severe given the difficulty of deriving a consistent estimate of AR parameters in that context[50]. The same problem of bias and

[49] This is called dynamic panel bias or Nickell's bias.

[50] It is worth emphasizing that only if T→ ∞ will the FE estimator be consistent in the DPD model.

inconsistency affects the Random Effects (RE) estimator. Even if we assume that the unobserved county specific effects (η_i) are uncorrelated with other regressors and thus a Random Effects (RE) estimator would be appropriate, this estimator will be biased too in a DPD model[51].

The Arellano-Bond (1991) estimator is appropriate in this case. It begins by specifying the model as a system of equations, one per each period (so it includes T equations) and allows the instruments applicable to each equation to differ (here for example, more lagged values of instruments may be used for more recent years). The instruments matrix includes both lagged level of endogenous variables and differenced strictly exogenous variables[52]. This estimator is designed for datasets with many panels and few periods (Large N, Small T), and it requires that there be no autocorrelation in the idiosyncratic errors.

In addition to controlling for the endogeneity of C, we can also control for the conventional endogeneity problem of the Probability of arrest and Police per capita. As mentioned above, instruments are suitable lags of the levels of the endogenous variables. This Arellano-Bond (1991) estimator can easily generate many instruments, since by period τ all lags prior to, say $\tau - 2$, might be individually considered as instruments. Here there is a danger of instrument proliferation which might cause bias and inconsistency. This caution has been taken care of in our estimation results by restricting the number of instruments[53].

---

[51] Technically, in order to apply GLS for estimation RE estimator, quasi-demeaning is applied in which $(C_{i,t-1} - \theta\overline{C}_{i,-1})$ is correlated with $(u_{it} - \theta\overline{u}_i)$ where u_{it} is error component including provincial specific effects and typical disturbance terms.

[52] It's also possible to add some excluded instrument, if available any, to the matrix of instruments. They are come in form of differences too.

[53] For more details on potential dangers of instrument proliferation in DPD models see Roodman (2009a).

There is a potential weakness in Arellano-Bond (1991) DPD estimator revealed by Arellano-Bover (1995) and Blundel-Bond (1998); the lagged levels are often rather poor instruments for first differenced variables. The solutions these authors propose is to include lagged levels as well as lagged differences as instruments. The augmented version of the difference GMM – the System GMM- can be considered that a two-equation systems is estimated (one in levels and the other in difference) where in levels endogenous variables are instrumented by the corresponding differenced lagged variables and in differenced equation endogenous variables are instrumented by suitable lagged levels of them as it is in Arellano-bond (1991) estimator. The original estimator, Arellano-Bond (1991) is often called "Difference GMM", whereas the expanded one suggested by Arellano-Bover (1995) and developed by Blundel-Bond (1998) is usually known as "System GMM". Again we have taken care of instruments proliferation in our system GMM estimator when we apply it.

The consistency of the parameters obtained by means of the GMM estimator depends crucially on the validity of the instruments. Two specification tests suggested by Arellano and Bond (1991) will be applied to check for validity of instruments. The first test is the Sargan/Hansen[54] test of over-identifying restrictions where the null hypothesis is overall validity of the instruments. Failure to reject this null hypothesis supports the choice of the instruments. We also report the test for serial correlation of the error term, where the null hypothesis is that the differenced error term is first and second order serially correlated. Failure to reject the null hypothesis implies that the moment conditions are correctly specified (for more details see Arellano and Bond (1991), Baltagi, 2005 and Baum, 2002).

[54] When we apply robust standard errors then Sargen test is not robust any more. In this case Hansen test is robust but it can be weakened by instrument proliferation. In all of our estimates standard errors are robust and so we automatically just report Hansen's over-identification J-test. We take care for instruments count to avoid reducing the power of Hansen test.

Considering the appropriate dimensions of North Carolina Dataset (T=7, N=90) which fits with “Small T, Large N”, we think Dynamic Panel Data (DPD) model has enough advantages to be applied. Past criminal experience affects the decision to commit a crime in several ways (Fajnzylber et al., 2002 a, b; Glaeser, Sacerdote, & Scheinkman, 1996; Sah, 1991, Bounanno & Montolio 2009, Enfort & Spengler 2000, Han et al, 2013). In other words, higher crime today is associated with higher crime tomorrow (i.e. persistence over time). There could be several reasons why crime rate can be thought to be correlated over time: (1) Criminals can learn by-doing and acquire an adequate criminal know-how level. This acquisition, in turn, means that the costs of carrying out criminal acts decreases over time (Case & Katz, 1991). (2) Recidivism caused by, among other things, negative expected payoffs from the labor market for being a criminal; convicted criminals have fewer opportunities for legal employment and a lower expected wage (Grogger, 1995); (3) Business cycle features such as recessions affecting the crime rate over successive periods. Furthermore, the lagged crime rate acts as a proxy for the lagged effects of variables such as lagged unemployment rate, lagged detection rate that explain the crime rate at current time. These arguments strongly suggest the possibility of criminal hysteresis or inertia. Furthermore, through a DPD model we can also deal with both conventional endogeneity and measurement error problems discussed in previous sections.

3.6 Results

In this section we present our estimates for a dynamic panel data model which is supposedly can better identify the economic model of crime. For comparative purposes the original estimate results of CT (1994), Baltagi (2006) will be reported. All our estimate results are for corrected version of dataset mentioned in previous sections. Like as original study of CT (1994) and its replication by Balatgi (2006), we also assume that explanatory variables Probability of arrest and Police per capita are endogenous as the endogeneity test couldn’t reject their joint endogeneity. Table 3.6 presents the estimate results. All GMM regressions are two-step and use robust standard errors, the Windmeijer (2005) finite sample corrected standard errors. For avoiding bias and inconsistency due to instrument proliferation in both 'Difference GMM" and "System GMM" the instrument matrix uses only two lags of endogenous variables as instrument. In all DPD models, county's tax per capita has been also included as excluded instrument.

In both Difference-GMM and System-GMM estimators, crime shows a significant inertia. Counties which have experienced more crimes in the past will continue to experience it in the future too. About law enforcement and its effects on the crime level, in both first and second columns law enforcement process including arrest, conviction and imprisonment show significant deterrent effects on the crime. However, their estimated elasticities in the case of System-GMM estimator (which is assumed to be superior to Difference-GMM) are lower[55]. In general, their estimated amounts are much lower than ones in the original CT (1994) and Baltagi (2006). For example, based on estimate results for System-GMM estimator, a 10% increase in probability of arrest (which is supposedly needs more police staff) decreases crime level only by 2.6%. Estimated elasticities for both conviction and imprisonment are also almost half of their associates in CT (1994) and Balatagi (2006). This bias in previous studies partly stems from using an insensible measure of conviction ($P_c > 1$) and partly due to their models weak identification. Severity of punishment proxied by average days of imposed sentence is small and statistically insignificant, possibly reflecting the fact that North Carolina has a

[55] The resulting "System GMM" estimator has been shown in Monte Carlo studies by e.g. Blundell and Bond (1998) and Blundell, Bond and Windmeijer (2000) to have much better finite sample properties in terms of bias and root mean squared error than that of the "Difference GMM" estimator. Blundell and Bond (1998) argued that the "System GMM" estimator performs better than the "Difference GMM" estimator because the instruments in the level equation (the lagged differenced) remain good predictors for the endogenous variables even when the series are very persistent.

policy of determinate sentencing. An alternative interpretation is that increasing in severity of punishment has no deterrent effect on criminals. Police per capita which is unexpectedly positive and significant in both Difference-GMM and EC2SLS columns (as it is in original CT (1994) and Balatgi (2006)) shows a much smaller positive effects in column System –GMM which is not even statistically significant.

Concerning to labor market incentives proxied by average weekly wages, only two of them show significant effects on the crime level. Based on estimate results of System-GMM estimator, a 10% increase in average weekly wage of employees in construction industry (which has one of the lowest average wages, see Table 3.1) will decrease crime rate by almost only 0.5% which is trivial. In a similar way, a 10% increase in average weekly wages paid in State government would decrease crime rate by 2.1%. This might reflect an increase in incentives of employees of law enforcement which are part of state government. Surprisingly, none of demographic and geographic covariates (except Central) shows significant effect on the crime.

The last rows of the Table 3.6 in first and second columns show the specification test for applied estimators along with instrument counts. Both sets of specification test (serial correlation and Hansen test) confirm that the model has identified well enough and over-identification moments are valid.

3.7 Conclusions

This chapter applies panel data on 90 counties in North Carolina over the period 1981-87 to estimate a Dynamic Panel Data model of crime. This dataset was originally used by CT (1994) to estimate a FE estimator. Baltagi (2006) replicated CT (1994) and after making some estimation correction, he suggested EC2SLS estimator as the appropriate one. Both CT (1994) and Baltagi (2006) considered conventional simultaneity problem for two explanatory variables - probability of arrest and police per capita. Both of these authors conclude that both law enforcement variables and some of labor market incentives matter in crime control.

First, we made some corrections in the dataset concerning the covariate probability of conviction which is proxied by the ratio of number of convictions to the number of arrests. Considering that any arrested suspect either convicts or acquits, this ratio must be between zero and one. However, there are 71 (almost 11%) observations for which this ratio is more than 1. This more likely happens due to mistakes in measurement of associated data. Our strategy for correction of dataset is to eliminate these observations. All of our estimates are based on this corrected version of the dataset. After doing this correction, Baltagi's suggested EC2SLS estimate for probability of arrest is much lower and statistically insignificant. However, CT's FE estimator shows negligible changes after correction.

Furthermore, CT and Baltagi's estimates crucially depend on the legitimacy of instruments applied by them. We have first queried the specification of their model and adopted alternative specifications that tackle both the problem of weak instruments and that of measurement errors.

We first queried the assumption of endogeneity of two key deterrence variables – probability of arrest and police staff per capita - and the quality of instruments used by CT and Baltagi to address the problem. We could confirm endogeneity using Davidson and Mackinnon (1993) test. Then different identification tests affirmed that the instruments they use turn out to be weak and econometric model they adopt turn out to be either under-identified or weakly identified. This implies that the estimates results are not reliable enough.

As an alternative we applied a Dynamic Panel Data (DPD) model to Becker's economic model of crime. In both Difference-GMM and System-GMM estimators, crime inertia is significant which is along with “hot spot” in criminology literature. In other words, counties with higher crimes rates likely will experience more crimes in the future. All deterrent variables have their expected sign and all of them, except severity of punishment (proxied by average sentence days) are highly significant. However, the elasticity of crime with respect to deterrent variables based on System-GMM estimator which is assumedly superior to Difference-GMM are much lower than ones estimated by CT (1994) and Baltagi (2006). The upward bias in deterrent variables in CT (1994) and Balatagi (2006) is partly due to using a probably mistaken measure of probability of conviction and partly due to weakly identified models. These models are either under-identified or weakly identified due to weak instrument applied by them. Our estimates for deterrent variables using a DPD model are likely more reliable than previous efforts done by CT (1994) and Balatgi (2006).

Table 3.6- Estimates results

	(1) Diff-GMM	(2) Sys-GMM	(3) EC2SLS	(4) Baltagi (2006)	(5) CT (1994)
L.Crime	0.329*	0.567***			
	(0.159)	(0.112)			
Pa	-0.396*	-0.264**	-0.0857	-0.413***	-0.355***
	(0.170)	(0.0876)	(0.113)	(0.0974)	(0.0322)
Pc	-0.225**	-0.175***	-0.148**	-0.323***	-0.282***
	(0.0748)	(0.0430)	(0.0452)	(0.0536)	(0.0211)
Pp	-0.184**	-0.127**	-0.121**	-0.186***	-0.173***
	(0.0578)	(0.0375)	(0.0413)	(0.0419)	(0.0323)
S	-0.0353	-0.0569	0.00760	-0.0102	-0.00245
	(0.0306)	(0.0351)	(0.0261)	(0.0270)	(0.0261)
Police	0.610***	0.159	0.349***	0.435***	0.413***
	(0.153)	(0.141)	(0.0994)	(0.0897)	(0.0266)
Density	0.0756	0.0935	0.463***	0.429***	0.414
	(0.549)	(0.0595)	(0.0534)	(0.0548)	(0.283)
Wcon	-0.0498	-0.0497*	-0.0566	-0.00748	-0.0378
	(0.0283)	(0.0214)	(0.0365)	(0.0396)	(0.0391)
Wtuc	0.0163	0.0215	0.0553**	0.0454*	0.0455*
	(0.0194)	(0.0226)	(0.0184)	(0.0198)	(0.0190)
Wtrd	-0.0345	-0.0221	-0.0248	-0.00814	-0.0205
	(0.0263)	(0.0442)	(0.0363)	(0.0414)	(0.0405)
Wfir	0.0170	0.0231	-0.00708	-0.00364	-0.00390
	(0.0107)	(0.0233)	(0.0501)	(0.0289)	(0.0283)
Wser	0.0164	0.0129	-0.00496	0.00561	0.00888
	(0.0257)	(0.0250)	(0.0181)	(0.0201)	(0.0191)
Wmfg	-0.00193	-0.0156	-0.0696	-0.204*	-0.360**
	(0.182)	(0.105)	(0.0859)	(0.0804)	(0.112)
Wfed	0.119	0.197	-0.102	-0.164	-0.309
	(0.200)	(0.154)	(0.143)	(0.159)	(0.176)
Wsta	-0.133	-0.210*	-0.0384	-0.0540	0.0529
	(0.140)	(0.0929)	(0.0976)	(0.106)	(0.114)
Wloc	-0.185	-0.0620	0.122	0.163	0.182
	(0.128)	(0.129)	(0.122)	(0.120)	(0.118)
Pym	0.229	0.00408	0.0618	-0.108	0.627
	(0.609)	(0.0757)	(0.124)	(0.140)	(0.364)
Pmin		0.0643	0.122**	0.189***	
		(0.0410)	(0.0426)	(0.0415)	
West		-0.0967	-0.292**	-0.227*	
		(0.0696)	(0.100)	(0.0996)	
Central		-0.0850*	-0.205***	-0.194**	
		(0.0387)	(0.0600)	(0.0598)	
Urban		-0.00253	-0.125	-0.225	
AR(1)	[0.071]	[0.001]	(0.107)	(0.116)	
AR(2)	[0.474]	[0.313]			
Instrument count	48	68			
Hansen J-test	[0.42]	[0.307]			
N	393	477	559	630	630

Robust Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Bibliography

- Ajani, G. Mattei, U. (1995). Reforming Property Law in the Process of Transition. Some Suggestions from Comparative Law and Economics, *Hastings International and Comparative Law Review, 19, 117–37*
- Angrist, J. Pischke, J. (2008). Mostly harmless econometrics. *Princeton University Press*.
- Arellano, M. Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and application to employment equation. *Review of Economic Studies* 58(2):277-297.
- Arellano, M. Bover, O. (1995). Another look at the instrumental-variable estimation of error component models. *Journal of Econometrics* 68(1): 29-52.
- Baltagi, BH. (2004). A Hausman test based on the difference between fixed effects two-stage least squares and error components two-stage least squares, *Econometric Theory* 20: 223–224.
- _________. (2006). Estimating an economic model of crime using panel data from North Carolina. *Journal of Applied Econometrics*. 21:543-547.
- _________. (2005). Econometric Analysis of Panel Data. 3th edition. *John Wiley & Sons Ltd.*
- Baum, Ch. (2006). An Introduction to Modern Econometrics Using Stata. *Stata Press, Texas.*
- Bean, P. (1981) Punishment: A Philosophical and Criminological Inquiry. *Oxford, England: Martin Robertson*.
- Beccaria, C. (1809). An Essay of Crimes and Punishment. *Philadelphia: William P. Farrand and Co.*

- Becker, G. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76(2), 169–217.
- Bentham, J. (1789). An Introduction to the Principles of Morals and Legislation, *Hafner Press (1948).*
- Blundell, R. Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87(1): 115-143.
- Blundell, R., S. Bond, and F. Windmeijer (2001). Estimation in dynamic panel data models: Improving on the performance of the standard GMM estimator. In, B.H. Baltagi, T.B. Fomby, and R. Carter Hill (eds.), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels. Advances in Econometrics*, Volume 15, Emerald Group Publishing Limited, 53–91.
- Bounanno, P. Montolio, D. (2008). Identifying the socio-economic and demographic determinants of crime across Spanish provinces. *International Review of Law and Economics* 28: 89-97
- Case, A. Katz, L. (1991).The Company you keep: the effects of family and neighborhood on disadvantaged youth. *NBER working paper 3705.*
- Cornwell, C. Trumbull, WN. (1994). Estimating the economic model of crime with panel data. *Review of Economics and Statistics* 76: 360–366
- Cragg, J. G, Donald, S. G. (1993). Testing Identifiability and Specification in Instrumental Variable Models. *Econometric Theory* 9: 222–240.
- Caterina, R. (2006). Comparative law and Economics, in Elgar Encyclopedia of Comparative Law, edited by Jan Smiths, *Edward Elgar Publishing, Inc. pp 161-172*
- Corman, T. Mocan, D. (2000). A time series analysis of crime, deterrence and drug abuse in New York City, *American Economic Review*, Vol 90, No.3, 584-604.
- Cherry, T. List, J. (2001). Aggregation bias in the economic model of crime, *Economics Letters 75,* 81-86.

- Dubber. M. (2006). Criminal Law in Comparative Context, *Journal of Legal Education*, Vol 56 (433-443).
- Devine, J. Sheley, J. Smith, M. (1988). Macroeconomic and social control policy influences on crime rate changes, 1948-1985, *American Sociological Review*, Vol 53, No 3, 407-420.
- Davidson, R. MacKinnon, J. (1993). Estimation and inference in Econometrics. *Oxford University Press*. New York.
- Eide, E. Rubin, P. Shepherd, M. (2006). Economics of crime, *Now Publishers Inc*.
- Ehrlich, I. (1973). Participation in illegitimate activities: a theoretical and empirical investigation. *Journal of Political Economy* 81: 521–567.
- Enfort, H. Spengler, H. (2000). Socioeconomic and demographic factors of crime in Germany: evidence form panel data of the German states. *International Review of Law and Economics* 20(1): 75-106.
- Fajnzilber, P et al. (2002a). Inequality and violent crime. *Journal of Law and Economics* 45(1): 1-40.
- Fajnzilber, P et al. (2002b). What causes violent crimes? *European Economic Review*. 46: 1323-1357.
- Forst, B. (1976), Participation in illegitimate activities: further empirical findings, 2 *Policy Analysis*, 477-492.
- Frank, H. (1991). Cocaine addition: *psychology and neurophysiology, Science*, 251: 1580–86.
- Freeman, R. B. (1996). Why do so many young American men commit crimes and what might we do about it? *Journal of Economic Perspectives*, 10(1), 25–42.
- Furlong, W. Mehay, S. (1981). Urban law enforcement in Canada: an empirical analysis, *The Canadian Journal of Economics,* Vol 14, No1, 44-57.

- Gawin, Frank H. 1991. "Cocaine addition: Psychology and Neurophysiology," Science, 251: 1580–86.
- Glaeser, E. et al. (1996). Crime and social interactions. *Quarterly Journal of Economics* 111(2): 507-548.
- Griliches, Z. Hausman, J. (1986). Errors in variables in panel data. *Journal of Econometrics* 31:93-118.
- Gogger, J. T. (1998). Market wages and youth crime. *Journal of Labor Economics*, 16(4), 756–791.
- ____________ (1995). The effects of arrest on the employment and earning of young men. *Quarterly Journal of Economic* 110(1): 51-72.
- Gellen, P. (2006). Aims of Comparative Law, in Elgar Encyclopedia of Comparative Law, edited by Jan Smiths, *Edward Elgar Publishing, Inc*. pp 57-66.
- Gilboa, I. (2010). Rational choice. *The MIT Press.*
- Gould, E. Weinberg, B. Mustard, D. (2002). Crime rates and local labor opportunities in the United States: 1979-1997, *The Review of Economics and Statistics*, Vol 84, No1, 45-61.
- Haist, M. (2009) Deterrence in a sea of "just desert": are utilitarian goals achievable in a world of limiting retributivism? *The Journal of Criminal Law and Criminology*, 99(3), 789-822.
- Han et al (2013). Determinants of violent and property crimes in England and Wales: a panel data analysis. Applied Economics. 45(34): 482-4830.
- Hayashi, F. (2000). Econometrics. *Princeton University Press.*
- Hillsman, Sally T. (1990), 'Fines and day fines', in Michael Tonry and Norval Morris (eds), Crime and Justice. A Review of Research, vol. 12, Chicago: *University of Chicago Press*, pp. 49–98
- Hirsch, A. (1986) Doing Justice. *Northeastern University Press.*
- Holtz-Eakin, D. W. Newey, and H. S. Rosen. (1988). Estimating vector autoregressions with panel data. *Econometrica* 56: 1371–1395.

- Hudson, B. (1996) Understanding Justice: An Introduction to Ideas, Perspectives and Controversies in Modern Penal Theory. *Buckingham, England: Open University Press*.
- Kant, I. (1991). The Metaphysics of Morals, *Mary Gregor trans., Cambridge*.
- Kaplow, L. Shavell, S. (2006). Fairness versus welfare. *Harvard University Press.*
- Kelly, M. (2000). Inequality and crime, *The Review of Economics and Statistics,* Vol 82, No. 4, 530-539.
- Kleibergen, F. Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition, *Journal of Econometrics*, 133(1): 97-126.
- Lewin, J. Trumbull, W. (1990). The social value of crime? *International Review of Law and Economics,* 10: 271-84.
- Levitt, S. (1998). Why do increased arrest rates appear to reduce crime: deterrence, incapacitation or measurement error? , *Economic Inquiry* 36, No 3, 353-72.
- Lowenstein, G. (2004). Out of control: visceral influences on behavior, ch26, Advances in behavioral economics, *Princeton University press.*
- Mattei, U. (1994). Efficiency in Legal Transplants: an Essay in Comparative Law and Economics, *International Review of Law and Economics.*
- Mustard, D. (2003). Reexamining criminal behavior: the importance of omitted variables bias, *The Review of Economics and Statistics,* Vol 85, No.1, 205-211
- Nagin, D. (1978). General deterrence: a review of the empirical evidence, in Blumstein, Cohen and Nagin (eds), Deterrence and incapacitation: estimating the effects of criminal sanctions on crime rates, *Washington, DC, National Academy of Sciences*, 95-139.

- Nagin, D. Pepper, J. (2012). Deterrence and the death penalty. *The National Academic Press.*
- Nickell, S. (1981). Biases in dynamic models with fixed effects, *Econometrica* 49:1417–1426.
- Polinsky, M. Shavell, S. (2000). The economic theory of public enforcement of law. *Journal of Economic Literature,* pp 45-76.
- Pratt, J. (2002). Punishment and Civilization: Penal Tolerance and Intolerance in Modern Society. *SAGE Publication.*
- Raphael S, Winter-Ember R. (2001). Identifying the effect of unemployment on crime, *Journal of Law and Economics*, Vol 44, No. 1, 259-283.
- Rawls, J. (1955). Two concepts of rules. *Philosophical Reviews.* Vol 10.
- Robinson, P. & Darley, J. (1997). The utility of desert. *Northwestern University Law Review*, 91(2), 453-499.
- Robinson, P. (2008) Competing Concepts of Modern Desert: Vengeful, Deontological and Empirical. *Cambridge Law Journal*, 67(1), pp. 145-175.
- __________. (2008) Distributive Principles of Criminal law: Who should be Punished How Much? *Oxford University Press.*
- Roodman, D. (2009a) A Note on the Theme of Too Many Instruments. *Oxford Bulletin of Economics and Statistics*, 71(1).
- __________. (2009b) How to do Xtabond2 in Stata: an Introduction to Difference and System GMM in Stata. *The Stata Journal*, 9(1), pp. 86-136.
- Rowan, M. (2011), Democracy and Punishment: a Radical View, *Theoretical Criminology*.
- Sah, K. (1991). Social osmosis and pattern of crime. *Journal of Political Economy* 99(6):1271-1295
- Saridakis, G. Spengler, H. (2012). Crime, deterrence and unemployment in Greece: a panel data approach, *The Social Science Journal*, Vol 49, 167-174.

- Schelling, Thomas. 1984. “Self-command in Practice, in Policy, and in Theory of Rational Choice.” American Economic Review, 74: 1–11.
- Schissel, B. (1992). The influence of economic factors and social control policy on crime rate changes in Canada, 1962- 1988, *The Canadian Journal of Sociology,* Vol 17, No.4, 405-428.
- Stigler, G.J. (1970). The Optimum Enforcement of Laws. *Journal of Political Economy*, 78, 526-536.
- Stock, J. Yogo, M. (2005). Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, edited by Donald W.K. Andrews and James H. Stock, Cambridge: *Cambridge University Press.*
- Stock, J. Wright, J. (2000). GMM with weak identification. *Econometrica* 66(5): 1055-1096.
- Tunick, M. (1992) Punishment: Theory and Practice. *University of California Press*.
- Tyler, T. Boeckmann, RJ. (1997). Three strikes and you're out, but why? The psychology of public support for punishing rule breakers. *Law & Society Review 31 (2): 237-65.*
- Ulen, T. (1997). Firmly Grounded: Economics in the Future of the Law, *Wisconsin Law Review*, 433–63.
- Vandaele, W. (1978). Participation in illegitimate activities: Ehrlich revisited. In: Deterrence and incapacitation, eds A. Blumstein, J. Cohen and D. Nagin, *National Academy of Sciences, Washington DC*, 270–335.
- Watson, Alan (1981), the Making of the Civil Law, Cambridge, MA: *Harvard University Press.*
- Wiegend, T. (2006). Criminal law and criminal procedure, in Elgar Encyclopedia of Comparative Law, edited by Jan Smiths, *Edward Elgar Publishing*, Inc. pp 214-228.

- Windmeijer, F. (2005). A finite sample correction for the variance of linear efficient two-step GMM estimators. *Journal of Econometrics*, 126, 25–51.
- Withers, G. (1984). Crime, punishment and deterrence in Australia: an empirical investigation, *Economic Records 60,* 176-185.
- Wolpin, K. (1978). An economic analysis of crime and punishment in England and Wales, 1894-1967, *The Journal of Political Economy*, Vol 86, No.5, 815-40.
- Zimmerman, P. (2010). The Economics of Capital Punishment and Deterrence, in Handbook on economics of crime edited by Benson & Zimmerman. *Edward Elgar Publishing, Inc.*