

Comparison of Criminal and Administrative Penalties for Environmental Offenses

Thomas Blondiau

TMLeuven, Diestsesteenweg 57, B-3010 Leuven, Belgium; E-mail: Thomas.Blondiau@tmleuven.be

Carole M. Billiet

U.Gent, CM&ER, Universiteitsstraat 4, B-9000 Gent, Belgium; E-mail: Carole.Billiet@ugent.be

Sandra Rousseau

KULeuven, CEDON, Warmoesberg 26, B-1000 Brussel, Belgium; E-mail: Sandra.Rousseau@kuleuven.be

CITE AS: Blondiau, T., Billiet, C.M. & S. Rousseau (2015). Comparison of criminal and administrative penalties for environmental offenses. *European Journal of Law and Economics*, vol. 39(1), p. 11-35. (Doi: 10.1007/s10657-014-9473-9)

Abstract

We conduct a counterfactual analysis to measure the treatment difference between administrative and criminal enforcement of environmental violations. Our aim is to control for the selection bias effect and make a comparison of how similar offenses are treated in both enforcement tracks. This analysis is relevant to answer the question of how deterrence and welfare in a combined criminal-administrative enforcement system compare with an enforcement system relying on criminal penalization only. This is an important question given the observed shift towards two-track penalization systems, combining administrative and criminal penalization, for environmental enforcement in the European Union. We apply matching techniques on a unique dataset of environmental enforcement cases to control for sample selection bias. We match different enforcement cases, based on case characteristics and estimate the average treatment effect for these cases. Overall, we find that the marginal penalty is slightly lower in administrative enforcement compared to criminal enforcement.

Keywords: Environmental law enforcement; Criminal and administrative enforcement; Propensity score matching; counterfactual analysis

JEL codes: C14 Semiparametric & nonparametric methods; K32 Environmental Law; K42 Illegal behavior and enforcement of law; Q58 Environmental economics – government policy

1 Introduction

The aim of this paper is to contribute to the debate on the relative costs and benefits of using criminal versus administrative enforcement for environmental policy. On the one hand, the official policy of the EU puts a strong emphasis on stimulating the use of criminal enforcement and aims at strengthening enforcement systems that rely strongly on administrative enforcement (European Commission, 2008). On the other hand, several European countries that predominantly relied on criminal enforcement in the past, such as the UK and Belgium, recently moved towards more administrative enforcement (Svatikova, 2012). The overall picture emerging from these developments is a move towards combined criminal-administrative enforcement systems (Billiet, 2012). In this paper we analyze a combined criminal-administrative enforcement system used in one region of Belgium and compare it to a criminal enforcement system used in another region of the country.

One of the main benefits of any enforcement policy is its capacity to prevent individuals or firms that plan on breaching the law from doing so, i.e. deterrence. According to the rational crime model (Becker, 1968), higher deterrence typically leads to a higher compliance rate¹. This is welfare increasing as long as the regulatory framework is well-designed and promotes efficiency (Svatikova, 2011). However, even if regulations are inefficient, effective enforcement can be desirable, for instance, to promote fairness or to ensure a level playing field. In this paper, we will generally assume that the regulations to be enforced are reasonable and desirable and thus that social welfare is promoted by enforcing them. In the context of environmental regulations, the empirical evidence concerning the effectiveness of enforcement policies (including inspections and sanctions) is vast. Gray & Shimshack (2011), for instance, argue that an effective monitoring and enforcement regime has been a key factor in the improvement of environmental quality over past decades. In contrast, the evaluation of voluntary, informational and cooperative compliance systems seems to provide mixed results. Thus, Gray & Shimshack (2011) argue that alternative regulatory strategies can, for now, at best be considered as a useful complement to an effective inspection and punishment regime.

However, the debate concerning the most appropriate enforcement path for environmental offenses is still unresolved. Several policy papers, such as Hampton (2005) and Macrory (2006) for the UK, argue that, in practice, exclusive reliance on criminal enforcement is not the most cost-effective path for environmental regulation enforcement. They found that the then applied UK system with its dominant criminal enforcement, was inefficient, costly and time-consuming. The results included low punishment probabilities, low expected financial penalties, and poor deterrence of true criminals. The EU enforcement policy, on the other hand, stresses that enforcement systems without a criminal track have not been able to achieve a sufficient protection of the environment and that criminal penalties are needed because these are thought to be more dissuasive (European Commission, 2008).

We now look at the advantages of using (more costly) criminal enforcement instead of (cheaper) administrative enforcement. As a normative benchmark, the concept of the enforcement pyramid (Ayres & Braithwaite, 1995) is quite useful. This concept states that only the offenders who commit the most serious and harmful violations should be criminally prosecuted, while less serious offenses can be dealt with using administrative enforcement. The largest group of small offenders may be treated in a rather co-operative way, using the possibility of administrative warnings. The main goal of criminal prosecution is deterrence, given that only the most serious cases should, in principle, be taken to court. To achieve this goal, criminal courts seem to be more effective given that all punishment tools (such as closure of firms and imprisonment of persons) are at the criminal judge's disposal. Bowles et al. (2008)

¹ For an overview of theoretical literature extending the optimal law enforcement framework, see Garoupa (1997) or Polinsky & Shavell (2007). For a view on empirical literature, one can consult Levitt & Miles (2007).

apply an economic approach to assess which case characteristics make criminal or administrative enforcement more efficient. They state that it is better to use the criminal enforcement system in cases concerning large scale damages, because the damages caused are disproportionate compared to prevention costs, or when behavior of the offender exceeds socially acceptable boundaries. Additionally, criminal penalties should also be preferred in situations where detection probabilities are low. There is also good reason to use criminal enforcement when the defendant is judgment-proof by lack of financial means, namely to provide marginal deterrence through prison sentences. Finally, criminal enforcement is more appropriate if high accuracy is required. The aim of the additional procedural safeguards of criminal enforcement and high burden of proof is precisely to prevent legal mistakes. In contrast, Bowles et al. (2008) also evaluate a number of disadvantages associated with criminal enforcement. They see it as less appropriate in enforcement policy areas where the possibility of over-deterrence exists (which may be the case for environmental offenses in contrast to, for example, murder or rape), when high standards of evidence lead to a disproportionate increase in enforcement costs, if specialization of an enforcement body is needed, or when a lot of flexibility is required in the enforcement procedures². Much attention has also been given to the potential deterrence effect of social stigma associated with criminal convictions. Shavell (1993) considers this as one of the important virtues of criminal law. Rasmusen (1996) also stresses private punishment following from social stigma as an important deterrence effect. Funk (2004) shows that social stigma may also have negative consequences, through increased possibility of recidivism and thus more criminal behavior. Finally, Galbiati & Garoupa (2007) analyze why social stigma is more often associated with a criminal conviction than with an administrative penalty. The reason is that criminal conviction is harder to obtain and thus provides a more reliable signal about someone being guilty.

Our aim is to evaluate how a combined criminal-administrative enforcement system compares to a single-track criminal enforcement system in terms of the penalties that they administer for similar offenses. Is it the case that there are wide disparities between both systems or do they actually perform quite similarly? To our knowledge, Svatikova (2011) is the first to analyze the relative performance of different enforcement systems in terms of their effect on (economic) welfare. She develops a formal model to compare the use of a criminal enforcement system with a combined criminal-administrative enforcement system. According to the basic deterrence model (Becker, 1968), potential violators will offend the law as long as the expected sanction is lower than the benefit obtained. Formally, with p defining the probability of a receiving penalty and F the penalty level, they violate if: $p \times F < Benefit$. The probability of receiving a penalty p is actually a reduced representation of a series of probabilities: namely, the probability of detecting the offense, the probability of a sanctioning procedure being started, the probability that a party is found to be guilty, and the probability that an enforcement action is taken for guilty parties. Each of these probabilities may be influenced by the type of enforcement system in place. For example, an environmental inspector might be more diligent in detecting minor offenses, if there is a high probability that an administrative fine will be imposed rather than facing a high probability of dismissal during a criminal procedure. Or, the probability of finding a party guilty can be lower in a criminal procedure than in an administrative procedure due to a higher standard of proof. Furthermore, we also need to consider the nature of the benefits obtained from an offense (Bowles et al., 2000). These benefits can include delayed costs, avoided costs, or economic benefits such as an increased market share or additional income from illegal activities (EPA, 2005).

Overall, Svatikova (2011) needs a few critical assumptions to develop her model. A first assumption is that, for similar offenses, the probability that a penalty is imposed is higher in a combined (criminal-administrative) prosecution system than in a criminal enforcement system. This assumption is quite

² Cooter & Ulen (2007) also discuss conditions for applying different types of enforcement procedures.

plausible (Macrory, 2006) and also seems to be corroborated by the available empirical evidence for Belgium (MHR, 2009). Secondly, she assumes that the marginal penalty is equal in an administrative and a criminal enforcement track. This implies that penalties for similar offenses are more or less equal. Combining both effects, Svatikova concludes that introducing administrative enforcement leads to higher deterrence, because of higher expected penalties. So in terms of deterrence, the hypothesized effect of moving to a combined enforcement system can be summarized as follows: the probability the an offense is sanctioned increases, thus the expected costs of non-compliance increase for potential offenders and this leads to a lower number of environmental offenses.

In this paper, we focus on measuring the relative difference in stringency between both enforcement systems for similar types of offenses. There is a lack of research on the (absence of a) treatment difference between criminal and administrative penalization for environmental violations. Therefore, it seems to us that the equality of penalties in both enforcement tracks has been assumed without much foundation. This leads us to our research question: *Is there any difference between the penalty imposed in criminal and administrative prosecution for similar violations? And, if so, how large is this difference?*

We restrict our attention to a comparison of the monetary penalties imposed through different enforcement systems. Thus, we abstract from issues such as social stigma and non-monetary penalties. This is mainly for feasibility reasons, as it is very difficult to exactly measure social stigma. Moreover, the use of non-monetary penalties is much more restricted in administrative than in criminal enforcement, as prison sentences do not exist in the administrative track, which complicates a fair comparison. Still, this restriction should not be inflated because non-monetary penalties are not frequently imposed for environmental offenses (e.g. Gray & Shimshack, 2011).

We use econometric tools to measure the relative treatment difference between both enforcement tracks. It is generally known that sample selection bias is often a serious concern in this type of empirical legal analysis. The empirical researcher would ideally want to use a randomized design to make correct inferences such as in Ashenfelter et al. (1995) or Telle (2013). Unfortunately, randomization may not always be feasible in practice. The use of matching for policy evaluation when only observational data are available is well established in the field of economics, especially for evaluating labor market programs³. However, to our knowledge, the application of matching in the context of empirical legal studies is relatively scarce. Therefore, this paper can provide an innovative way to take sample selection bias into account, which is a problem that is prevalent in empirical legal analysis. Our empirical strategy consists of identifying and matching similar cases where similarity is based on a set of explanatory variables. Then, we measure treatment differences between criminal and administrative enforcement using our matched pairs. We use a unique dataset of Belgian environmental law enforcement cases, in which penalties are imposed by judges and administrations for environmental offenses in 2003-2006.

The paper is outlined as follows. In section 2 we give some general information on our dataset and the legal context of environmental enforcement in Belgium. In section 3, we present the methodology. Section 4 defines the variables used. In section 5, we discuss the results. Section 6 concludes.

2 Data and legal context

Our empirical analysis can be relevant for environmental regulation enforcement in other jurisdictions since an increasing number of Western European countries use a mix of administrative and criminal enforcement regimes. A recent OECD study (OECD, 2009) finds that national enforcement strategies are

³ For an overview of the theoretical literature on statistical matching, see Heckman et al. (1999) or Imbens & Wooldridge (2009). For an application in regulatory impact analysis, see Milimet & List (2004). For some examples of matching in environmental economics, see Ferraro et al. (2007) or Deng et al. (2011).

often substantially different. However, international comparisons tend to be blurred by a sample selection effect. For instance, if one compares average criminal penalties in the US (about €178 000) with those in the UK (around €10 000), one gets the impression that courts in US are much stricter. However, comparing these numbers is biased because the selection of environmental cases into the criminal enforcement track as well as the determination of the sanction levels are likely to be different in both countries (see, e.g., UK Sentencing Advisory Panel, 2000, and US Sentencing Commission, 2008). Moreover, the objectives pursued by the regulators can also differ. For example, the US non-compliance penalties are typically designed to take away any benefit from non-compliance and to make potential offenders indifferent between compliance and non-compliance (Cohen, 2000). Thus, sample selection problems make the comparison of legal regimes in terms of treatment effects more complex. Next, we provide a description of our dataset and of the environmental enforcement process in Belgium.

2.1 Data

We use data from the SBO project “Environmental law enforcement: a comparison of practice in the criminal and the administrative tracks” (www.environmental-lawforce.be). Subsets of this dataset have been used in other papers, such as Billiet et al. (2009), Blondiau & Rousseau (2010) and Billiet et al. (2014). Specifically, we use first instance court judgments and administrative penalties that were imposed between 2003 and 2006 in Flanders and Brussels⁴. Table I presents some key characteristics.

Table I: Descriptive statistics of our dataset

| | | Criminal court cases | Administrative cases |
|---------------------|-------------------------------------|----------------------|----------------------|
| Full dataset | <i>Number of cases</i> | 1034 | 624 |
| | <i>Number of defendants</i> | 1617 | 640 |
| Used subset | <i>Number of firms</i> | 251 | 262 |
| | <i>Number of convicted firms</i> | 189 | 174 |
| | <i>Number of postponed verdicts</i> | 32 | 81 |
| | <i>Number of acquitted firms</i> | 30 | 7 |

The criminal cases in our dataset are collected in the Flemish Region, more specifically within the area of the Court of Appeal of Ghent. The administrative cases all relate to the Brussels-Capital Region. For each of these cases, we have information related to offense characteristics (such as cause of the offense, type of contamination, etc.) and offender characteristics (such as whether the offender is a repeat offender, whether he made efforts to clean-up or mitigate environmental damages, etc.). As the sanctions are imposed per defendant and not per prosecution case, our observations for the empirical analysis concern the numbers of defendants. Multiple defendants can be accused in one single court case.

We restrict our sample by excluding the administrative cases dealing with noise caused by airplanes. These cases represent a specific regional effect since the Belgian national airport in Zaventem is located next to Brussels. Further, we only include firms to improve the comparability of administrative and criminal cases in our dataset. In criminal prosecution law there are powerful incentives to prosecute natural persons who are involved in an environmental crime together with the legal persons involved. Therefore, in the criminal enforcement path, a significant share of defendants consists of natural persons. Similar incentives are not present in the administrative track, which thus mainly prosecutes

⁴ Belgium is a federal state with three regions: the Flemish Region (Flanders), the Walloon Region and the Brussels-Capital Region (Brussels).

legal persons. Since penalties on legal persons are typically much higher than those on natural persons, it is clear that focusing on firms significantly improves comparability.

Judges decide in each case between conviction, acquittal and postponement of the verdict. Acquittal is only possible if either the offense or the culpability of the defendant cannot be proven sufficiently. Postponement actually means that the case is closed without penalty, because the judge does not think that a particular offense should be penalized. Equivalent choices are available to the administrative penalization official and these decisions are called conviction, technical acquittal and opportunity acquittal respectively. It is mandatory to acquit defendants for which the offense could not sufficiently be proven, both in the administrative and the criminal track. Therefore, we exclude the acquittals⁵ in our empirical analysis to focus on treatment differences caused by issues over which judges or penalization officials have discretionary decision power. However, we add a sensitivity analysis where we do include acquittals in our dataset to analyze whether differences in the burden of proof affect our results.

2.2 The criminal and the administrative tracks in a combined relation

In Belgium the federal and regional legislators tend to combine criminal and administrative enforcement systems in a complementary way. The introduction of administrative fining systems is a key element in the legislative policy over recent years. This evolution has been made especially clear during the last fifteen years, in environmental law as well as other policy fields. These combined enforcement systems are often based on the example of the enforcement system introduced some forty years ago in labor and social security law. The co-existence of criminal and administrative penalization is organized using a priority rule that gives priority to criminal penalization.

The procedure starts with a notice of violation, documenting one or more particular offenses, which is sent to the public prosecutor's office. The public prosecutor has a certain time period to decide on the treatment of the case: criminal prosecution, administrative fining prosecution, a settlement offer or dismissal of the case. Public prosecutors in Flanders decide to prosecute in about 8 % of the environmental cases that reach their office (Billiet et al., 2014)⁶. Prosecution guidelines regarding the decision to prosecute and the type and level of sanctions to request from the bench do not exist in Belgian criminal procedure law. In May 2000 the Council of Prosecutors-General did approve a memorandum identifying the environmental offenses that rate priority in the prosecution policy. The prioritized offenses are essentially those that have or might have serious consequences for public health and the environment, have an organized crime character, are committed in a professional context, or concern the exploitation of a facility or activity without the required environmental permits. This memorandum, however, does not bind the prosecutors at all. Thus, the prosecutor's discretion concerning the decision to prosecute or not and, if prosecuting, the selection of sanctions to request, is very broad. Note that the public prosecutor's role is somehow to act as a playmaker. In the empirical analysis we take the decision process of the prosecutor as given. When he decides to keep away from criminal procedures, the case can be transferred to an administrative fining official.

The competence of this official to impose a fine only arises after the prosecutor has decided not to pursue criminal penalization. His competence in this type of procedure is twofold. The official can impose a fine if he considers that the offense and the culpability for the offense are sufficiently proven. However, he is not obliged to impose a fine. He can also choose to close the case without fining the offender for opportunity reasons, for instance because the original (minor) problem is adequately solved

⁵ Next we use the term 'acquittal' to represent acquittals in the criminal track as well as technical acquittals in the administrative track. We use 'opportunity acquittal' to present opportunity acquittals in the administrative track.

⁶ Some 60% of the Flemish environmental cases end with a dismissal and 14% with a settlement (Billiet et al., 2014). The remainder is merged with previously existing files or referred to other courts.

(opportunity acquittal). Beyond this first decision, to punish or not, the fining official who decides to punish also has the competence to decide on the level of the penalty, anywhere between a maximal and a minimal fine. These competences of the fining official are somehow parallel to the competence that a formal judge has in this matter. The main difference is that some additional sanctioning tools are available to criminal judges, such as prison sentences for natural persons, forfeiture of illegal gains, etc.

This is the description of a typical Belgian combined criminal-administrative enforcement system for controlling environmental crime, as it was introduced in the Brussels-Capital Region in 1999. Ten years later, in 2009, this system was also introduced in the other two Belgian regions, namely the Flemish and Walloon Regions. Before 2009, the choice set of the public prosecutor in both of those regions was limited to criminal prosecution, settlement proposal or opportunity dismissal. We exploit this difference in timing by comparing the treatment of cases administered in the Flemish Region where an ‘old enforcement system’ was still in place with the cases in Brussels where a ‘new enforcement system’ was already operational. Our goal is to compare the criminal and administrative penalties within the framework we just sketched, in order to assess whether both enforcement systems are similar in their treatment of environmental offenders.

3 Methodology

In this section, we explain our empirical strategy. Our aim is to estimate the average difference between the treatment of defendants in the administrative and in the criminal track. From now on, when we refer to the ‘treatment group’, we mean the group of administrative cases; these are the cases that may receive administrative treatment in the Brussels-Capital Region instead of criminal enforcement. The ‘control group’ is a set of criminal cases in the Flemish Region. These are used as a benchmark to evaluate what would have happened to the administrative defendants if they would have been criminally prosecuted. The relative difference between both groups is the *average treatment effect for the treated*. Next, we explore three approaches to compare both groups: a comparison of means, propensity score matching, and Mahalanobis-metric matching.

3.1 Comparison of means

If cases were to be assigned randomly to ‘treatment’ and to ‘control’, we could simply compute the average penalties imposed in both groups and calculate the difference to estimate the average treatment effect. However, it is likely that there is a selection bias in treatment assignment. The reason for the selection bias is the role of the public prosecutor the key player who decides which cases require a criminal procedure and which cases can potentially be left to the administrative enforcement (as explained in section 2.2). It is likely that, on average, more serious offenses are taken to court and that selection of the two groups is non-random. The results of comparing the average penalty in both groups are presented in Table II. We see that the average monetary penalty is thus roughly €12000 higher in criminal than in administrative enforcement, if we look at the full amount. If we only take into account the part of the penalty that is effectively imposed, we find a difference of some €9500. Both numbers are statistically significant at the 95% level.

Table II: Comparison of mean penalties in different treatment groups

| | | Number of obs. (n) | Mean (euro) | Standard error |
|---|----------------------|--------------------|----------------|----------------|
| Full fine levels (i.e. effective and suspended fine) | Criminal track | 221 | 12607.6 | 3064.9 |
| | Administrative track | 255 | 654.9 | 52.6 |
| | Combined | 476 | 6204.4 | 1447.6 |
| | Difference | | 11952.7 | 2853.4 |

| | | | | |
|------------------------------|----------------------|-----|---------------|--------|
| Effective fine levels | Criminal track | 221 | 10185.3 | 2876.8 |
| | Administrative track | 255 | 654.9 | 52.6 |
| | Combined | 476 | 5079.7 | 1352 |
| | Difference | | 9530.4 | 2678.3 |

3.2 Propensity score matching

Given the emphasis of this paper on accurately measuring the treatment effect, we focus on matching techniques through which we can explicitly evaluate the reduction in the sample selection effect. Propensity score matching allows us to match cases of different treatment groups, based on similarity with respect to a set of observable characteristics, and then measure the treatment effect for similar cases by calculating the difference in average treatment for these similar cases. This technique also allows us to restrict the used observations of our dataset to the common support region (of treatment and control group) and allow for estimation of non-linear relationships.

The idea of propensity score matching is to approximate a randomized control experiment using observational data. The identifying assumptions are the following. Firstly, the conditional independence assumption states that we can find a list of explanatory variables X such that, conditional on the set X , the cases included in both treatment and control group only differ in some random component that does not affect the outcome variable Y . Let D denote the treatment status ($D = 1$ if administrative enforcement) and define Y_0 as the counterfactual outcome, i.e. the outcome if the case would remain untreated (criminal enforcement). The conditional independence assumption is then formally:

$$E(Y_0 | X, D = 1) = E(Y_0 | X, D = 0).$$

Secondly, the common support assumption states that there should be a certain region (defined on the space of the explanatory variables X) with common support. Formally:

$$P(D = 1 | X) < 1.$$

Matching methods may suffer from the ‘curse of dimensionality’ problem. This problem arises when the number of explanatory variables X on which the matching is based becomes increasingly large. A propensity score (Rosenbaum & Rubin, 1983) can solve this problem by acting as a balancing score to assess similarity between different cases. The definition of a balancing score is that, conditional on this score, treatment is uncorrelated with all the explanatory variables. Formally:

$$X \perp D | p(X)$$

The propensity score can be estimated by regressing the treatment probability ($D = 1$) on the set of explanatory variables X . Intuitively, we first estimate the factors determining treatment assignment. Then, the estimated propensity score $p(X)$ acts as an aggregate indicator of the variables X ⁷.

Smith & Todd (2005) provide an overview of the conditions under which propensity score matching estimates generally provide accurate and unbiased estimation coefficients in the context of evaluation of labor market programs. They state that the estimates based on propensity score matching are sensitive to the variables included in the score model as well as to the sample used. They find that

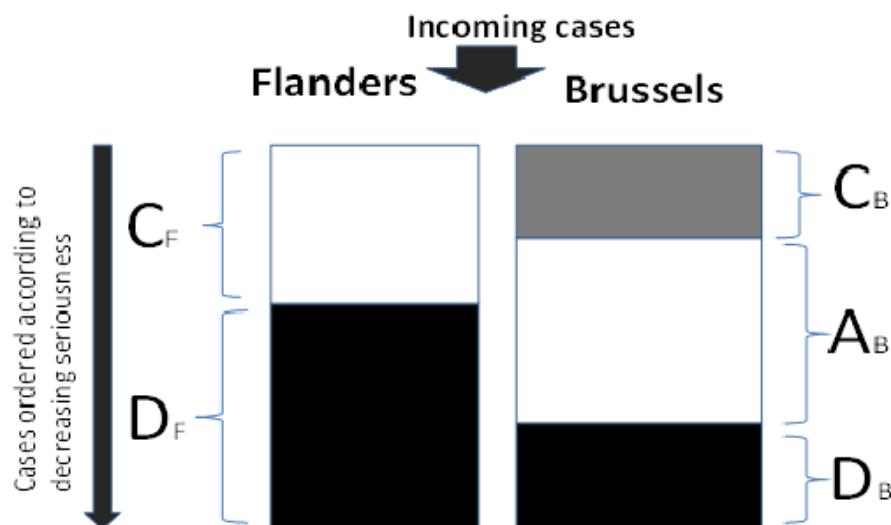
⁷ We check whether this aggregation in one measure $p(X)$ results in a balanced set for all the covariates X .

accurateness of propensity score matching is not universal, but should be assessed in each particular context. The most important criteria for application of propensity score matching are the following:

- A rich set of covariates should be available to explain the selection decision and these should be included in the selection model.
- There should be significant overlap between the treatment and the control observations in the dataset used.
- The control group should be, if possible, drawn from the same geographical location as the treatment group. If this is not possible, preferably a difference-in-difference matching estimator should be used to reduce geographic mismatch as much as possible.
- Dependent variables should be measured in the same way for participants and non-participants (this is the equivalent of the SUTVA, or stable unit treatment value, assumption).

Firstly, we have a relatively rich set of covariates to implement propensity score matching. While in empirical legal analysis data availability is generally quite limited, we can use micro-level data on the individual cases.

Figure I: Overview of enforcement decisions in Brussels and Flanders (C = Criminal enforcement, A = Administrative enforcement, D = opportunity dismissal/transaction offer, F = Flanders, B = Brussels)



Secondly, we expect that there is significant overlap between both groups in terms of severity of the offense, which is caused by the difference in timing in the introduction of administrative enforcement competences. While these were introduced in the Brussels-Capital Region in 1999, their introduction in the Flemish Region occurred only in 2009. At the time during which we collected our data, in 2003-2006, there is thus scope for significant overlap between the Brussels administrative cases and the Flemish criminal cases (see Figure I). First, the cases are entering a public prosecutor's office. In Figure I, we assume an ordering of violations in terms of seriousness. The prosecutor then decides whether to proceed by using criminal prosecution (=C), using administrative prosecution (=A), or by awarding an opportunity dismissal or transaction offer (=D). The subscripts indicate the region. We can thus expect some overlap between seriousness of criminal cases in Flanders and administrative cases in Brussels in the time span of our dataset (2003-2006). A matching test allows us to check whether such overlap can

indeed be found. We exploit this overlap to learn about treatment differences. Specifically, we obtain data for criminal cases in Flanders and administrative cases in Brussels (the zones represented in white in Figure I), match similar cases and compare their treatments in the two enforcement paths.

Thirdly, the condition that treatment and control groups should be drawn from the same geographical area, if possible, is not fulfilled in our context. Ideally, we would use a difference-in-difference matching estimator to address the fact that enforcement processes are situated in different regions, but this is not possible given the cross-sectional nature of our dataset. Thus, it is possible that part of the treatment result that we obtain is driven by a regional-specific effect, rather than by the treatment effect itself. Therefore, we must assess whether there might be any regional-specific characteristics which can potentially impair the results of our analysis. Firstly, the environmental legislations which we consider in our dataset are highly similar in Flanders and Brussels. The foundation of environmental law in the different regions in Belgium originates from one common legislation which was uniform up to the regionalization of the environmental competences in 1980. Further, most recent legislation originates from European Union legislation and is therefore also very similar. Regarding enforcement legislation and practice, we note that the criminal code and criminal prosecution code are the same for the whole country. With respect to administrative enforcement, regional legislators tend to look at each other and base legislation on a similar model. Therefore we can conclude that ex-ante there is no immediate reason to suspect that regional-specific effects will affect the results of our analysis⁸.

Fourthly, the final condition concerning the measurement units of the dependent variable, is fulfilled if we focus only on the monetary component of the penalties imposed. This is somewhat of a restriction because the judge in a criminal court has a richer set of sanctioning tools (imprisonment and additional fines of individuals that are active at accused firms, closure of firms, forfeiture of illegal gains) at his disposal than his counterpart, the administrative penalization official. Thus we are only making a partial comparison of relative deterrence by comparing fines between both enforcement tracks. However, we should not overrate this restriction for a number of reasons. Firstly, deterrence obtained through the threat of prison sentences in criminal enforcement is likely to be quite low in Belgium since sentences of at most three years are not executed due to lack of space in prisons (Billiet & Rousseau, 2014). Secondly, while closure of a firm is an additional sanctioning tool at the judge's disposal, we observe that this measure is often temporary and mainly serves as a remediation mechanism rather than as a deterrence mechanism. In addition, this penalty is imposed in only 5% of all cases (Billiet et al, 2009). The last additional type of penalty available in criminal and not in administrative enforcement is the forfeiture of illegal gains. If this type of penalty is imposed, it tends to be significant in monetary terms. However, its application is limited to 5% of all cases (Billiet et al. 2009)⁹. Compare this to the 95% share of cases in which a fine is imposed as (part of) the penalty and we can conclude that our focus on a comparison of fines cannot be interpreted as overly restrictive.

In order to implement this matching technique, we first estimate an auxiliary probit model to assess which case characteristics are generally considered as 'gravity indicators' by judges and prosecutors. If we can control for these 'gravity indicators', and compare cases which are considered as similar (to a certain extent) by judges, we can get an accurate measurement of treatment differences. In the auxiliary probit model we estimate which case characteristics lead to higher prevalence of imprisonment

⁸ We can correct for a possible difference in relevance of a particular legislation/type of offense If we include this information in the propensity score indicator. In addition, Billiet et al. (2009) found that regional disparities in terms of the offended environmental regulations are not as large as one may intuitively expect. They observed that breaches of the environmental permit decrees and the waste decrees are quite common in both regions. Besides these, a large share of cases concern noise nuisance in both enforcement tracks.

⁹ Also, alternative penalties such as community service are applied in fewer than 2% of all cases.

sentences. These characteristics can thus be considered as ‘gravity indicators’. We proceed by estimating another auxiliary probit model to construct the propensity score indicator. This can be done through a simple probit model in which the treatment indicator is the dependent variable. Formally:

$$P(Adm|X) = \Phi(\hat{\beta}_0 + X\hat{\beta} + \varepsilon)$$

The matching can be one-dimensional; matching will then be done according to propensity score values $p(X)$. Or, as we explain in the next section, it can be multi-dimensional, using the propensity score $p(X)$ along a number of other variables X and minimizing the statistical distance.

3.3 Mahalanobis-metric matching

While matching cases based on propensity score values allows reducing the dimensionality problem by aggregating all information into one number, it is also possible to match cases using several characteristics on multiple dimensions. We then include a number of variables that can lead to selection bias into a vector X and use the Mahalanobis distance to assess similarity between different cases. It is well-known that Mahalanobis-metric matching is particularly well suited for matching datasets which include many dummy variables, as is the case in our dataset. Formally, the statistical (=Mahalanobis) distance can be expressed by, with V representing the variance-covariance matrix of (W_i, W_j) :

$$d(i, j) = (W_i - W_j)V^{-1}(W_i - W_j)'$$

We implement the matching procedure using the `psmatch2`-package in Stata which has been developed by Leuven & Sianesi (2003).

Finally, we will use a balancing test to assess the quality of the matching procedure. Basically, in this test we compare the average values for each of the explanatory variables (included in X) in both treatment groups. If the balancing test shows that all the explanatory variables are balanced between treatment groups, we can calculate the average treatment difference (for similar cases) in our matched sample.

4 Variables and expected results

First we define the dependent and explanatory variables that we use in our estimations. Next, we outline the results that we expect to obtain.

4.1 Dependent variables

Our main dependent variables represent the size of the monetary penalty that is imposed to guilty firms in administrative and criminal enforcement and will be used to evaluate treatment effects. Because part of this penalty can be imposed conditionally, we specify two different dependent variables for measuring the treatment effect in criminal enforcement. The variable `AMOUNTALL` represents the entire fine for criminal enforcement, while the variable `AMOUNTEFF` represents only the part of the fine that is effectively imposed (the unconditional part of the fine). Since it is not possible to impose a conditional administrative fine, both variables represent the entire penalty for the administrative cases.

We include the treatment indicator `ADMIN` (=1 for administrative cases; =0 for criminal cases) as the dependent variable in the probit regression to estimate the propensity score values. Finally, we use an indicator for the imposition of prison sentences `PRISON` (=1 if a prison sentence was imposed to a natural person in the case for which the firm is being judged) to evaluate which case characteristics can be considered as ‘gravity indicators’ (by a judge or public prosecutors).

4.2 Explanatory variables

The explanatory variables consist of a number of environmental enforcement offense characteristics and a number of offender characteristics. First we define some offense characteristics. The variable RISK is equal to 1 if the violation did not result in any environmental damage (yet), but led to significant environmental hazards and potential damages. When the offense did result in environmental damages, RISK is equal to 0. A related variable is DAMAGE, which equals 1 if the punishment motivation accompanying the verdict demonstrates that the judge/fining official considered this offense to be particularly damaging. HEALTH is equal to 1 if the violation had a significant impact on public health. The variables WASTE, NOISE, SOILWATER and ODORAIR indicate the offense's type of contamination. Furthermore, COMPLAINT is equal to 1 if the offense was detected after the filing of a complaint by the surrounding community. Finally, PERMIT DUTY and PERMIT COND are equal to 1 for specific offenses against the obligation to file for an environmental permit or against the conditions of this permit. PERMIT DUTY violations reflect the situation where a firm does not have a (valid) environmental permit. Offenses such as not supplying data or not filling out forms are not included in this category, but are often categorized as violations of permit conditions. While the lack of a (valid) environmental permit does not necessarily imply that large environmental damages are caused, legally the firm should not be active without a permit and thus all emissions associated with its activities are by definition illegal.

With respect to the characteristics of the offender, the dummy variable PREVCONV is equal to 1 if the offender had a negative compliance history in terms of previous convictions for environmental issues. The count variable OFFENSES indicates the number of infringements committed by the offender in the actual case. We also include the dummy POSITIVE which takes a value of one if the violator made significant efforts to reduce the environmental damages caused or to reduce the environmental hazards imposed. Another explanatory variable LNASSETS represents the firms' asset level, which is the log transformation of the firm assets by book value in the year in which the infringement occurred. This is thus a proxy variable for firm size. Finally, INTENT is equal to one if the judge or fining official has a negative perception of the violator's intent. This variable is also deduced from the judge's respectively from the fining official's motivation, which accompanies the verdict.

4.3 Expected results

We first discuss the gravity indicators, i.e. which case characteristics should be considered as indicators for more or less serious environmental offenses? We expect that the severity of an offense increases with the variables DAMAGE and HEALTH, because they indicate that a certain offense has been more damaging and that gravity decreases with RISK, since no actual damage occurred. We also know that violations of the PERMIT DUTY obligation are considered as more serious (Billiet et al., 2014). Furthermore, with respect to characteristics of the offender, we know that an offender with a negative compliance history (PREVCONV), who committed multiple OFFENSES, or who showed negative INTENT is normally considered a more serious offender. In contrast, if the offender has shown signs of POSITIVE actions, this is generally considered a reason to treat him less stringently.

Next, we discuss the expectations with respect to the average treatment difference between administrative and criminal prosecution. Based on the normative analyses of Ayres & Braithwaite (1995) and Bowles et al. (2008) we conclude that administrative penalization should normally be less stringent. Also, using the simple concept of the optimal fine for environmental liability, outlined in Polinsky & Shavell (2000), we can expect this to be the case. Formally, this optimal fine can be written as follows:

$$F^* = \frac{harm}{p} + k .$$

In this formula, p equals the inspection & sanctioning probability, while k represents the fixed costs of the prosecution procedure. Given that the fixed prosecution costs k are generally higher in criminal procedures and that the inspection & sanctioning probability p may be lower, we derive the expectation that penalization should be more stringent in criminal than in administrative enforcement (for similar offenses). We thus expect to find a negative treatment difference.

5 Results from the empirical analysis

We now discuss the results of estimating treatment difference by using propensity score matching and Mahalanobis-metric matching. We end with sensitivity analysis in which we include the acquittals.

5.1 Propensity score matching

Propensity score matching consists of three steps: first we identify indicators of the seriousness or gravity of the offense, next we estimate a propensity score, and finally we are able to estimate an average treatment effect.

5.1.1 Auxiliary regression to estimate the ‘gravity indicators’

When we implement propensity score matching, we first want to get an idea about which case characteristics can be considered as ‘gravity indicators’ and which not. It will be particularly important to do matching in such a way that our matched samples are balanced for these ‘gravity indicators’, as otherwise the measurement of treatment effects could be significantly biased. We implement an auxiliary regression in which we regress a set of explanatory variables X against the dependent dummy variable PRISON in order to identify the ‘gravity indicators’. The results are shown in Appendix A.

We find that the following case characteristics can be considered as proxy indicators for the gravity of a certain case and that therefore balancing of these variables should be ensured in order to have a good matched sample. A first highly significant variable is POSITIVE (measures have been taken to mitigate damages or regularize the situation, a positive attitude towards enforcement officials), which has a negative effect on the propensity to impose imprisonment. Secondly PERMIT DUTY (a firm has performed activities which are potentially environmentally damaging without obtaining the necessary permits) has a significant positive impact on probability of prison sentences. Other variables which were somewhat significant include: a negative compliance history (PREVCONV), an offense with a serious impact on public HEALTH, a violation related to illegal WASTE treatment and a case that originated from a COMPLAINT (and has thus led to a certain tangible damage or discomfort). For all these variables, we check whether they are equally represented in both treatment groups during the balancing test.

5.1.2 Propensity score estimation

We proceed by estimating the propensity scores. These can be calculated by taking the fitted values of the probit model where the dependent variable is the indicator ADMIN. This model is presented in Appendix B.

5.1.3 Estimation of average treatment effect

We use a number of different matching specifications to estimate the average treatment effect (ATE). The results are shown in Table VI and Table VII.

Table III: Average Treatment Effect for AMOUNTALL based on Propensity Score Matching

| | | <i>ATE</i> | <i>T-stat</i> | <i>n</i> | <i>Mean bias</i> | <i>Median bias</i> | <i>Balanced (chi²-test)</i> |
|-----------------------------------|------------------------|---------------|---------------|------------|------------------|--------------------|--|
| Means comparison | | -11952 | -4.19 | 476 | 30.1 | 23.9 | No |
| One-to-one matching | No common support | -12104 | -0.78 | 476 | 21.2 | 21 | No |
| | Common support | -13388 | -0.92 | 454 | 21.8 | 18.5 | No |
| | With caliper (0.005) | -14987 | -1.48 | 373 | 10.7 | 8.5 | +/- |
| Smoothed matching | Nearest neighbours (5) | -11579 | -1.38 | 454 | 12 | 9.3 | No |
| | Radius (0.05) | -9748 | -1.2 | 476 | 13.7 | 11 | No |
| Weighted smoothed matching | Kernel (epanechnikov) | -11218 | -1.48 | 454 | 15.8 | 12 | No |

Table IV: Average Treatment Effect for AMOUNTEFF based on Propensity Score Matching

| | | <i>ATE</i> | <i>T-stat</i> | <i>n</i> | <i>Mean bias</i> | <i>Median bias</i> | <i>Balanced (chi²-test)</i> |
|-----------------------------------|------------------------|--------------|---------------|------------|------------------|--------------------|--|
| Means comparison | | -9530 | -3.56 | 476 | 30.1 | 23.9 | No |
| One-to-one matching | No common support | -11316 | -0.76 | 476 | 21.2 | 21 | No |
| | Common support | -12525 | -0.89 | 454 | 21.8 | 18.5 | No |
| | With caliper (0.005) | -13873 | -1.42 | 373 | 10.7 | 8.5 | +/- |
| Smoothed matching | Nearest neighbours (5) | -10390 | -1.31 | 454 | 12 | 9.3 | No |
| | Radius (0.05) | -8757 | -1.15 | 476 | 13.7 | 11 | No |
| Weighted smoothed matching | Kernel (epanechnikov) | -10134 | -1.42 | 454 | 15.8 | 12 | No |

For each estimation we provide a number of key statistics: we show the statistical t-value (which is significant at 5% level from a value of 1.64), the number of observations on which this estimation is based and the average and mean biases. These last two numbers are the average numbers of the percentage biases for each of the explanatory variables included in \bar{X} . Finally, we also mention whether the balancing procedure passes a joint balancing test based on a Chi²-distribution.

In general, we see that the propensity score matching procedure does not work very well in our context. The joint balancing test is almost always rejected which means that we did not succeed in balancing all the key explanatory variables. When we compare one-to-one matching with the means comparison, we do not only see that the balancing test is rejected, but also that the bias reduction is generally quite limited. It is only by imposing a caliper (a restriction on the allowable distance between a treatment observation and its control variable) that we succeed in reducing the bias somewhat. Even though the caliper model succeeds in the joint balancing test, there are still a number of individual variables that do not pass the balancing test. Moreover, we also find that the treatment gap is increasing in comparison to a means comparison (see also section 3.1). This is not very encouraging, because we would expect exactly the opposite: that the treatment gap should decrease once we control for selection bias.

In the smoothed matching techniques, we are no longer comparing one treatment observation to its single nearest control observation (based on propensity scores), but we are rather comparing it to some weighted average of the observations that are in its surroundings. In the simple smoothed matching the weights are uniform, whereas the weights are inversely proportionate to distance in case of a kernel.

When we look to these techniques, we do find that the treatment gap is somewhat decreasing (at least for the variable AMOUNTALL). This is encouraging as this is what we would expect to occur when controlling for selection bias. The bias reduction is also considerably higher than in case of one-to-one matching. However, the balancing test is still not passed for the smoothed matching techniques. Thus, we are still uncomfortable to start interpreting the estimation results.

5.2 Mahalanobis metric matching

Given the disappointing results of propensity score matching, we move on to the multi-dimensional matching technique of Mahalanobis matching. We show the results in Table VIII and Table IX.

We include the estimated propensity score value $p(X)$ along with the explanatory variables X into a vector to implement Mahalanobis matching. A treatment observation is then coupled to its closest control observation so as to minimize the Mahalanobis statistical distance. Mahalanobis matching is known to provide good matching results in case of dealing with many dummy variables. However, our matched dataset still does not pass the balancing test since there are still a number of explanatory variables unbalanced, among which the variable PERMIT DUTY. The selection bias is significantly reduced, however. As expected, we find that the average treatment effect decreases considerably.

Table V: Average treatment effect for AMOUNTALL using Mahalanobis matching

| | | <i>ATE</i> | <i>T-stat</i> | <i>n</i> | <i>Mean bias</i> | <i>Median bias</i> | <i>Balanced (chi²-test)</i> |
|-------------------------|-------------------------|---------------|---------------|------------|------------------|--------------------|--|
| Means comparison | | -11952 | -4.19 | 476 | 30.1 | 23.9 | No |
| Mahalanobis | One-to-one | -5028 | -1.57 | 476 | 11.9 | 8.6 | No |
| | Kernel | -1580 | -0.79 | 249 | 0.3 | 0 | Yes |
| | Kernel - bw(0.3) | -2588 | -1.4 | 273 | 0.1 | 0 | Yes |

Table VI: Average treatment effect for AMOUNTEFF using Mahalanobis matching

| | | <i>ATE</i> | <i>T-stat</i> | <i>n</i> | <i>Mean bias</i> | <i>Median bias</i> | <i>Balanced (chi²-test)</i> |
|-------------------------|-------------------------|--------------|---------------|------------|------------------|--------------------|--|
| Means comparison | | -9530 | -3.56 | 476 | 30.1 | 23.9 | No |
| Mahalanobis | One-to-one | -3302 | -1.75 | 476 | 11.9 | 8.6 | No |
| | Kernel | -1135 | -0.56 | 249 | 0.3 | 0 | Yes |
| | Kernel - bw(0.3) | -1783 | -0.95 | 273 | 0.1 | 0 | Yes |

Once we combine Mahalanobis matching with weighted smoothed matching through use of a kernel, we finally succeed in passing the balancing test. In fact, we obtain the result that we have reduced the selection bias up to almost 100%. Our matched samples are so similar that mean bias has dropped to 0.1% (down from 30.1% in case of unmatched samples). When we look at the average treatment effect, we find that the treatment gap has narrowed significantly. Thus, taking into account selection bias makes the treatment difference go down. This is exactly as we expected. In the last specification, we loosen the bandwidth somewhat (= we allow a larger distance between treatment and control observations) up from 0.06 to 0.3, in order to increase the number of observations (n) in our matched sample. We then find an estimate for the treatment difference between administrative and criminal

enforcement of €2588 when we look at the entire monetary penalty imposed in court and of €1783 when we only look at the effective component of the monetary penalty.

We implement a bootstrap procedure to calculate standard errors for our preferred specification, namely the sample that was matched using Mahalanobis matching with a kernel of bandwidth 0.3. The results are shown in Table VII.

Table VII: Average treatment difference for AMOUNTALL and AMOUNTEFF using Bootstrap standard errors

| | <i>Coef.</i> | <i>Std. Err.</i> | <i>z</i> | <i>P>z</i> | <i>[95% Conf. Interval]</i> | |
|-----------------------|--------------|------------------|----------|---------------|-----------------------------|------|
| ATE(AMOUNTALL) | -2588.4 | 1247.2 | -2.08 | 0.038 | -5033 | -143 |
| ATE(AMOUNTEFF) | -1783.8 | 1229.4 | -1.45 | 0.147 | -4193 | 625 |

We find that the average treatment difference (ATE) is statistically significant if we include the full amount of the monetary fine. If we only take the effective component of the penalty into account, then we find that the treatment difference is not statistically significant. In any case, there is some evidence that criminal judges are more severe for similar cases than administrative fining officials. However, the difference is not very large. The apparent treatment difference when looking at sample means seems to be attributable to the largest extent to a selection bias effect, and only to a small extent to an actual treatment difference.

5.3 Sensitivity analysis with respect to inclusion of acquitted defendants

Finally we implement a small sensitivity analysis. We (re-)include the acquittals in our dataset to find out how the different enforcement paths compare in terms of ex-ante deterrence, i.e. before the decision is made of whether enough ‘burden of proof’ is present to convict a defendant. We make the comparison with the results that we obtained when we discarded acquittals. We only do this comparison using one specification, namely the preferred Mahalanobis matching (with kernel) result. The results of this comparison are shown in Table VIII.

We find that treatment effects are smaller when we include the acquittals in our dataset. This confirms our expectations, as the burden of proof is generally considered to be higher in criminal enforcement than in administrative enforcement. However, the difference with the earlier results (without acquittals) is not very large. This indicates that the difference in burden of proof in the administrative and criminal enforcement practice does not seem to be very important in our legal context.

Table VIII: Comparison of treatment differences taking acquittals (not) into account

| | <i>Coef.</i> | <i>Std. Err.</i> | <i>z</i> | <i>P>z</i> | <i>[95% Conf. Interval]</i> | |
|--|--------------|------------------|----------|---------------|-----------------------------|------|
| ATE(AMOUNTALL) - without acquittals | -2588.4 | 1247.2 | -2.08 | 0.038 | -5033 | -143 |
| ATE(AMOUNTEFF) - without acquittals | -1783.8 | 1229.4 | -1.45 | 0.147 | -4193 | 625 |
| ATE(AMOUNTALL) - with acquittals | -2101.5 | 1096.5 | -1.92 | 0.055 | -4250 | 47 |
| ATE(AMOUNTEFF) - with acquittals | -1464.0 | 1045.9 | -1.4 | 0.162 | -3514 | 586 |

5.4 Discussion of results

From the empirical analysis we find that there are some treatment differences between administrative and criminal enforcement of similar environmental offenses: a similar type of offense leads to lower penalties under administrative treatment than in criminal enforcement. From this we conclude that in terms of providing deterrence against regulation offenders, the criminal track still seems to be a more powerful instrument in practice. However, the difference that we found is not very large and it also turns out that the largest part of apparent treatment differences can be attributed to a selection bias effect. Still, in this analysis we have focused on the fines and thus abstracted from additional criminal deterrence mechanisms such as the stigma effect or the possibility of prison sentences. Therefore, the relative treatment difference which we estimated should be interpreted as a lower bound for the relative difference between these enforcement paths. One should also be aware that our evidence can be context-specific and it is currently unclear to what extent it carries over to other jurisdictional areas.

Related to our objective to compare a combined administrative-criminal enforcement path with the sole use of criminal enforcement, we can conclude that the choice between a combined enforcement path and an enforcement system using criminal procedures only should be seen as a trade-off in terms of deterrence. On the one hand, it is quite likely that prosecution and penalization will increase in a combined system. On the other hand, the expected penalty may decrease, because part of the offenses will be going to administrative rather than criminal enforcement; leading to significantly lower expected penalties for similar offenses. It follows that the effect on the expected costs of committing a violation ($= pF$) is unclear ex-ante.

One has to be aware, however, that the simple multiplication of pF assumes that offenders are risk-neutral. As discussed by Becker (1968), risk-averse offenders will be more deterred by a higher penalty F , whereas risk-loving violators can be deterred more effectively by increasing penalty probability p ¹⁰. In the light of this observation, we are inclined to conclude that a combined enforcement system may still be more effective in providing deterrence (against risk-loving offenders) than the single use of criminal enforcement.

On a more general level, we think that matching techniques are an interesting and feasible approach for analyzing treatment effects in different enforcement systems, while correcting for selection bias. This is an issue that may prove to be relevant in many areas of law & economics, some of which may be very different from our context of environmental regulation enforcement.

6 Conclusion

Our paper aims to make two main contributions: one on a specific level and one on a more general level. On a specific level, we measure the treatment difference between administrative and criminal enforcement of environmental violations, for similar offenses. We find that criminal judge impose tougher monetary penalties for similar offenses, but that the difference is not very large (+/- €2000). Most of the apparent treatment difference is attributable to a selection bias effect. Only a small share turns out to be an actual treatment difference for similar offenses.

From this, we can formulate some tentative policy advice. A move from a criminal enforcement system towards a combined criminal-administrative system is not necessarily an improvement and should be carefully considered since two counteracting effects are likely. Firstly, additional deterrence can be

¹⁰ The observation that a higher p is more effective in providing deterrence, then follows from the fact that most law violators are risk-loving individuals.

expected from the combined system since the probability of imposing a sanction after an offense was detected is likely to increase. Secondly, if fines for similar offenses are indeed lower in the administrative track, this additional deterrence effect may not be as large as one would initially expect. As a result, the combined effect on environmental compliance from a move towards a combined enforcement system is ambiguous and may be negative. The overall compliance rate could reduce if the deterrence effect of the decrease in fine levels outweigh the effect of an increase in sanction probability. This situation becomes even more likely, if we take the additional stigmatization effect of criminal sanctions into account. On the other hand, if regulatory offenders are risk-loving, and thus better deterred by increasing the sanction probability rather than by increasing sanction level, combined enforcement could still have positive effects on compliance. Overall, additional empirical research regarding the compliance effects of different enforcement systems is clearly desirable. So when considering a move towards a combined enforcement system, the policy maker should tread carefully. The impact on the probability of imposing a sanction on a guilty party as well as on the related probabilities of detecting an offense, of starting a sanctioning procedure and of finding a party guilty should be assessed. In addition, the effect on the stringency of sanctions should also be considered and it may be advisable to build in legal guarantees to protect the deterrence effect of environmental enforcement.

On a more general level, we explore an innovative method for evaluating legal treatment (differences) while taking into account selection bias. Given the prevalence of selection bias in these types of empirical analysis and given the difficulty to deal with it, we think that our approach is a significant contribution to come to better evaluation of legal (enforcement) programs. Application of this methodology can be relevant in many fields of empirical law & economics and should not be restricted to the area of environmental regulation enforcement. In short, the steps involved in this methodology are the following: 1) Construct a dataset documenting individual cases and represent this information in a set of dummy variables or (preferably) continuous variables; 2) Find an appropriate matching methodology to identify similar cases in such a way to pass a balancing test; and 3) Obtain treatment differences by calculating average treatment effect.

As an illustration, in Belgium recently a discussion emerged about whether there is a treatment difference for similar violations between different regions. The argument was made that in a certain region judges would be more lenient than in others. Subsequently, when looking at some case data, the observation was made that no conclusions could be drawn given the fact that the caseload that is compared need not necessarily be similar. This is a typical situation in empirical legal analysis. The methodology that we propose could provide a potential solution for this problem and allow for conclusions to be made on these questions.

Acknowledgements

We would like to thank Stef Proost, Timo Goeschl, Donatella Porrini, Tom Ulen and participants of PE-ETE seminars at KU Leuven as well as participants of the EALE conference in Hamburg for providing useful comments. We would also like to thank the participants of the project *'Environmental law enforcement: a comparison of practice between the administrative and the criminal track'* for collecting the data used in this paper. This project was funded by the Agency for Innovation by Science and Technology of the Flemish Government, 'Strategic Basic Research' program.

Note

The experimenting and investigation conducted for this article comply with the current Belgian laws. The authors declare that they have no conflict of interest.

References

- Ashenfelter, O., Eisenberg, T. & Schwab, S. (1995). Politics and the judiciary: the influence of judicial background on case outcomes. *Journal of Legal Studies* 24, 257-281.
- Ayres, I. & Braithwaite, J. (1995). *Responsive regulation: Transcending the deregulation debate*. Oxford University Press.
- Becker, G.S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy* 76, 169-217.
- Billiet, C.M. (2012). Satellite images as evidence for environmental crime in Europe. A judges perspective, in Purdy, R. and Leung, D.(ed.), *Evidence from Earth Observation Satellites. Emerging Legal Issues*. Leiden – Boston, Brill.
- Billiet, C.M., Blondiau, T. & S. Rousseau (2014). Punishing environmental crimes: An empirical study from lower courts to the court of appeal. *Regulation and Governance* 8(4), 472-496
- Billiet, C. & Rousseau, S. (2014). How real is the threat of imprisonment for environmental crime ? *European Journal of Law and Economics* 37(2), 183-198
- Billiet, C., Rousseau, S., Balcaen, A., Meeus, R., Styns, K., De Meyer, G., Vander Beken, T. & Lavrysen, L. (2009). Milieurechtshandhaving: een databestand voor onderzoek naar penale en bestuurlijke sanctioneringspraktijk. *Tijdschrift voor Milieurecht* 18 (2), 128-150.
- Blondiau, T. & Rousseau, S. (2010). The impact of the judicial objective function on the enforcement of environmental standards. *Journal of Regulatory Economics* 37(2), 196-214.
- Bowles, R., Faure, M. & Garoupa, N. (2000). Economic analysis of the removal of illegal gains. *International Review of Law and Economics* 20, 537-549.
- Bowles, R., Faure, M. & Garoupa, N. (2008). The scope of criminal law and criminal sanctions: an economic view and policy implications. *Journal of Law and Society* 35(3), 389-416.
- Cohen, M. (2000). Monitoring and enforcement of environmental policy. In: Tietenberg, T. and Folmer, H. (eds.), *International yearbook of environmental and resource economics*, Volume III, Edward Elgar Publishers.
- Cooter, R. & Ulen, T. (2007). *Law & Economics, Fifth Edition*. Pearson International Edition.
- Deng, X., Huang, J., Uchida, E., Rozelle, S. & Gibson, J. (2011). Pressure cookers or pressure valves: Do roads lead to deforestation in China? *Journal of Environmental Economics and Management* 61(1), 79-94.
- EPA (2005). Calculation of the economic benefit of noncompliance in EPA's civil penalty enforcement cases. *Federal Register* 70(165)
- European Commission (2008). Directive 2008/99/EC on the protection of the environment through criminal law. Available through: <http://ec.europa.eu/environment/legal/crime/index.htm>

- Ferraro, P., McIntosh, C. & Ospina, M. (2007). The effectiveness of the US endangered species act: An econometric analysis using matching methods. *Journal of Environmental Economics and Management* 54(3), 245-261.
- Funk, P. (2004). On the effective use of stigma as a crime deterrent. *European Economic Review* 48, 715-728.
- Galbiati, R. & Garoupa, N. (2007). Keeping stigma out of administrative law: An explanation of consistent beliefs. *Supreme Court Economic Review* 15, 273-283.
- Garoupa, N. (1997). The theory of optimal law enforcement. *Journal of Economic Surveys* 11, 267-295.
- Gray, W. & Shimshack, J. (2011). The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence. *Review of Environmental Economics and Policy* 5(1), 3-24.
- Hampton, P. (2005). *Reducing administrative burdens: effective inspection and enforcement*. Retrieved: <http://www.bis.gov.uk/files/file44593.pdf>
- Heckman, J., LaLonde, R. & Smith, J. (1999). The economics and econometrics of active labour market programmes, in Ashenfelter, O. & Card, D. (eds.), *The handbook of Labour Economics*, Volume III.
- Imbens, G. & Wooldridge, J. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature* 47, 5-86.
- Leuven, E. & Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing and covariate imbalance testing. <http://ideas.repec.org/c/boc/bocode/s432001.html>. Version 4.0.6 - 17may2012.
- Levitt, S. & Miles, T. (2007). Empirical study of criminal punishment. In: Polinsky, M. & Shavell, S. (Eds.) *Handbook of Law and Economics*, Elsevier.
- Macrory, R. (2006). *Regulatory justice: Making sanctions effective*. Retrieved: <http://www.bis.gov.uk/files/file44593.pdf>
- MHR - Milieuhandhavingrapport (2009). Vlaamse Overheid: Departement Leefmilieu, Natuur & Energie. Retrieved from: <http://www.lne.be/themas/handhaving/afdeling-milieu-inspectie/milieuhandhavingrapport>
- Milimet, D. & List, J. (2004). The case of the missing pollution haven hypothesis. *Journal of Regulatory Economics* 26(3), 239-262.
- OECD (2009). *Ensuring environmental compliance: Trends and good practices*. Paris: OECD Publishing.
- Polinsky, M. & Shavell, S. (2000). The economic theory of public law enforcement. *Journal of Economic Literature* 38, 45-67.
- Polinsky, M. & Shavell, S. (2007). The theory of public enforcement of law. In: Polinsky, M. & Shavell, S. *Handbook of Law and Economics*, Elsevier.
- Rasmusen, E. (1996). Stigma and self-fulfilling expectations of criminality. *Journal of Law & Economics* 39, 519-544.
- Rosenbaum, P. & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1), 41-55.
- Shavell, S. (1993). The optimal structure of law enforcement. *Journal of Law & Economics* 36(1), 255-287.

- Smith, J. & Todd, P. (2005). Does matching overcome Lalonde's critique of nonexperimental estimators? *Journal of Econometrics* 125, 305-353.
- Svatikova, K. (2011). Combined use of criminal and administrative fines in enforcing environmental violations. In: *PhD Thesis: Economic criteria for criminalization: Optimizing enforcement in case of environmental violations*, Chapter 6.
- Svatikova, K. (2012). Criminal or administrative law to protect the environment? Evidence from Western Europe. *Journal of Environmental Law* 24(2), 253-286.
- Telle, K. (2013). Monitoring and enforcement of environmental regulations. *Journal of Public Economics* 99, 24-34
- UK Sentencing Advisory Panel (2000). Environmental Offences: The Panel's Advice to the Court of Appeal. [Last accessed 8 November 2013.] Available from URL: <http://www.banksr.co.uk/images/Guidelines/Advisory%20Panel%20Reports/Environmental%20Offences.pdf>.
- US Sentencing Commission (2008). Federal Sentencing Guidelines Manual. Chapter 2, Part Q, US Sentencing Commission, Washington, DC.

Appendix A

Table A: Estimation of 'Gravity indicators' using a Probit model with prison sentence as dependent variable

| | Coefficient | p-value |
|------------------|-------------|---------|
| PREVCONV (d) | 0.205* | 0.056 |
| DAMAGE (d) | 0.00603 | 0.848 |
| RISK (d) | -0.0101 | 0.805 |
| INTENT (d) | -0.0249 | 0.49 |
| POSITIVE (d) | -0.126*** | 0.004 |
| HEALTH (d) | 0.122* | 0.067 |
| WASTE (d) | 0.0770* | 0.077 |
| SOILWATER (d) | 0.0185 | 0.665 |
| ODORAIR (d) | 0.0299 | 0.644 |
| COMPLAINT (d) | 0.109* | 0.074 |
| PERMIT DUTY (d) | 0.117*** | 0.005 |
| PERMIT COND (d) | -0.0165 | 0.608 |
| OFFENSES | 0.00458 | 0.401 |
| LNASSETS | 0.00490 | 0.472 |
| Observations | 221 | |
| Pseudo R^2 | 0.283 | |
| chi ² | 47.56 | |
| Pred_1 | 75 | |
| Pred_0 | 77.20 | |
| Pred_Total | 76.92 | |

Marginal effects; p -values in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix B

Table B: Estimation of Propensity Score Value using a Probit model with Administrative Enforcement as Dependent Variable

| | Coefficient | p-value |
|-------------------|-------------|---------|
| PREVCONV | 1.214*** | (0.000) |
| DAMAGE | -0.456** | (0.036) |
| RISK | 1.959*** | (0.000) |
| INTENT | -0.425** | (0.046) |
| POSITIVE | 0.284** | (0.044) |
| HEALTH | 0.669*** | (0.007) |
| WASTE | 0.622*** | (0.000) |
| SOILWATER | -0.0611 | (0.776) |
| ODORAIR | 0.630** | (0.011) |
| COMPLAINT | 1.232*** | (0.000) |
| PERMIT DUTY | -0.0712 | (0.632) |
| PERMIT CONDITIONS | -0.501*** | (0.002) |
| OFFENSES | 0.0585** | (0.042) |
| LNASSETS | -0.0756*** | (0.000) |
| Constant | -1.182*** | (0.000) |
| Observations | 513 | |
| Pseudo R^2 | 0.358 | |
| Chi ² | 254.7 | |

p-values in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$