View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Illinois Digital Environment for Access to Learning and Scholarship Repository

© 2012 by Gauthier Robert Louis Pitois. All rights reserved.

DETERMINANTS OF FACILITY LEVEL ENVIRONMENTAL INSPECTIONS

BY

GAUTHIER ROBERT LOUIS PITOIS

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Agricultural and Applied Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2012

Urbana, Illinois

Master's Committee:

Professor Nicholas Brozović, Chair
Professor Alex Winter-Nelson
Professor Hayri Önal

Abstract

As environmental regulatory agencies have limited resources to enforce compliance, they tend to optimize the efficiency of resource allocation by employing targeting strategies. This thesis investigates the scheduling of Clean Water Act inspections in Illinois and the extent to which these inspections are memoryless. Using facility level and local EPA agency level data, we test inspection strategies for common decision factors, such as environmental performance, and compare them across the different jurisdictions in Illinois.

Our analysis has several key results. First, at the facility level, a majority of inspections are memoryless, though they are targeted according to local jurisdiction parameters. Second, although some facilities are targeted for more frequent non-memoryless inspections, none of our environmental performance parameters seem to influence this sorting. Finally, different inspection types are implemented in different ways, suggesting that they serve distinct purposes in the regulatory process.

Acknowledgments

I would like to thank to my adviser, Dr. Nicholas Brozović, for his guidance throughout the last couple of years. I am grateful and appreciative of his support as well as his enthusiasm and have greatly benefited from his experience and boundless curiosity and creativity. My sincere thanks also go to my committee members, Dr. Hayri Önal and Dr. Alex Winter-Nelson, for helping me improve my research. I would also like to thank all my fellow graduate students in the Department of Agricultural and Consumer Economics. I appreciate your kindness and help and wish you all the best. In particular, I owe special thanks to Taro Mieno for sharing his extended knowledge and his pertinent ideas. Finally, I want to thanks my parents, Jean-Louis and Anabelle Pitois, my sister, Corentine, and my fiancée, Najwa Obeid. Your continuous support has always been precious.

Table of Contents

List of Tables

List of Figures

1

Introduction And Literature Review

In the United States, environmental regulations are generally decided at the federal level and the implementation of these regulations falls to federal and state agencies. State environmental agencies have limited resources and cannot monitor thoroughly all regulated industries, thus requiring regulators to adopt strategies to try to optimize the efficiency of their regulatory actions. Different forms of resource allocation strategies may be referred to as targeting. Targeting generally consists of defining how the components of regulatory action should be used. Generally, these components are site inspections and enforcement actions. Numerous empirical studies measure the efficiency of different kinds of targeting strategies (Laplante and Rilstone, 1996; Helland, 1998; Eckert, 2004). They report how regulators use past and current environmental performance information from regulated firms to allocate their limited resources. The key element of these studies is to investigate the relationships between violations of environmental standards, site inspections, enforcements and relevant specific characteristics. In this study, we address the following questions: what kind of targeting strategy is used by the Environmental Protection Agency (EPA) in Illinois to implement the regulations of the Clean Water Act (CWA), and what are its determinants?

Harrington (1988) is one of the first authors to introduce a theoretical framework for targeting strategies for environmental compliance. His model consists of sorting polluting facilities in two groups on the basis of their past environmental performance and allocating regulatory resources accordingly. The environmental agency applies a greater control on the 'bad' firms, thus increasing inspection probability and expected fines. An increasing compliance level can be observed from the 'bad' firms, however, the 'good' firms have a lower incentive to comply since the probability of having a violation detected decreases. Magat and Viscusi (1990) underline the importance of an appropriate enforcement strategy and intensity given a desired environmental standard and suggest that inspections should play a role in encouraging the self-reporting of violations.

Following early theoretical contributions on targeting, a number of empirical studies address two important aspects of environmental regulations. The first issue is the efficiency of targeting strategies in deterring violations. The second is the expected regulatory pressure that a plant faces given its polluting activity. Gray and Deily (1996) analyze the interaction between regulatory decisions and compliance decisions at the plant level for steel plants. They find enforcement increases compliance, but at the same time compliance behavior induces fewer enforcement decisions. Although inspections do not always appear to be effective at increasing compliance, the threat of an inspection and the potential detection of a violation associated with it have an effect on a firm's decisions. For example, Earnhart (2004) finds a significant impact of enforcement on the emission levels of wastewater treatment plants in Kansas. Conversely, he finds that inspections have no deterrence effect while the threat of future inspections or enforcement actions has a deterrence effect. Eckert (2004) shows that past warnings increase the probability of an inspection as well as deter future violations. Likewise, Magat and Viscusi (1990) and Helland (1998) find that targeting potential violators has a significant effect on self-reporting of violations but does not lead to significant deterrence effects on emission levels. Similarly, Laplante and Rilstone (1996) find that increased inspections induce improved self-reporting of emissions and also find that both inspections and the threat of inspections reduce pollution emissions. Nadeau (1997) distinguishes the extent and the duration of harm caused by emissions violations. He shows that the EPA effectively reduces violation time, notably by allowing for separate strategies between compliant and non-compliant firms. Shimshack and Ward (2008) suggest that the randomness and jointness[1] of pollutant emissions generate compliance from noncompliant

[1] Randomness refers to the uncertainty of the extent of pollutant emissions and jointness expresses the fact that a reduction in emissions of one pollutant is often linked to a reduction in emissions of another.

firms as well as over-compliance from already compliant firms. Their study of pollutant discharges of the pulp and paper industry points out how enforcement can generate greater welfare than expected. Another type of indirect effect of environmental regulation is due to the spatial location of firms. Eckert and Eckert (2010) use the geospatial dimension of inspections of petroleum storage sites in Manitoba to demonstrate that inspections can be spatially correlated and that a plant is less likely to violate when its neighbors have been recently found to violate.

As these studies suggest, the empirical literature has shown quite similar results regarding the effects of enforcement actions but contrasting results concerning the effects of regulatory inspections. One reason for differences in the observed effects of inspections may reside in differences in the role played by inspections. In his early paper, Harrington (1988) considers inspections as the only means to detect violations while in most of the studies cited above, violations are self-reported and inspections have rather a routine control role and a threatening role. In our context, violations are self-reported, and therefore, the role of inspections is primarily to ensure that facilities are truthfully reporting their emissions and possibly to detect other kinds of defaults uncovered by reporting requirements. Consequently to this change in the role of inspection activity, a change in inspection strategy is to be expected. In a study focused on inspections, Rousseau (2007) distinguishes three different types of inspection to analyze the strategy of the Flemish environmental agency towards the textile industry in Flanders (Belgium). In order to determine if the environmental agency uses targeting, she uses survival analysis to estimate the probability of having an inspection given the amount of time since the last inspection as a function of past compliance behavior, together with relevant characteristics of the firms. As the different types of inspection[2] are estimated separately, the results show that the factors influencing the different types of inspection are

In particular, Shimshack and Ward (2008) indicate that Biological Oxygen Demand (BOD) reductions have important implications for other pollutant levels.

[2]Rousseau (2007) distinguishes reactive, routine and project-related inspections and estimates the probability of inspection for each type to show that the types are treated differently by the inspection agency.

also different.

Finally, the literature identifies the decision level at which environmental investigation is decided, generally by distinguishing federal inspections from state inspections. While CWA regulations are decided at the federal level, implementation frequently falls at the state level and studies generally assume consistent decisions for regulatory actions within a state. For example, Earnhart (2004) differentiates federal and state inspections and enforcement threats and actual actions and shows that the threat of federal actions has a greater deterrence effect. He also analyzes interactions between federal and state inspections and how they follow each other. Federal inspections follow state inspections as EPA may collect its own evidence, and state inspections follow federal inspections as the state agency may revisit after federal inspections.

The aim of this thesis is to investigate empirically the EPA's regulatory activities under the Clean Water Act towards major[3] point source polluters in Illinois. We analyze regulatory inspection decisions from two different angles: at the facility level by investigating each facility's inspections history using a fine time scale (quarters) and at the jurisdiction level by comparing local agencies' inspection schemes.

There are three major contributions of this thesis. First, although the state EPA implements CWA, day to day operational decisions are taken by local offices. For example, in Illinois the state is divided into seven independent jurisdictions. To the best of our knowledge, no published study has considered this finer scale of decision making. Our results suggest that jurisdiction-level heterogeneity of both industry type and regulators' preferences do affect inspection decisions.

Second, both previous studies and the federal EPA's current recommendations emphasize the importance of targeting inspections based on previous performance. Our results show a weak relationship between inspection frequency and location of facilities on impaired water

[3]Water discharging facilities are divided into two categories: major and non-major, based on specific rating criteria. Major facilities discharge equal to or greater than one million gallons per day and are required to self-report their emissions.

bodies[4] and previous violations and enforcement actions. However, there is evidence that local agencies target according to other parameters such as travel time and industry type.

Third, EPA uses different types of inspections that differ in terms of complexity and time for both the inspected facilities and the inspector. Previous studies have not carefully distinguished between inspection types. Our results show that local jurisdictions use different inspection types in different ways that are consistent with their relative costliness.

The remainder of the thesis is laid out as follows. First we present an institutional background where we detail implementation of the CWA and practices of local EPA agencies. Then, we justify and develop the econometric methods used based on results from the inspection targeting literature. In particular, we describe econometric methods to analyze the extent to which inspections are memoryless (using a test for a Poisson process and probit regression) and the determinants of facility level inspection rates (using OLS). A data section explains the dataset and presents summary statistics. Following this, results are presented and the thesis concludes with a discussion of results and policy implications.

[4] A large number of major water dischargers in Illinois discharge in streams listed under the 303(d) list of impaired waters of the CWA.

2

Institutional Background

Enacted in 1972, the Clean Water Act (CWA) seeks to "restore and maintain the chemical, physical, and biological integrity of the nation's waters." To serve this goal, the CWA established the National Pollutant Discharge Elimination System (NPDES). Under the NPDES program, all facilities discharging pollutants from any point source[5] into the waters of the United States are required to obtain a permit issued by the U.S. EPA or by an authorized state agency. As a result, all municipal and industrial facilities are issued a permit with numerical limits for regulated chemicals and are required to self-report their emission levels monthly. Facilities that cannot meet the NPDES permit limits still have the option to obtain a permit along with a compliance schedule. Compliance schedules allow dischargers to proceed with the installation of needed abatement technology while continuing to operate in noncompliance. Failure to meet with compliance schedule deadlines results in a violation.

The Integrated Compliance Information System National Pollutant Discharge Elimination System (ICIS-NPDES) gathers data from a large number of states and territories in a common database scheme. Different types of violation are recorded in ICIS-NPDES. Noncompliance with an effluent limit or a failure to report effluent levels automatically results in a permit violation. Noncompliance with previously scheduled work or compliance levels is also recorded in the corresponding category.

Inspections and enforcement actions are the two common regulatory actions conducted by EPA. For our purpose, we are interested in inspections and more particularly in under-

[5]Point source discharges refer to facilities with an identified connection to water bodies. They typically comprise industrial facilities and Publicly Owned Treatment Works (POTW). On the other hand, non-point sources (e.g. agricultural fields) are exempt from the NPDES program.

standing how the regulator employs them. NPDES inspections are of several different types. Reconnaissance inspections With Sampling or withOut Sampling (RWS/ROS) involve a visual control of the permittee's installation including the proper use of abatement technology. Given the short amount of time and procedures needed for RWS and ROS inspections, they are considered the least costly inspection type. Comprehensive inspections involve a more stringent review of facilities' monitoring records, interviews of the personnel, and inspection of wastewater treatment processes. The principal types are Compliance Evaluation Inspections (CEI), Compliance Biomonitoring Inspections (CBI), Compliance Sampling Inspections (CSI), Performance Audit Inspections (PAI), Diagnostic Inspections (DI), and Toxic Sampling Inspections (XSI). Because of their extensive requirements, comprehensive inspections are relatively costly for industries as well as for regulators in terms of time and effort. The federal EPA currently requires that a comprehensive inspection is performed at least once every two fiscal years for each facility.[6]

The other type of regulatory actions are enforcement actions. Informal enforcements are commonly used when facilities are in violation and consist of a warning through a phone call or a letter. Formal enforcements are less frequent[7] and are only applied to a small proportion of facilities in violation. Formal enforcements are generally administrative compliance orders and a smaller proportion concerns penalty orders.

While environmental regulations are decided at the federal level through the CWA, EPA may authorize states, territories, or tribes to implement the NPDES program. In most of the country, EPA has delegated to states the implementation of the program. States then operate with little intervention from the federal EPA. As a consequence, there may be different strategies for enforcing compliance across states corresponding to different political contexts or different budget constraints. In Illinois, the state EPA performs most regulatory actions. As mentioned in the previous section, there is also an even more local level of decision

[6]Frequency goal set by the Office of Enforcement and Compliance Assurance as of 2008.

[7]In our dataset, there are 247 formal enforcements versus 648 informal enforcements.

making at the jurisdiction level. A jurisdiction generally features one or two agencies from which facilities are located within fairly short driving distances. Illinois is divided into seven non-overlapping jurisdictions (Figure 1).

Finally, as public wastewater treatment plants represent most of the major water dischargers in Illinois, it is relevant to provide a short description of this industry. Most facilities are old and need to be expanded or rehabilitated and the collection system is over 50 years old.[8] Many public wastewater agencies are financially constrained and have difficulties performing maintenance and upgrade work on their facilities. The EPA maintains relevant information about the state of wastewater plants as they vary in size and work load. In particular the 'critical list' indicates those plants that are close to reaching their maximum treatment capacity and cannot be connected to additional sewer systems without a careful investigation.[9] A common expansion practice for current plants is to install pretreatment equipment. Such installations may be integrated into a compliance schedule.

[8]A more detailed report is available at: http://www.infrastructurereportcard.org/node/182

[9]There is also a 'restricted list' concerning plants that have reached their maximum capacity, but none of the major facilities in Illinois is on the restrited list.

3

Methods

We are concerned with the scheduling and determinants of inspections by local regulatory agencies. The federal EPA clearly states that the goal of CWA inspections is to "ensure and document whether entities possessing NPDES permits are complying with their CWA obligations."[10] It also recommends state administrators to "direct resources towards the most important noncompliance and environmental problems." On the other hand, local agencies are budget constrained, must meet minimum frequencies for comprehensive inspections for all facilities, and may be concerned with potential strategic behavior from facilities. The interaction of these features of the local regulatory process guides our choice of econometric strategy.

As seen before, the timing of inspections can be a function of firms' environmental performance (Rousseau, 2007) and the threat of future inspections can deter the occurrence of violations as well as decrease their duration. In some cases, the regulator may lack relevant information about firms' environmental performance or face uncertainty regarding when non-compliance may occur.[11] Also, from facilities' point of view, uncertainty regarding the timing of the next inspection may give an incentive for staying constantly in compliance, even after a recent inspection.

We assume that a strategic firm is particularly interested in the probability of having an inspection at a given time $P(I_t)$. If inspection probabilities vary over time in a way that is predictable to the firm, the firm will have time varying incentives to adjust its effort. This

[10] Documents related to CWA inspections available, as of June 2012, at: http://cfpub.epa.gov/compliance/resources/policies/civil/cwa/

[11] In particular, heavy rainfall events may cause greater pollution of water bodies.

is a challenge for a budget constrained regulator that cannot inspect all facilities regularly. One solution for the regulator would be not to condition the inspection probability on the time since the last inspection:

$$P(I_t) = P(I_t|I_{t-1}) \tag{3.1}$$

Note that even if inspections are memoryless, as equation (3.1) suggests, regulators can still target facilities according to their environmental performance. In the case of noncompliant facilities, more frequent memoryless inspections could encourage a faster return to compliance. In the case of compliant facilities, less frequent memoryless inspections would encourage the maintenance of compliance even after a recent inspection. Then, the probability of inspection would be constant at each facility and as a consequence the regulatory pressure would be the same at any time t for that facility.

The property of memorylessness is equivalent to having inspection times generated by a Poisson process. In a theoretical study on the availability of maintained systems subject to random failures, Wortman and Klutke (1994) explain how a random inspection strategy, based on a Poisson process, can maintain a constant availability. This idea can be transferred to environmental inspections: when information is reduced or uncertain, the regulator could adopt a Poisson process to inspect facilities.[12]

For these reasons, we test if inspections are memoryless, which is equivalent to testing the hypothesis that the regulator does not choose any particular times to inspect each facility or that each facility is inspected randomly over time. Our analysis first tests if the history of inspections of each facility is compatible with a Poisson process. If the times of inspections are independent, we can consider that the regulator randomly chooses the time when a facility is inspected. Thus, this test allows us to sort between the facilities that are randomly inspected over time and those that are not. Note that this methodology differs from survival analysis and, to the best of our knowledge, has not been used in environmental economics.

[12] When inspections are not a response to already known violations, a Poisson process could be used to maximize the findings of inspections, particularly in the case of reconnaissance inspections.

A Poisson process refers to a series of events for which the time between two events has a parameterized exponential distribution (Poisson distribution). In order to test for a Poisson process, we choose to use the Conditional Chi-squared statistic described in equation (3.2) where X is a random variable, in our case the number of inspections per quarter. Under the null hypothesis described in equation (3.3) T_{CC} has a Chi-square distribution with $n-1$ degrees of freedom (i.e. $T_{CC} \sim \chi^2_{n-1;1-\alpha}$).

$$T_{CC} = \sum \frac{(X_i - \overline{X})^2}{\overline{X}} \tag{3.2}$$

$$\begin{cases} H_0 : X_i \sim Poiss(\lambda_i),\ \lambda_1 = ... = \lambda_n \\ H_a : X_i \sim Poiss(\lambda_i),\ \sum \left(\lambda_i - \bar{\lambda}\right)^2 > 0 \end{cases} \tag{3.3}$$

This methodology is described by Cochran (1954) as the test of variance for the Poisson distribution and has proven to be reasonably sensitive for this purpose. Under the null hypothesis, we cannot reject that the random variable X is drawn from a Poisson distribution of mean λ, and under the alternative hypothesis, we reject that the observations of X are issued from the same Poisson distribution. This procedure enables us to consider individually each facility in the sample and to test if each sequence of inspections is compatible with a Poisson process.

It is relevant to note that failing to reject the hypothesis that inspections are compatible with a Poisson process does not mean that inspections are totally random and that the regulator does not have a targeting strategy. Indeed, it implies, as mentioned before, that inspections are memoryless[13] and that, in our model, the number of quarterly inspections is exponentially distributed with parameter λ.[14] In other words, the regulator chooses the rate of inspections for each facility as a result of his targeting strategy. In the following part of this study, we are interested in investigating the determinants of the rate of inspections.

[13] Or, the time intervals between two inspections are independent random variables.

[14] The mean number of inspections per quarter is then λ.

We calculate the rate of inspections as the mean number of inspections per quarter. As this rate is a continuous variable, we choose to use an OLS model with the rate of inspection as the dependent variable in order to investigate what facility level characteristics are involved in the regulator's choice. With such a model, we seek to verify if our data are compatible with a targeting strategy of increased pressure towards noncompliant facilities (Magat and Viscusi, 1990).

For those facilities whose inspections are consistent with a Poisson process, we regress the rate of inspection r using facility level attributes. Equation (3.4) describes our OLS regression model where r_i is the rate of inspection for facility i and $r_i = \lambda_i$, RA_i is a vector representing counts of regulatory actions towards facility i over the period, T_i is the travel time faced by the regulator to reach facility i, I_i is a vector of binary variables indicating the industry type of facility i (wastewater treatment plant, electric services, plastic industry, others), and μ_i is the error term. Since we differentiate reconnaissance and comprehensive inspections, r_i is either the rate of reconnaissance inspections or the rate of comprehensive inspections.

$$r_i = \alpha + \beta RA_i + \gamma T_i + \delta I_i + \mu_i \tag{3.4}$$

Since environmental inspections and regulatory actions are parallel processes over time, decisions can be made by the regulator regarding both processes simultaneously. In order to avoid endogeneity issues, we estimate the rate of inspection using prior regulatory actions. Thus, variables in RA_i are lagged. Ultimately, this model permits estimation of factors influencing the inspectors' decisions about how much regulatory pressure should be applied on each plant. As we estimate reconnaissance and comprehensive inspections separately, we are able to reveal any difference in the factors influencing inspection frequency for the two types of control.

In our first modeling approach, we build on previous economic theory that supports the determination of groups within the industry (Harrington, 1988; Magat and Viscusi, 1990) and separates plants according to environmental performance. In our second modeling approach, we inspect how information available to the regulator may be used to produce such a sorting. In particular, prior environmental performance is taken into account as we test the importance of the most common chemicals involved in water quality evaluation in the regulator's decision. Our choice of variables includes cumulative violation or enforcement counts as encountered in Rousseau (2007) or Eckert (2004) and chemical-specific violations as in Laplante and Rilstone (1996) or Earnhart (2004), so that we are able to investigate the precision level or pollutant specific concerns of the regulator in his decision process. Notably, the current EPA recommendations for targeting insist on general noncompliance, but also emphasize site impairment (i.e. impaired receiving water bodies) that generally relates to pollution from wastewater treatment plants with specific pollutants (e.g. BOD, coliform). Because the levels of different pollutants are often linked, the regulator may focus only on a few of them believed to be the best indicators of plants' behavior. In order to measure if individual pollutant events make a difference, we estimate the probability for facilities to belong to the group that is not inspected randomly over time. If inspections do not occur randomly over time, we suppose that the regulator is reactive to certain parameters or events and adjusts his rate of inspection accordingly. Thus, we use a probit model containing lagged individual pollutant violation events and enforcement events.

$$\begin{aligned} P(Facility_i \in NotRandom) = & \beta_0 + \beta_1 FormalEnf_i + \beta_2 InformalEnf_i \\ & + \beta_3 BOD_i + \beta_4 Chlorine_i + \beta_5 Coliforms_i \\ & + \beta_6 TSS_i + \beta_7 Overdue_i + \nu_i \end{aligned} \tag{3.5}$$

Equation (3.5) describes the probit model discussed above. Regressors are composed of formal and informal enforcement counts, violation counts of three common pollutants: BOD, chlorine, and fecal coliform, and violation counts of overdue monitoring reports.

4

Data

The data used in this study are composed of environmental performance records and geospatial information. The environmental performance data are extracted from the EPA's Integrated Compliance Information System-NPDES (ICIS-NPDES). This database comprises records for all discharging facilities in the United States.[15] These facilities are divided into two categories: major and non-major facilities, according to specific ratings criteria developed by EPA or States. Major facilities include facilities discharging equal to or greater than one million gallons per day. Such facilities are required to self-report their emissions levels. Reporting is voluntary for non-major facilities and as a result emissions data are often missing. In this thesis, we use only major facility data for consistency. Even though non-major facilities are more numerous, emissions from major facilities constitute the largest proportion of emissions and are thus more important to the regulator.

This study focuses on the 275 major facilities located in Illinois, their environmental performance and the related inspection activity of the EPA during the period 2001-2009. As Figure 1 shows, the state of Illinois is divided into seven jurisdictions and each jurisdiction is locally supervised by an EPA office in charge of monitoring regulated facilities. The facilities are unevenly concentrated in space with five jurisdictions having a lower concentration of facilities and the other two having a higher concentration. This fact reflects the disparities between rural and urban areas. Indeed, the jurisdiction comprising the Chicago area concentrates a greater number of facilities compared to the other jurisdictions.

[15]The modernized ICIS-NPDES is used by a large proportion of U.S. states and territories, although the older Permit Compliance System (PCS) is still used in the rest of the country.

Our primary variable of interest represents the inspection activity of the regulator. Based on the date of all inspection events, we aggregate inspections by quarter and use the count of inspections per quarter. We consider inspection types separately so that inspection counts are grouped by type: reconnaissance and comprehensive. During the first part of the analysis, we test for a Poisson process in the occurrences of inspections over time. The result of that test is then represented in a binary variable used in the next stages of the study with a zero value when the series is compatible with a Poisson process and a value of one when it is not.

Secondly, a set of variables describes facilities' characteristics. Three binary variables indicate the type of industry: sewerage systems, electric services, and plastic materials plants. The other types of plants are not flagged in the analysis. In order to signal whether or not a plant is located in a sensitive area, we use a binary variable indicating whether the plant is discharging in a stream listed under the 303(d) list of impaired waters of the CWA.[16]

We also use dummy variables to account for the jurisdiction each facility belongs to. The last characteristic is the travel time from the local EPA agency to the facility. We favor travel time over distance in order to represent better the difference in travel cost between rural and urban areas for the regulator. Travel time is calculated using the Directions tool of Google Maps.

Finally, a set of variables denotes facilities' environmental performance over time. Facility level violations, enforcement actions, and penalty fines are reported per year. Violations are reported for the most important water quality pollutants such as chlorine, fecal coliform, nitrogen, BOD (5-day BOD at 20°C), TSS, and pH. Violations are also recorded when self-reports of emissions are overdue. Enforcement actions are represented by three variables to distinguish the count of informal enforcements, the count of formal enforcements, and the sum of penalties accompanying a formal enforcement. All these variables are aggregated by year so that we can use them as lagged values.

[16]The CWA requires water bodies not meeting water quality standards to be listed. Water bodies in the 303(d) list are prioritized for restoration and protected by Total Maximum Daily Load (TMDL) programs. Currently the U.S has 14,153 waters on the 303(d) list, including 1,057 in Illinois.

As Table 1 shows, most of the major facilities are wastewater treatment plants[17] (208/275), followed by electric services plants (30/275). We also note that more than 50% of the major facilities discharge into impaired waters. Given CWA requirements for impaired waters, we expect these facilities to receive more oversight from the regulator. It is not surprising that travel time is variable across facilities as facilities are distributed across jurisdictions (Figure 1). Only jurisdiction 3 presents two obvious industrial spatial clusters, though each of them comprises an EPA agency.

Between 2001 and 2009, the regulator performed 2,519 comprehensive inspections and 11,776 reconnaissance inspections. This activity was focused on wastewater facilities with respectively 2,025 and 9,849 inspections for comprehensive and reconnaissance. Note that there is a decreasing trend in the amount of inspections during our sample time frame with reconnaissance inspections strongly declining as Figure 2 shows. The data also reveal that the regulator increased water sampling during reconnaissance inspections.[18]

Table 1 also provides violation records as two separate types: overdue violations and limit exceedance violations. An overdue violation is recorded when a facility fails to transmit all its required monitoring values to EPA. An exceedance violation is recorded when a facility reports emission levels greater than the value specified by its NPDES permit; a percentage of exceedance is associated with the violation record. Wastewater treatment plants account for most of the exceedance violations with 5,652 violations. These violations are also on average greater in exceedance percentage than for other industries. This may be attributed to the uncertainty of weather events and the variability of water flows as well as the limited capacity of wastewater treatment plants, thus making the wastewater industry an important concern for the regulator. As opposed to other studies (Rousseau, 2007), violations are automatically generated and are not necessarily linked to inspections.

The unit of observation used to count inspections is a quarter of a year. We believe that

[17] Table 2 presents specific variables for this industry.

[18] Inspections records show that ROS are replaced by RWS over time.

using quarterly data is a realistic approach considering that inspections may not happen more than a few times a year for most facilities. Table 3 shows when violations occur across quarters. As expected, water pollution shows some seasonality effects for limit exceedance violations. This is particularly observable for chlorine and coliform with higher counts for quarters 3 and 4 (April to June and July to September) which correspond to periods when rainfall is heavier and when wastewater treatment plants may face greater loads. Values for coliform are significantly greater from April to September at the 0.001 level and values for chlorine are greater at least at the 0.1 level. Surprisingly, overdue violation counts increase over the calendar year and may be linked to different levels of administrative or financial constraints as the year progresses.

The last regulatory action reported in this study is enforcement actions. We observe that violations do not necessarily lead to enforcement actions, even informal ones, as the count of enforcements is less than that of violations. Perhaps the regulator considers that the issuance of a violation is a sufficient signal to the industry, or that if a facility receives an enforcement after a violation, it will not be enforced again for its following violations. We can also suppose that enforcements concern only the worst violations with high level of exceedance. Another explanation for the lower number of actions is they are very costly to the regulator. The same pattern is observed with financial penalties as formal enforcement actions are not necessarily accompanied by a penalty. Only 23 formal enforcements in the wastewater industry out of 247 resulted in a total $74 million in penalties. Penalties are generally issued after a court decision which could explain their scarcity. The large number in the righthand column of Table 1 ($201 million) is entirely due to a single petroleum plant.

Finally, Table 4 presents the data as used in our analysis. We use annual formal and informal enforcement counts as well as financial penalties from 2001 to 2004 to estimate regulatory actions from 2005 to 2009. This lag between variables is adopted in order to avoid endogeneity. Similarly, we use counts of violations for relevant pollutants from 2001 to 2004. Chlorine, fecal coliforms, and TSS related violations are provided as well as overdue violations.

5

Results

We begin by reporting the results for the test of variance for the Poisson distribution (Equation 3.2). Recall that the analysis is carried out using facility-level inspection data. For each jurisdiction, Table 5 presents the number of facilities for which the hypothesis that inspections are memoryless can be rejected (i.e. are significantly different from a Poisson process) over the period 2001 to 2009. For each type of inspection, the proportion of all facilities is given as well as the proportion among wastewater treatment plants. In the case of reconnaissance inspections, almost all inspection timings seem to be memoryless. In the case of comprehensive inspections the proportion of facilities for which the hypothesis that inspections are memoryless can be rejected is between 20% and 50%. This means that for this subset of facilities, the probability of a comprehensive inspection is conditioned on the time since the last inspection. Also, wastewater treatment plants do not seem to be treated differently as their proportions resemble the total proportions (Table 5). This trend suggests that as the cost of inspections increases, the regulator increases the number of facilities non-randomly inspected.

The most striking result is the proportion of non-memoryless reconnaissance inspections in jurisdiction 3 (81%), corresponding to the Peoria area in western Illinois. As mentioned before, this jurisdiction presents some distinct characteristics. It has two spatial clusters of facilities, each of them having an EPA agency near to them. We also observe the greatest decrease in the average number of inspections over time in this jurisdiction (see Table A.1). These two facts suggest that jurisdiction 3 has adopted different targeting strategies than seen in other jurisdictions.

Overall, we note the difference between the two types of inspections. At the facility level, reconnaissance inspections are randomly distributed over time. This type of inspection can serve as a way to maintain a threat of receiving an intervention (Earnhart, 2004). This practice also encourages self-reporting (Magat and Viscusi, 1990; Laplante and Rilstone, 1996; Helland, 1998) as a facility is unable to predict its next inspection.

On the other hand, scheduling comprehensive inspections may have different factors as nearly one out of three facilities has inspections that are not memoryless. Comprehensive inspections have different purposes than reconnaissance inspections. In particular, the state regulator is required to perform at least one comprehensive inspection every two years for each facility. Comprehensive inspections also play an important role in enforcing environmental compliance. Because they involve a rigorous scrutiny of facilities' equipment and records, they may reveal concealed violations. Moreover, given their higher cost, it is not surprising that the regulator employs a different scheduling strategy. It is interesting to note that having memoryless comprehensive inspections and having memoryless reconnaissance inspections are significantly independent. The Pearson Chi-squared test for dependence is equal to $\chi^2 = 5.30e-3$ with p value $= 0.94$, so we cannot reject the hypothesis that memoryless reconnaissance inspections over time and memoryless comprehensive inspections over time are independent,[19] which reinforces the idea that the two types have different purposes for the regulator.

Theoretical results have shown that memoryless inspections are desirable when the underlying failure rate is random (Wortman and Klutke, 1994). Next, we analyze whether or not the occurrences of violations reported by facilities are memoryless (Table 6). We observe that for most facilities, overdue violations and limit exceedance violations are not randomly distributed in time. This suggests that those events are either smoothly distributed over time or concentrated at certain dates. Since a lot of facilities (e.g. wastewater) may have under-dimensioned equipment and some major weather events may overload plants'

[19] See Appendix A for complete independence test results.

treatment capacity, both scenarios are plausible. Conversely, when looking at particular pollutant violations we can observe a variety of patterns (Table 6). Overall, chlorine and BOD_5 violations appear to be randomly distributed over time. Depending on jurisdiction, coliform violations are more or less randomly distributed, with higher rates of non-random patterns in jurisdiction 1, 3, and 7. This may be due to disparities in wastewater treatment equipment.[20] Concerning TSS, a consistent 30-40% of facilities have non-randomly occurring violations. Overall, violations do not appear to be randomly distributed over time, which is an advantage for the regulator if he wants to follow up on violations. However, if the regulator focuses on particular pollutants (e.g. BOD_5), his strategy should account for the random character of their occurrences.

It is interesting to note that, for the physical and chemical violation types, the highest rates of non-memoryless violations occur in jurisdiction 3. Because this pattern is seen for both wastewater and other facilities, this may be interpreted as consistent with strategic behavior in emissions violations by facilities.

Table 7 presents OLS estimations for our inspection rate model. The purpose of this model is to explain the rate of inspection for facilities that were identified as having memoryless inspections in the previous part. We use the methodology presented above to determine which facilities have a sequence of inspections consistent with a Poisson process over a three year period. Then we create several subsets of our dataset in order to extract facility-level observations. Finally, we calculate the quarterly average rate of inspection during the same three year period to obtain the response variable. The model is completed with lagged enforcement count variables, travel time in minutes, and industry-specific binary variables.

The results are presented for different periods of the response variable for robustness[21] and we observe some consistent outcomes. Again, we note that the two types of inspection exhibit important differences.

[20] Coliform violations concern only the wastewater treatment industry.

[21] See Appendix A for more detailed results.

We find heteroskedasticity for both reconnaissance rate estimations using the Breusch-Pagan test, with p values of 0.01 and 0.0004. Therefore, we adjust the standard errors with White's heteroskedasticity-consistent estimator. The other two estimations for comprehensive inspection rates do not present any sign of heteroskedasticity. Models are also tested for multicollinearity using the Variance Inflation Factor and show no evidence of multicollinearity (we observe VIF values no greater than 2.3).

The rate for reconnaissance inspections appears to be significantly positively correlated with lagged informal enforcements of a prior year, indicating that the regulator may be increasing effort towards those facilities that have recently been warned. We note that informal enforcements from 2003 are significant in both time settings, which suggests that the regulator is interested in events that occurred during 2003 rather than in events occurring with a specific lag. None of the formal enforcement variables are significant indicating that reconnaissance inspections may not be considered as follow-up visits after an administrative decision. Unsurprisingly, we note that inspection rate is negatively correlated with travel time. It is likely that for budget or time constraints, further distances discourage the regulator and there may be a trade-off between spending time driving and spending time inspecting as reconnaissance inspections are rather short in time. As the coefficient of the travel time variable is -0.004, we estimate that a facility located 100 minutes further away from the EPA agency has 1.6 less inspections per year compared to a facility located nearer to the agency. Note that previous studies using state level data have not been able to use travel time as a variable because the location of the inspectors was not defined. Concerning industry type, it appears that wastewater treatment plants are more subject to reconnaissance inspections. Many reasons may motivate the regulator to visit this industry more frequently, notably because it has a large pollution potential and its equipment requires careful attention and operation. Moreover this industry is particularly challenged by uncertainty due to weather events. Our results are consistent with the intuition given by Rousseau (2007) who finds that, for certain lags, violations significantly increase (or decrease) the probability of routine inspections.

Comprehensive inspection rates present some similarities concerning past enforcement records. Although in this analysis no coefficient for informal enforcement is significant, some lagged formal enforcements have weak positive significance. As opposed to reconnaissance inspections, comprehensive inspections may serve as follow-up visits after a formal enforcement is issued, requiring the regulator to inspect facilities more thoroughly. However, this significance of formal enforcement is not consistent across time settings (Table 7). Travel time and wastewater variables are not as significant compared to reconnaissance inspections with coefficients an order of magnitude less. Indeed, the mandatory character of comprehensive inspections may reduce the importance of travel cost, while wastewater treatment plants may not require increased comprehensive inspection effort.

We now consider differences between facilities with and without memoryless comprehensive inspections.[22] For comprehensive inspections during the period 2005 to 2009 the rate of inspection is significantly higher for non-memoryless inspections, with an average of 0.29 inspections per quarter, versus 0.21 memoryless inspections per quarter. The Welch t-test for the difference in means is $t = 3.00$ and p value = 3.65e-3. This trend is consistent for our different time periods.

The higher rate of comprehensive inspections for facilities with non-memoryless comprehensive inspections suggests that the regulator targets these facilities by allocating more of his resources towards them. Recall also that in all jurisdictions during 2001 to 2009, 20 to 50% of facilities have non-memoryless comprehensive inspections (Table 5). Next, we use shorter intervals to observe whether facilities enter and exit the non-memoryless group. We observe respectively 59, 48, and 49 non-memoryless patterns over the periods 2003-07, 2004-08, and 2005-09, with a turnover of 30.5% between the first two periods, 25% between the two last ones, and finally with 52.5% of the initial group remaining through all periods.

[22]Differences in rates of reconnaissance inspections are significant but almost all non-memoryless inspections are in jurisdiction 3, which has a low average reconnaissance inspection frequency.

This suggests that when a facility is in the non-memoryless group (i.e. the targeted group), it tends to stay in this group in the following periods. Consequently, the regulator appears to be targeting the same facilities for higher inspection frequencies.

To try to understand the determinants of this targeting, we use a probit model to estimate the probability of non-memoryless comprehensive inspections. Building on previous literature (Eckert, 2004; Earnhart, 2004; Rousseau, 2007) as well as EPA's current recommendations for targeting we use a set of environmental performance variables and facility characteristic variables. We also separate the dataset into two groups in order to investigate wastewater treatment plants separately to allow the usage of a set of explanatory variables specific to wastewater plants.

Both previous literature and EPA's current recommendations claim that targeting should be applied to noncompliant facilities and environmental problems. However, the results of our estimations do not show any consistent significant estimates for environmental performance variables or enforcement action variables (Tables 8 and 9). Similarly, impaired water body dummies are not significant. Other tested variables (not shown in the tables) include travel time and percentage of minorities within three miles; neither of these variables is significant. Wastewater treatment plant-specific variables (e.g. flow capacity, inflow) also do not provide any evidence (Table 8). Overall, the information provided by the ICIS-NPDES does not permit us to identify the determinants of targeting for comprehensive inspections.

Our results show that although there is a sub group of facilities that are targeted for higher inspection frequencies over long periods, none of the environmental performance factors that we would expect to be determinants of targeting are significant. This suggests that the local jurisdictions are using information that is not available in the ICIS-NPDES database for targeting. Potential relevant variables involved in targeting for this type of inspection may be other facility characteristics that are not publicly available or other types of variable such as citizen complaints, demographics, or other political factors.

6

Discussion And Concluding Remarks

Economic theory indicates that environmental agencies should use targeting in order to increase compliance. This targeting should involve an analysis of environmental performance and specific characteristics of firms to determine an adapted assignment of regulatory agencies' limited resources.

In this thesis, we look for evidence of a sorting strategy of water-polluting facilities by the regulator. We consider inspections performed under the Clean Water Act in the seven local EPA jurisdictions of Illinois. For the period 2001 to 2009, we analyze inspection and violation data for 275 facilities classified as major water dischargers and regulated under the Clean Water Act. We report three main results.

First, we take into account the planning of inspections for each facility and identify those inspection schemes that are memoryless. With the exception of one jurisdiction, we find that reconnaissance inspections (the quickest, cheapest inspection type) are memoryless. For these inspections, we find that the rate of inspection increases for facilities closer to the local EPA agency, for wastewater treatment plants, and for facilities that previously received warnings. Interestingly, we find that in the jurisdictions that have memoryless reconnaissance inspections, violations of physical and chemical parameters tend to be memoryless too. In the one jurisdiction that has mostly non-memoryless reconnaissance inspections, a higher proportion of violations are non-memoryless too, which may indicate strategic behavior by facilities.

On the other hand, a significant proportion of comprehensive inspections is found not to be memoryless, so that inspections are not randomly distributed over time at the facility level. Memoryless comprehensive inspections, though, have higher inspection frequencies for the wastewater industry and shorter travel time. Together these results echo Rousseau (2007) by showing that different inspection types serve different purposes in the regulator's strategy. Future work on understanding regulatory objectives should account for different inspection types.

Second, non-memoryless comprehensive inspections occur at a higher frequency than memoryless comprehensive inspections which suggests that the regulator is targeting a subset of facilities. This kind of targeting is consistent with previous theoretical studies. Targeting is encouraged by EPA's current guidelines. However, we cannot identify any determinant for the observed targeting within the data extracted from the ICIS-NPDES database, such as inspection, enforcement, and violation histories, as well as firms' characteristics, such as discharging into an impaired water body. This suggests the local agencies are using other information in targeting facilities for inspections.

Third, even though most inspection schedules are memoryless, this does not mean that there is no targeting. In particular, we find that wastewater treatment plants and facilities that are located closer to local offices are inspected more frequently. This result shows that jurisdiction level parameters are important in determining regulatory behavior. Future studies should be undertaken at the jurisdiction level if possible.

The implications of this study concern primarily inspection strategies and resource allocation. At the local level jurisdictions do not seem to respond to parameters emphasized in EPA's current targeting recommendations and previous theoretical studies (Harrington, 1988; Magat and Viscusi, 1990). Budget constraints seem to be a limiting factor in the regulator's activity as both our findings for travel time and the general decrease in inspection frequencies over the years suggest. Nevertheless, inspections remain an important means of ensuring that wastewater treatment equipment is working properly. This role is particularly important in the wastewater industry because of limited treatment capacity and uncertain effluent concentrations due to variable wet weather flows. However, our results do not suggest that the regulator allocates more resources to noncompliant plants. Since the majority of plants are publicly owned, the regulator may not want to spend resources where the major concern is to upgrade or replace old equipment.

As the determinants of targeting for comprehensive inspections remain unclear, future work may investigate other variables such as detailed demographic parameters or political actions from the public or from interest groups. Also, because Illinois uses a combination of reconnaissance and comprehensive inspections, it may be interesting to compare targeting strategies with other states using the same combination of types of inspection or to other states using only comprehensive inspections. Another step in the analysis of inspection activity would be to interview a sample of EPA inspection officers to collect data about the objectives, demands, and contraints related to their activity. From a broader perspective, as we observe that travel time plays an important role in inspectors' activity, we could change the jurisdictions' boundaries and the inherent distribution of facilities within them and optimize the assignment of facilities to EPA agencies taking into account inspection schedules and travel time.

Tables

Table 1: Summary Statistics for Major Facilities in Illinois

Industry		Wastewater	Electric	Plastic & Resin	Other
Count		208	30	6	31
Impaired		122	21	5	19
Travel Time	Mean	55.37	62.42	80.01	53.00
(minutes)	SD	28.06	31.66	23.95	31.11
	Min	5.45	11.53	51.10	16.67
	Max	140.38	120.38	114.18	130.28
Inspection	Compr.	2025	208	51	235
	Recon.	9849	983	168	776
Violation	Overdue	1730	262	89	361
	Exceed.	5652	265	171	724
Enforcement	Informal	648	34	11	84
	Formal	247	8	7	33
Penalty	Count	23	2	2	2
($)	Sum	74,776,808	13,000	8,187,951	201,914,089
	Mean	302,740.11	1,625.00	1,169,707.29	6,118,608.76
	SD	3,816,508.09	3,543.10	3,090,595.70	35,135,269.92

Upper table shows facility related values.

Lower table shows regulatory action values over the industry.

Table 2: Complementary Summary Statistics for Wastewater Treatment Plants in Illinois

	Count	Mean	SD	Min	Max
Facilities	208				
Design Flow (million Gal/day)		15.8	89.61	< 1	1200
Pretreatment	69				
NON-POTW[1]	5				
POTW[1]	203				
Critical	13				

[1] (Non) Publicly Owned Treatment Works.

Table 3: Seasonality of Violations

Quarter[1]	Overdue	Limit	TSS	Chlorine[2]	BOD_5	Coliforms[3]
1	442	1666	501	34	90	255
2	508	1901	327	89	80	449
3	681	1754	239	87	45	457
4	811	1491	367	64	80	295

[1] Values represent sums per quarter during 2001-2009.

[2] Regression of yearly counts with quarter dummies show that quarter 2 and 3 are significantly greater than quarter 1 at the 0.001 level and than quarter 4 at the 0.1 level.

[3] Values for quarter 2 and 3 are significantly greater than 1 and 4 at the 0.001 level.

Table 4: Summary Statistics for Regulatory Action

	Mean	SD	Max	Min
Formal 2001	0.11	0.36	2	0
Formal 2002	0.10	0.31	2	0
Formal 2003	0.08	0.31	2	0
Formal 2004	0.11	0.35	2	0
Informal 2001	0.39	0.75	5	0
Informal 2002	0.39	0.70	4	0
Informal 2003	0.41	0.77	5	0
Informal 2004	0.25	0.56	2	0
Overdue 2001	0.56	2.33	24	0
Overdue 2002	0.59	2.35	33	0
Overdue 2003	0.58	1.71	14	0
Overdue 2004	0.64	2.01	21	0
Chlorine 2001	0.13	0.73	10	0
Chlorine 2002	0.15	0.64	8	0
Chlorine 2003	0.15	0.56	6	0
Chlorine 2004	0.10	0.50	6	0
BOD_5 2001	0.23	2.10	31	0
BOD_5 2002	0.24	1.65	22	0
BOD_5 2003	0.17	0.92	9	0
BOD_5 2004	0.13	0.75	8	0
Coliform 2001	0.52	1.69	12	0
Coliform 2002	0.61	1.68	12	0
Coliform 2003	0.47	1.44	12	0
Coliform 2004	0.61	1.71	12	0
TSS 2001	0.55	2.05	19	0
TSS 2002	0.60	1.74	13	0
TSS 2003	0.49	1.46	13	0
TSS 2004	0.45	1.24	9	0
Comp. Insp. Rate 2003-07	0.26	0.13	0.80	0
Comp. Insp. Rate 2004-08	0.25	0.12	0.75	0
Comp. Insp. Rate 2005-09	0.23	0.11	0.70	0
Reco. Insp. Rate 2003-07	1.28	0.83	3.35	0
Reco. Insp. Rate 2004-08	1.05	0.78	3.35	0
Reco. Insp. Rate 2005-09	0.86	0.73	3.25	0

Upper section shows formal and informal enforcements.

Middle section shows specific pollutants.

Lower section shows inspection rates for randomly inspected facilities.

Table 5: Non-Memoryless Inspections of Facilities (Proportions)

	Facilities[1]		Recon. Insp.[2]		Compr. Insp.[2]	
Jurisdiction	Total	Wastewater	Total	Wastewater	Total	Wastewater
1	28	21	0.00	0.00	0.32	0.38
2	121	98	0.02	0.00	0.19	0.18
3	32	21	0.81	0.81	0.50	0.48
4	25	17	0.04	0.06	0.36	0.47
5	20	13	0.00	0.00	0.35	0.31
6	28	21	0.00	0.00	0.43	0.52
7	21	17	0.00	0.00	0.38	0.47

[1] Counts of facilities. Values are presented for all facilities (Total) and for wasterwater treatment plants only (Wastewater).

[2] Proportions of facilities with non-memoryless inspections. Values are presented for all facilities (Total) and for wasterwater treatment plants only (Wastewater).

Table 6: Non-Memoryless Violations (Proportions)

Jurisdiction		1	2	3	4	5	6	7
Facilities	T	28	121	32	25	20	28	21
	W	21	98	21	17	13	21	17
Overdue	T	0.79	0.67	0.62	0.80	0.85	0.68	0.33
	W	0.76	0.67	0.48	0.76	0.77	0.67	0.29
Limit	T	0.61	0.60	0.78	0.52	0.60	0.75	0.67
	W	0.62	0.59	0.81	0.53	0.69	0.71	0.65
TSS	T	0.29	0.37	0.56	0.28	0.45	0.39	0.43
	W	0.38	0.38	0.57	0.24	0.38	0.33	0.41
Chlorine	T	0.00	0.03	0.16	0.04	0.05	0.07	0.00
	W	0.00	0.03	0.14	0.06	0.00	0.10	0.00
BOD_5	T	0.04	0.06	0.22	0.12	0.15	0.04	0.05
	W	0.00	0.02	0.14	0.00	0.08	0.00	0.00
Coliforms	T	0.18	0.12	0.31	0.12	0.05	0.07	0.29
	W	0.24	0.14	0.48	0.18	0.08	0.10	0.35

T: Total, W: Wastewater only

Table 7: OLS Regression of the Rate of Inspection

	Recon. 2004-08	Recon. 2005-09	Comp. 2004-08	Comp. 2005-09
(Intercept)	0.629***	0.624***	0.214***	0.197***
	(0.178)	(0.155)	(0.025)	(0.024)
Formal 2002	0.006		0.034	
	(0.144)		(0.024)	
Formal 2003	−0.271	−0.137	0.060**	0.031
	(0.161)	(0.207)	(0.020)	(0.020)
Formal 2004		−0.221		0.031
		(0.135)		(0.019)
Informal 2002	0.110		0.004	
	(0.074)		(0.011)	
Informal 2003	0.242***	0.193**	−0.009	−0.004
	(0.067)	(0.067)	(0.009)	(0.009)
Informal 2004		−0.002		−0.024*
		(0.098)		(0.012)
Travel Time (min)	−0.004*	−0.004**	$-5.024e{-4}$*	$-3.802e-4$
	(0.002)	(0.002)	($2.251e{-4}$)	($2.218e{-4}$)
Wastewater	0.676***	0.674***	0.051*	0.051*
	(0.150)	(0.130)	(0.022)	(0.021)
Electric	0.362	0.376*	0.001	0.013
	(0.188)	(0.167)	(0.028)	(0.027)
Plastic	0.219	0.221	0.033	−0.014
	(0.293)	(0.317)	(0.048)	(0.047)
Log-likelihood	−269.785	−229.953	213.377	217.890
N	248	215	227	226

standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Probability of Non-Memoryless Comprehensive Inspections for Wastewater Facilities

	Comp. 2003-07	Comp. 2004-08	Comp. 2005-09
(Intercept)	-0.780^{***}	-0.947^{***}	-0.920^{***}
	(0.173)	(0.187)	(0.192)
Impaired	-0.547^{**}	-0.357	-0.211
	(0.212)	(0.223)	(0.220)
Design Flow	-0.004	-0.008	-0.005
(million Gal/Day)	(0.005)	(0.010)	(0.008)
NON-POTW[1]	0.121	-4.274	0.176
	(0.673)	(262.395)	(0.662)
Pretreatment	0.722^{**}	0.424	0.140
	(0.223)	(0.243)	(0.242)
Critical	-0.129	-0.883	-0.722
	(0.420)	(0.606)	(0.552)
Informal 2001	0.325^{**}	0.217	0.452^{***}
	(0.126)	(0.127)	(0.131)
Informal 2002	0.025	0.153	-0.007
	(0.143)	(0.140)	(0.148)
Informal 2003		0.088	0.149
		(0.134)	(0.137)
Informal 2004			-0.319
			(0.220)
Log-likelihood	-103.141	-91.347	-92.529
N	208	208	208

standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

[1] Non Public Owned Treatment Works.

Table 9: Probability of Non-Memoryless Comprehensive Inspections for Other Facilities

	Comp. 2003-07	Comp. 2004-08	Comp. 2005-09
(Intercept)	-1.404^{**}	-1.654^{**}	-1.151
	(0.461)	(0.571)	(0.639)
Impaired	0.607	0.764	0.176
	(0.504)	(0.598)	(0.626)
Informal 2001	-0.133	0.054	0.177
	(0.412)	(0.406)	(0.393)
Informal 2002	-1.128	-0.619	-0.955
	(0.838)	(0.685)	(1.103)
Informal 2003		-0.812	-0.785
		(0.639)	(0.667)
Informal 2004			0.943
			(0.581)
Mean Exceed. 2001	0.103	0.138^{*}	-0.006
	(0.063)	(0.069)	(0.038)
Mean Exceed. 2002	0.006	0.036	0.030
	(0.004)	(0.020)	(0.021)
Mean Exceed. 2003		0.015	0.028
		(0.011)	(0.040)
Mean Exceed. 2004			-0.393
			(0.626)
Max. Exceed. 2001	-0.071	-0.096^{*}	-0.003
	(0.044)	(0.048)	(0.013)
Max. Exceed. 2002	-0.002	-0.017	-0.015
	(0.001)	(0.011)	(0.012)
Max. Exceed. 2003		-0.003	-0.023
		(0.002)	(0.038)
Max. Exceed. 2004			0.084
			(0.129)
Log-likelihood	-22.870	-19.988	-14.378
N	67	67	67

standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figures

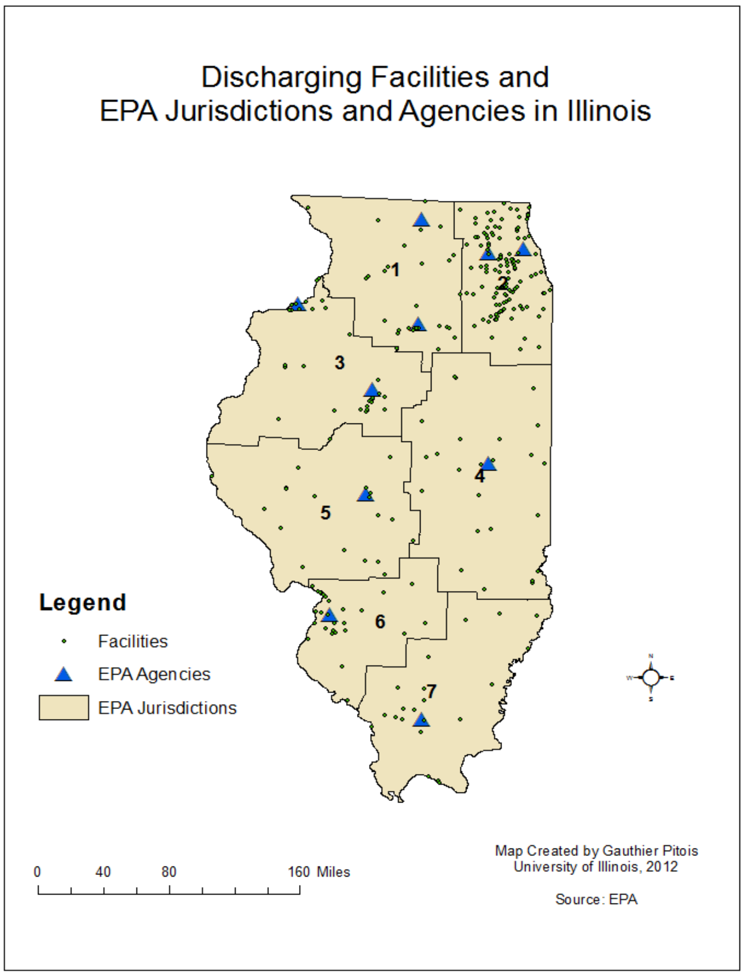

Figure 1: Industry and EPA in Illinois

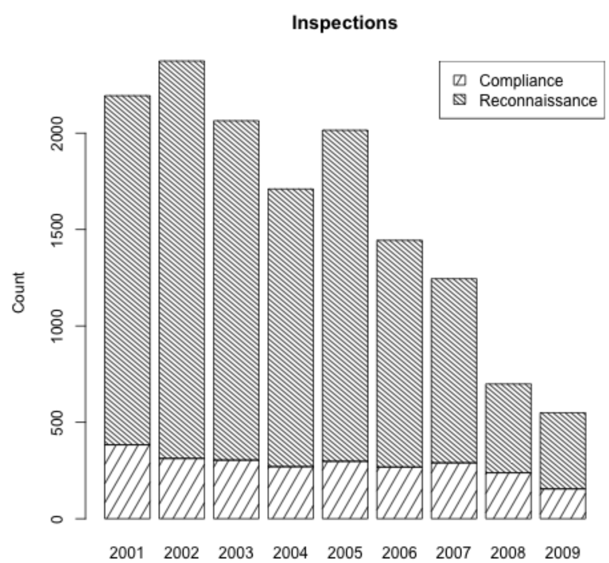

Figure 2: Counts of Inspections over Time

Appendix A

Additional Material

A.1 Independence Test for Inspection Types

The following are the results of the test for independence of non-memoryless (different from a Poisson process) comprehensive inspections and non-memoryless reconnaissance inspections performed on the whole dataset (275 facilities). The null hypothesis states that variables outcomes are statistically independent.

Pearson's Chi-squared test with Yates' continuity correction:

$$\chi^2 = 0.0053, df = 1, pvalue = 0.942$$

We cannot reject the null hypothesis and conclude that having non-memoryless comprehensive inspections and having non-memoryless reconnaissance inspections are independent processes.

A.2 Evolution of Environmental Inspections

Inspection activities appear to have changed over the recent years in Illinois. There is a general decreasing trend in the number of inspection in all jurisdictions (Table A.1).

Table A.1: Facility Average Number of Inspections by Year across Jurisdictions

Jurisdiction	2001	2002	2003	2004	2005	2006	2007	2008	2009
1	9.04	8.68	8.39	8.00	7.50	7.86	7.89	4.18	2.46
2	5.02	6.17	6.17	5.13	5.87	2.21	1.40	1.02	0.85
3	12.69	12.12	5.19	2.84	5.78	3.78	1.25	1.09	0.34
4	8.48	8.36	6.72	6.28	8.56	7.00	7.24	3.00	3.20
5	12.30	10.85	10.30	9.05	9.00	9.15	9.85	4.65	3.05
6	7.82	11.21	10.96	8.96	10.61	9.93	9.36	5.04	4.71
7	11.95	12.24	11.24	8.81	10.43	9.48	8.33	5.48	4.48

Usage of inspection types also change over the period 2001-2009. Figure A.1 shows the proportions of each inspection type for jurisdiction 1, corresponding to the northwestern corner of Illinois. We observe that reconnaissance inspections started to comprise water sampling in 2007. Figure A.2 shows that jurisdiction 2, corresponding to the Chicago area, has a different evolution with an increase of the proportion of comprehensive inspections.

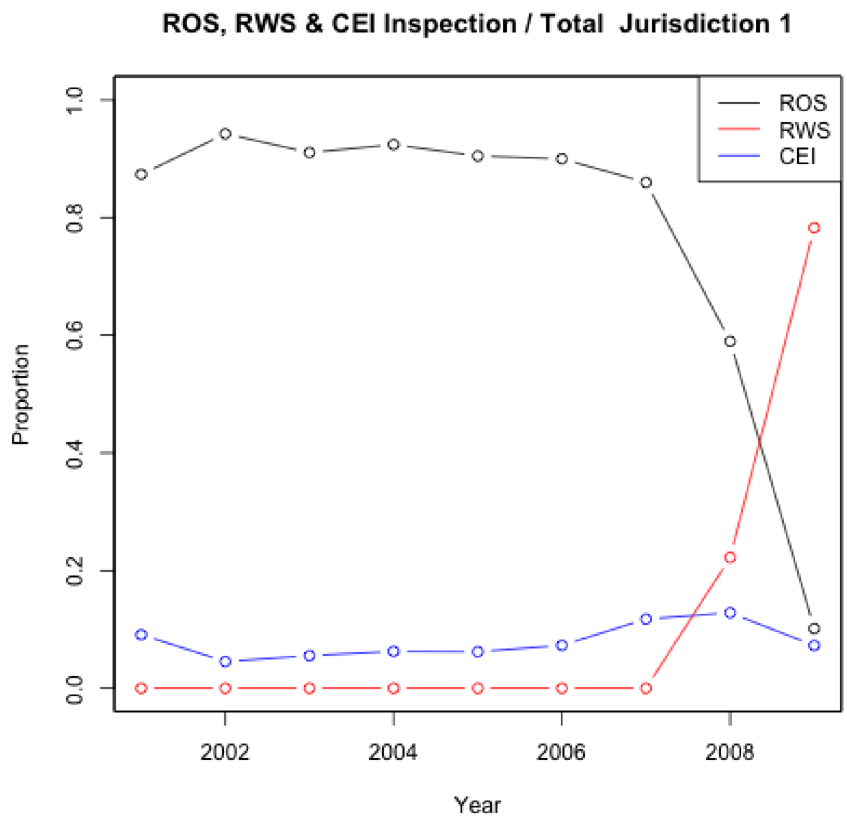

Figure A.1: Evolution of Inspection Types in Jurisdiction 1

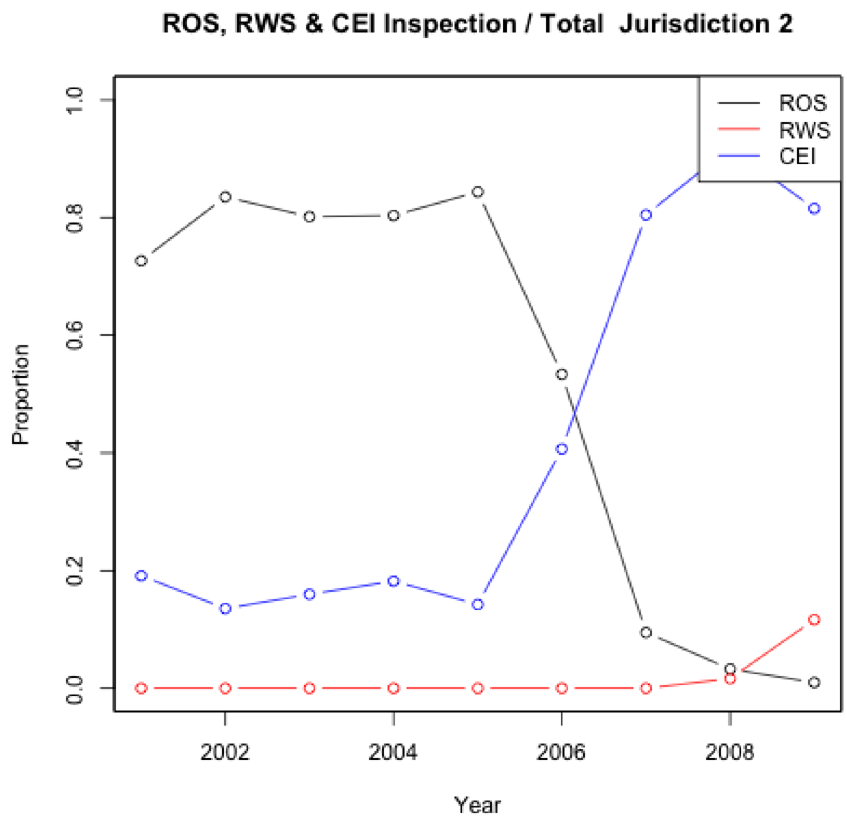

Figure A.2: Evolution of Inspection Types in Jurisdiction 2

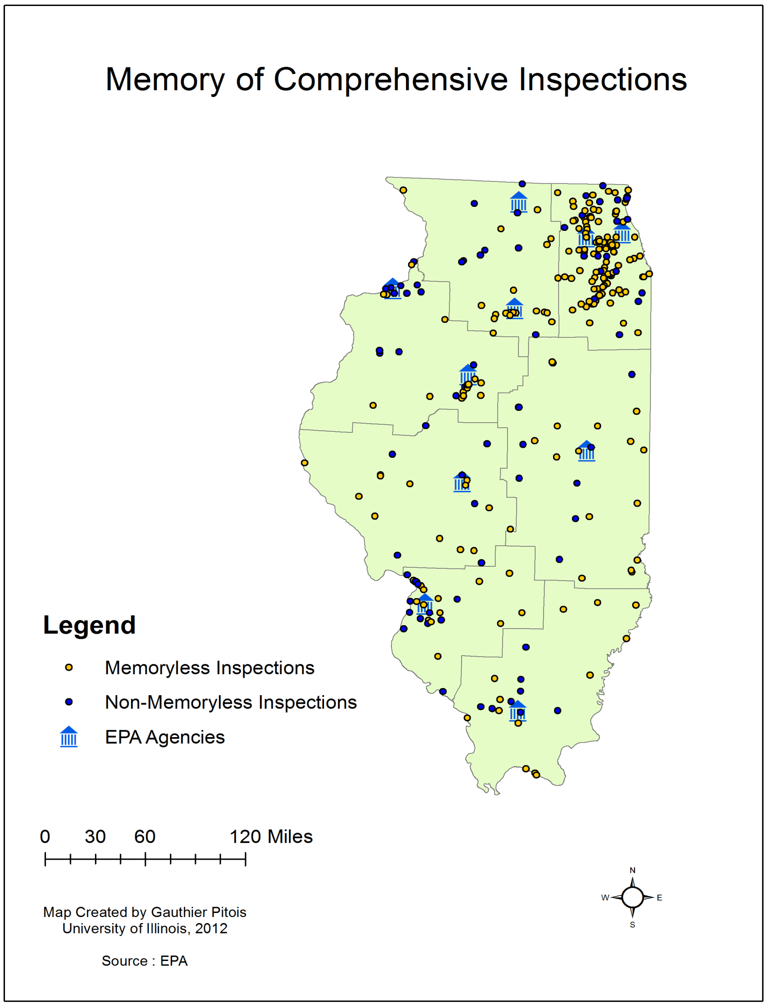

Figure A.3: Facilities with Memoryless Comprehensive Inspections

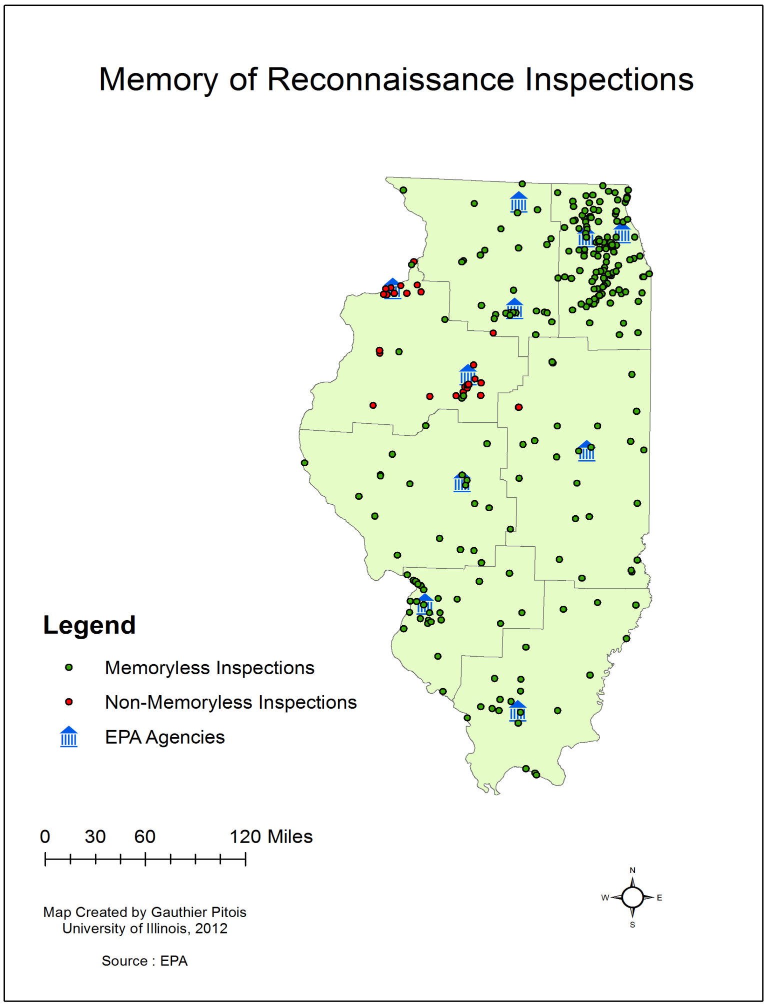

Figure A.4: Facilities with Memoryless Reconnaissance Inspections

A.3 Evolution of Violations

Tables A.2 and A.3 present respectively the counts of violation per year per jurisdiction and the facility-average counts of violation per year per jurisdiction.

Table A.2: Number of Violations per Year

Jurisdiction	2001	2002	2003	2004	2005	2006	2007	2008	2009
1	146	109	77	95	102	83	82	143	128
2	284	479	503	360	426	412	404	461	610
3	150	144	92	113	150	123	140	219	272
4	70	61	70	120	52	54	74	68	115
5	51	71	97	68	87	61	65	99	127
6	40	75	66	72	60	70	75	168	256
7	94	92	69	66	57	74	35	91	77

Table A.3: Violations per Facility (Average)

Jurisdiction	2001	2002	2003	2004	2005	2006	2007	2008	2009
1	5.21	3.89	2.75	3.39	3.64	2.96	2.93	5.11	4.57
2	2.35	3.96	4.16	2.98	3.52	3.40	3.34	3.81	5.04
3	4.69	4.50	2.88	3.53	4.69	3.84	4.38	6.84	8.50
4	2.80	2.44	2.80	4.80	2.08	2.16	2.96	2.72	4.60
5	2.55	3.55	4.85	3.40	4.35	3.05	3.25	4.95	6.35
6	1.43	2.68	2.36	2.57	2.14	2.50	2.68	6.00	9.14
7	4.48	4.38	3.29	3.14	2.71	3.52	1.67	4.33	3.67

A.4 Robustness Tests for OLS Regressions

As we use an OLS model to estimate the factors involved in the rate of inspection for facilities randomly inspected over time, we perform a series of robustness tests. We first test the model on three different time settings: we regress the rate of inspection during 2003-07 with 1-year and 2-year lagged variables and repeat the estimation after shifting dependent variable to 2004-08 and then 2005-09. Results are shown in Tables A.4 and A.5.

Our next robustness tests concern multicollinearity and heteroskedasticity. The former is tested using the variance inflation factor (VIF) and the later is tested with the Breusch-Pagan test. Both are performed within R with already existing functions. Table A.6 shows the VIF results and Table A.7 shows the Breusch-Pagan test. There is no sign of mulcollinearity, however models for reconnaissance inspections for 2004-08 and 2005-09 have heteroskedasticity. Thus, these models are corrected using White's standard errors in Table 7.

A.5 Robustness Tests for Probit Regressions

Table A.8 shows the predictions of our different probit models. The fourth column contains the percentage of correct predictions and the fifth column contains the percentage of the largest observed value.

Table A.4: OLS Regression of the Rate of Reconnaissance Inspection

	2003-07	2004-08	2005-09
(Intercept)	0.809***	0.629***	0.624***
	(0.174)	(0.173)	(0.172)
Formal 2001	0.132		
	(0.153)		
Formal 2002	−0.112	0.006	
	(0.171)	(0.161)	
Formal 2003		−0.271	−0.137
		(0.173)	(0.195)
Formal 2004			−0.221
			(0.144)
Informal 2001	0.089		
	(0.075)		
Informal 2002	0.162*	0.110	
	(0.071)	(0.070)	
Informal 2003		0.242***	0.193**
		(0.065)	(0.068)
Informal 2004			−0.002
			(0.096)
Travel Time (min)	−0.004*	−0.004*	−0.004**
	(0.002)	(0.002)	(0.002)
Wastewater	0.795***	0.676***	0.674***
	(0.158)	(0.153)	(0.149)
Electric	0.310	0.362	0.376
	(0.208)	(0.203)	(0.199)
Plastic	0.116	0.219	0.221
	(0.376)	(0.361)	(0.390)
R-squared	0.175	0.178	0.159
Log-likelihood	−288.202	−269.785	−229.953
N	253	248	215

standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5: OLS Regression of the Rate of Comprehensive Inspections

	2003-07	2004-08	2005-09
(Intercept)	0.225***	0.214***	0.197***
	(0.026)	(0.025)	(0.024)
Formal 2001	0.032		
	(0.023)		
Formal 2002	0.059*	0.034	
	(0.026)	(0.024)	
Formal 2003		0.060**	0.031
		(0.020)	(0.020)
Formal 2004			0.031
			(0.019)
Informal 2001	−0.015		
	(0.011)		
Informal 2002	0.011	0.004	
	(0.010)	(0.011)	
Informal 2003		−0.009	−0.004
		(0.009)	(0.009)
Informal 2004			−0.024*
			(0.012)
Travel Time (min)	$-5.314e{-4}^{*}$	$-5.024e{-4}^{*}$	$-3.802e-4$
	$(2.405e{-4})$	$(2.251e{-4})$	$(2.218e{-4})$
Wastewater	0.041	0.051*	0.051*
	(0.023)	(0.022)	(0.021)
Electric	−0.012	0.001	0.013
	(0.028)	(0.028)	(0.027)
Plastic	0.051	0.033	−0.014
	(0.050)	(0.048)	(0.047)
R-squared	0.115	0.128	0.096
Log-likelihood	192.440	213.377	217.890
N	216	227	226

standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.6: Variance Inflation Factor for OLS Regressions

Model	FE01	FE02	FE03	FE04	IE01	IE02	IE03	IE04	TT	W	E	P
Rec. 2003-07	1.39	1.17			1.44	1.09			1.03	1.94	1.81	1.17
Rec. 2004-08		1.11	1.16			1.16	1.17		1.04	1.98	1.86	1.19
Rec. 2005-09			1.11	1.12			1.22	1.19	1.05	1.83	1.81	1.15
Comp. 2003-07	1.28	1.08			1.31	1.09			1.02	2.15	2.01	1.20
Comp. 2004-08		1.15	1.12			1.19	1.18		1.02	2.28	2.15	1.22
Comp. 2005-09			1.11	1.15			1.30	1.25	1.03	2.18	2.09	1.22

FE: Formal Enforcement, IE: Informal Enforcement, TT: Travel Time

W: Wastewater, E: Electric Services, P: Plastic

Table A.7: Heteroskedasticity Test for OLS Regressions

Model	Breusch-Pagan Test	df	p value
Rec. 2003-07	8.156	8	0.4184
Rec. 2004-08	19.087	8	0.0144
Rec. 2005-09	28.687	8	0.00036
Comp. 2003-07	10.942	8	0.205
Comp. 2004-08	5.168	8	0.7395
Comp. 2005-09	4.575	8	0.8019

Table A.8: Prediction of Probit Models

Model	N	Hit	% Correct	% Largest Value
Comprehensive 2003-07 W	208	161	77.40	75.96
Comprehensive 2004-08 W	208	171	82.21	81.73
Comprehensive 2005-09 W	208	169	81.25	80.77
Comprehensive 2003-07 O	67	57	85.07	86.57
Comprehensive 2004-08 O	67	58	86.57	85.07
Comprehensive 2005-09 O	67	61	91.04	86.57

W: Wastewater, O:Other

A.6 Proportions of Memoryless Inspections

The following tables show the proportions of facilities with non-memoryless inspections over different time windows.

Table A.9: Non-Memoryless Inspections of Facilities from 2001 to 2005 (Proportions)

	Facilities		Recon. Insp.		Compr. Insp.	
Jurisdiction	Total	Wastewater	Total	Wastewater	Total	Wastewater
1	28	21	0.00	0.00	0.25	0.29
2	121	98	0.02	0.00	0.22	0.22
3	32	21	0.12	0.14	0.47	0.57
4	25	17	0.04	0.06	0.32	0.41
5	20	13	0.00	0.00	0.30	0.38
6	28	21	0.00	0.00	0.46	0.48
7	21	17	0.00	0.00	0.33	0.41

Table A.10: Non-Memoryless Inspections of Facilities from 2002 to 2006 (Proportions)

	Facilities		Recon. Insp.		Compr. Insp.	
Jurisdiction	Total	Wastewater	Total	Wastewater	Total	Wastewater
1	28	21	0.00	0.00	0.18	0.19
2	121	98	0.02	0.00	0.16	0.15
3	32	21	0.31	0.33	0.50	0.48
4	25	17	0.04	0.06	0.36	0.53
5	20	13	0.00	0.00	0.15	0.15
6	28	21	0.00	0.00	0.25	0.29
7	21	17	0.00	0.00	0.29	0.35

Table A.11: Non-Memoryless Inspections of Facilities from 2003 to 2007 (Proportions)

	Facilities		Recon. Insp.		Compr. Insp.	
Jurisdiction	Total	Wastewater	Total	Wastewater	Total	Wastewater
1	28	21	0.04	0.05	0.29	0.33
2	121	98	0.02	0.01	0.13	0.13
3	32	21	0.56	0.67	0.34	0.33
4	25	17	0.00	0.00	0.28	0.41
5	20	13	0.00	0.00	0.05	0.08
6	28	21	0.00	0.00	0.36	0.43
7	21	17	0.00	0.00	0.29	0.35

Table A.12: Non-Memoryless Inspections of Facilities from 2004 to 2008 (Proportions)

	Facilities		Recon. Insp.		Compr. Insp.	
Jurisdiction	Total	Wastewater	Total	Wastewater	Total	Wastewater
1	28	21	0.00	0.00	0.25	0.29
2	121	98	0.08	0.08	0.08	0.08
3	32	21	0.50	0.52	0.34	0.24
4	25	17	0.04	0.00	0.12	0.18
5	20	13	0.00	0.00	0.15	0.23
6	28	21	0.00	0.00	0.29	0.33
7	21	17	0.00	0.00	0.29	0.35

Table A.13: Non-Memoryless Inspections of Facilities from 2005 to 2009 (Proportions)

	Facilities		Recon. Insp.		Compr. Insp.	
Jurisdiction	Total	Wastewater	Total	Wastewater	Total	Wastewater
1	28	21	0.00	0.00	0.32	0.38
2	121	98	0.31	0.37	0.08	0.09
3	32	21	0.66	0.71	0.31	0.19
4	25	17	0.04	0.00	0.16	0.24
5	20	13	0.05	0.08	0.20	0.31
6	28	21	0.00	0.00	0.18	0.24
7	21	17	0.00	0.00	0.33	0.35

Appendix B

Code

B.1 SQL Extraction From the ICIS-NPDES Database

This section presents the SQL queries used to extract data from the ICIS-NPDES. After downloading records from the EPA ECHO website, we create a database comprising one SQL table for each ICIS-NPDES table (e.g. inspections, violations) and load the data into them. The following queries then extract relevant variables for statistical analyses.

SQL Queries

```
select perexno, viotype, viocode, vioparm, vioparm_text, viomvdt, vioepct,
 viovtcd, year(viomvdt) as year, month(viomvdt) as month
 from ECHO.dmr_violations
 where 2001 <= year(viomvdt) and year(viomvdt) <= 2009;

select perexno, acttypc, cpoa, f_sttldt, actsepa, f_enfcoac_text, t_enfcfpa,
 t_enfcslp, t_enfctsa, t_enfccaa, year(f_sttldt), month(f_sttldt)
 from ECHO.formal_enforcement_actions enf
 where 2001 <= year(f_sttldt) and year(f_sttldt) < 2010;

select perexno, enftypc, enftype, enfdate, enfsepa, year(enfdate),
 month(enfdate)
 from ECHO.informal_enforcement_actions enf
 where 2001 <= year(enfdate) and year(enfdate) < 2010;

select ins.perexno, codes.sic, cmpmoty, cmpaced,
 year(cmpaced), month(cmpaced)
 from ECHO.inspections ins
 join ECHO.snc_naic_code codes on (ins.perexno = codes.perexno)
```

```
where 2001 <= year(cmpaced) and year(cmpaced) < 2010;
```

B.2 Travel Duration with GoogleMaps in Python

```
# Facilities file contains lat/lon coordinates for each plant
# and lat/lon coordinates of the corresponding EPA office

from geopy import distance
from geopy import geocoders
from geopy.point import Point
from googlemaps import GoogleMaps
import csv
import re

gmaps = GoogleMaps()

outWriter = csv.writer(open('facitiliestraveltime.csv', 'wb'), \
                           quoting=csv.QUOTE_MINIMAL)

with open('../data/facilities.csv', "rb" ) as theFile:
    reader = csv.DictReader( theFile )
    for line in reader:
        start = gmaps.latlng_to_address(float(line['jurlat']), \
                                    float(line['jurlon']))
        dest  = gmaps.latlng_to_address(float(line['lat']), \
                                    float(line['lon']))
        correctedDest = re.sub('Unnamed Rd, ','',dest)
        dirs  = gmaps.directions(start, correctedDest)
        time  = dirs['Directions']['Duration']['seconds']
        dist  = dirs['Directions']['Distance']['meters']
        res   = [line['perexno'],line['jurisdiction'],time, dist]
        print res
        outWriter.writerow(res)
```

B.3 Statistics and Econometrics with R

Poisson Process Analysis of Inspections

```
library(doBy)
library(xtable)
library(reshape)

# Conditional Chi Sq
TCC <- function(X)
{
  X.mean = mean(X)
  n = length(X)
  t.cc = 0
  S2 = var(X)
  t.cc = (n-1) * S2 / X.mean
  integral = pchisq(t.cc, n-1)
  pval = 1 - integral
  return(c(t.cc, integral, pval))
}

facilities = read.csv("facilitiessimple.csv")
quarterid = seq(1,36)
inspquarterly = merge(facilities,quarterid)
names(inspquarterly)[names(inspquarterly)=="y"]<-"quarterid"
Recon = seq(0,1)
inspquarterly = merge(inspquarterly, Recon)
names(inspquarterly)[names(inspquarterly)=="y"]<-"Recon"

inspcountquart = read.csv("2001-2009-all-insp-recon-quarterid.csv")
inspcountquart = merge(inspquarterly,inspcountquart, all=TRUE)
inspcountquart[is.na(inspcountquart$nbinsp),]$nbinsp = 0

nbplant = dim(facilities)[1]
inspintvl = subset(inspcountquart, 17 <= quarterid & quarterid <= 36)
resultpoisson = data.frame(matrix(nrow=nbplant, ncol=7))
names(resultpoisson) <- c("perexno","juris","SIC","Rec.pval","Nrec.pval",
                          "meanrec","meancom")

for (i in 1:nbplant)
{
  plantsub = subset(inspintvl,
```

```
                  perexno == facilities$perexno[i] & Recon == 1)
  Rec = TCC(plantsub$nbinsp)
  mrec = mean(plantsub$nbinsp)
  plantsub = subset(inspintvl,
                  perexno == facilities$perexno[i] & Recon == 0)
  Nrec = TCC(plantsub$nbinsp)
  mcom = mean(plantsub$nbinsp)
  resultpoisson[i,1] <- as.character(facilities$perexno[i])
  resultpoisson[i,c(2:7)] <- c(facilities$juris[i], facilities$SIC[i],
                             Rec[3], Nrec[3], mrec, mcom)
}

resultpoisson$wastewat = 0
resultpoisson[resultpoisson$SIC == 4952,]$wastewat = 1
resultpoisson$RecNP = 0
resultpoisson[is.na(resultpoisson$Rec.pval)==FALSE &
              resultpoisson$Rec.pval <= 0.1,]$RecNP = 1
resultpoisson$ComNP = 0
resultpoisson[is.na(resultpoisson$Nrec.pval)==FALSE &
              resultpoisson$Nrec.pval <= 0.1,]$ComNP = 1
write.csv(resultpoisson, "facilities-poissonness.csv")


datacum = resultpoisson
datacum = merge(datacum, resultpoisson)
names(datacum)[names(datacum)=="RecNP"]<-"NPR0509"
names(datacum)[names(datacum)=="ComNP"]<-"NPC0509"
datacum = datacum[c(-4,-5,-6,-7)]
datacum = datacum[c(-8,-9,-10,-11)]

# Table Non Poissonness of Inspections
resultpoisson$wastewat = 0
resultpoisson[resultpoisson$SIC == 4952,]$wastewat = 1
outputpoisson = summaryBy(perexno ~ juris + wastewat, data=facilities,
                          FUN=c(length))
names(outputpoisson)[names(outputpoisson)=="perexno.length"] = "nbPlants"

# Number of Non-Reconnaissance Inspections that do not follow
# a Poisson Process
plantsub = subset(resultpoisson, Nrec.pval <= 0.1)
temp = summaryBy(perexno ~ juris + wastewat, data=plantsub, FUN=c(length))
names(temp)[names(temp)=="perexno.length"] = "NrecNotPoisson"
outputpoisson = merge(outputpoisson, temp, all=TRUE)
outputpoisson[is.na(outputpoisson$NrecNotPoisson),]$NrecNotPoisson = 0
```

```
# Number of Reconnaissance Inspections that do not follow Poisson Process
plantsub = subset(resultpoisson, Rec.pval <= 0.1)
temp = summaryBy(perexno ~ juris + wastewat, data=plantsub, FUN=c(length))
names(temp)[names(temp)=="perexno.length"] = "RecNotPoisson"
outputpoisson = merge(outputpoisson, temp, all=TRUE)
outputpoisson[is.na(outputpoisson$RecNotPoisson),]$RecNotPoisson = 0

temp = melt(outputpoisson, c("juris", "wastewat"))

result = cast(temp, juris ~ wastewat + variable)
result$nbPlantTot = result$`0_nbPlants` + result$`1_nbPlants`
result$nbWaste = result$`1_nbPlants`
result$nbNotRandRecon = result$`0_RecNotPoisson` +
                        result$`1_RecNotPoisson`
result$nbWasteNotRandRecon = result$`1_RecNotPoisson`
result$nbNotRandNrecon = result$`0_NrecNotPoisson` +
                         result$`1_NrecNotPoisson`
result$nbWasteNotRandNrecon = result$`1_NrecNotPoisson`

temp = result[c(-2,-3,-4,-5,-6,-7)]
temp$propTotNRndRecon = temp$nbNotRandRecon / temp$nbPlantTot
temp$propWasNRndRecon = temp$nbWasteNotRandRecon / temp$nbWaste
temp$propTotNRndNrecon = temp$nbNotRandNrecon / temp$nbPlantTot
temp$propWasNRndNrecon = temp$nbWasteNotRandNrecon / temp$nbWaste
```

Poisson Process Analysis of Violations

```
# Poisson Test fot violations
vio = read.csv("2001-2009-all-vio-date-v2.csv")
v1 = merge(vio, facilities)
v1$quarterRel = (vio$month-1)%/%3 + 1
v1$quarterAbs = (vio$year-2001)*4 + (vio$month-1)%/%3 + 1

quarterid = seq(1,36)
vioquarterly = merge(facilities,quarterid)
names(vioquarterly)[names(vioquarterly)=="y"]<-"quarterid"

tmp = subset(v1, viocode != "E90")
t = summaryBy(perexno ~ perexno + quarterAbs, data=tmp, FUN=length)
names(t)[names(t)=="perexno.length"]<-"nbVioOverdue"
vioquarterly = merge(vioquarterly, t, by.x=c("perexno","quarterid"),
```

```
                     by.y=c("perexno","quarterAbs") ,all.x=T)
tmp = subset(v1, viocode == "E90")
t = summaryBy(perexno ~ perexno + quarterAbs, data=tmp, FUN=length)
names(t)[names(t)=="perexno.length"]<-"nbVioLimit"
vioquarterly = merge(vioquarterly, t, by.x=c("perexno","quarterid"),
                     by.y=c("perexno","quarterAbs") ,all.x=T)
#TSS
tmp = subset(v1, vioparm==530)
t = summaryBy(perexno ~ perexno + quarterAbs, data=tmp, FUN=length)
names(t)[names(t)=="perexno.length"]<-"nbVioTSS"
vioquarterly = merge(vioquarterly, t, by.x=c("perexno","quarterid"),
                     by.y=c("perexno","quarterAbs") ,all.x=T)
#Chlorine
tmp = subset(v1, vioparm==50060)
t = summaryBy(perexno ~ perexno + quarterAbs, data=tmp, FUN=length)
names(t)[names(t)=="perexno.length"]<-"nbVioChlo"
vioquarterly = merge(vioquarterly, t, by.x=c("perexno","quarterid"),
                     by.y=c("perexno","quarterAbs") ,all.x=T)
#BOD5
tmp = subset(v1, vioparm==310)
t = summaryBy(perexno ~ perexno + quarterAbs, data=tmp, FUN=length)
names(t)[names(t)=="perexno.length"]<-"nbVioBOD5"
vioquarterly = merge(vioquarterly, t, by.x=c("perexno","quarterid"),
                     by.y=c("perexno","quarterAbs") ,all.x=T)
#coliform
tmp = subset(v1, vioparm==74055)
t = summaryBy(perexno ~ perexno + quarterAbs, data=tmp, FUN=length)
names(t)[names(t)=="perexno.length"]<-"nbVioColi"
vioquarterly = merge(vioquarterly, t, by.x=c("perexno","quarterid"),
                     by.y=c("perexno","quarterAbs") ,all.x=T)

vioquarterly[is.na(vioquarterly)] <- 0
vioquarterly$quarterRel = vioquarterly$quarterid %% 4
vioquarterly[vioquarterly$quarterRel==0,]$quarterRel <- 4

tmp = summaryBy(nbVioOverdue + nbVioLimit + nbVioTSS + nbVioChlo
                + nbVioBOD5 + nbVioColi ~ quarterRel,
                data=vioquarterly, FUN=sum)
xtable(tmp)

resultPoisTotal = facilities
resultPoisVio = data.frame(matrix(nrow=nbplant, ncol=5))
names(resultPoisVio) <- c("perexno","juris","SIC","pval","meanVioColi")
for (i in 1:nbplant)
```

```
{
  plantsub = subset(vioquarterly, perexno == facilities$perexno[i])
  pval = TCC(plantsub$nbVioColi)
  mvio = mean(plantsub$nbVioColi)
  resultPoisVio[i,1] <- as.character(facilities$perexno[i])
  resultPoisVio[i,c(2:5)] <- c(facilities$juris[i],
                              facilities$SIC[i], pval[3], vio)
}

resultPoisVio$NPColi = 0
resultPoisVio[is.na(resultPoisVio$pval)==FALSE &
              resultPoisVio$pval <= 0.1,]$NPColi = 1
resultPoisTotal = merge(resultPoisTotal, resultPoisVio[c(-4)])

write.csv(resultPoisTotal, "ResultsPoissonViolation.csv")

output = summaryBy(perexno ~ juris, data=resultPoisTotal, FUN=length)
names(output)[names(output)=="perexno.length"]<-"nbFacility"
output = merge(output, summaryBy(perexno ~ juris,
               data=subset(resultPoisTotal, sicGroup==1), FUN=length),
               all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbWasteWater"
output = merge(output, summaryBy(perexno ~ juris,
               data=subset(resultPoisTotal, NPOverdue==1), FUN=length),
               all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPOverdue"
output = merge(output, summaryBy(perexno ~ juris,
               data=subset(resultPoisTotal,
                           NPOverdue==1 & sicGroup==1), FUN=length),
               all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPOverdueWW"
output = merge(output, summaryBy(perexno ~ juris,
               data=subset(resultPoisTotal, NPLimit==1), FUN=length),
               all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPLimit"
output = merge(output, summaryBy(perexno ~ juris,
               data=subset(resultPoisTotal, NPLimit==1 & sicGroup==1),
               FUN=length), all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPLimitWW"
output = merge(output, summaryBy(perexno ~ juris,
               data=subset(resultPoisTotal, NPTSS==1), FUN=length),
               all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPTSS"
output = merge(output, summaryBy(perexno ~ juris,
```

```
                data=subset(resultPoisTotal, NPTSS==1 & sicGroup==1),
                FUN=length), all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPTSSWW"
output = merge(output, summaryBy(perexno ~ juris,
                data=subset(resultPoisTotal, NPChlo==1), FUN=length),
                all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPChlo"
output = merge(output, summaryBy(perexno ~ juris,
                data=subset(resultPoisTotal, NPChlo==1 & sicGroup==1),
                FUN=length),  all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPChloWW"
output = merge(output, summaryBy(perexno ~ juris,
                data=subset(resultPoisTotal, NPBOD5==1), FUN=length),
                all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPBOD5"
output = merge(output, summaryBy(perexno ~ juris,
                data=subset(resultPoisTotal, NPBOD5==1 & sicGroup==1),
                FUN=length),  all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPBOD5WW"
output = merge(output, summaryBy(perexno ~ juris,
                data=subset(resultPoisTotal, NPColi==1), FUN=length),
                all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPColi"
output = merge(output, summaryBy(perexno ~ juris,
                data=subset(resultPoisTotal, NPColi==1 & sicGroup==1),
                FUN=length), all.x=T)
names(output)[names(output)=="perexno.length"]<-"nbNPColiWW"

output[is.na(output)] <- 0
output$pctOverdue = output$nbNPOverdue / output$nbFacility
output$pctOverdueWW = output$nbNPOverdueWW / output$nbWasteWater
output$pctLimit = output$nbNPLimit / output$nbFacility
output$pctLimitWW = output$nbNPLimitWW / output$nbWasteWater
output$pctTSS = output$nbNPTSS / output$nbFacility
output$pctTSSWW = output$nbNPTSSWW / output$nbWasteWater
output$pctChlo = output$nbNPChlo / output$nbFacility
output$pctChloWW = output$nbNPChloWW / output$nbWasteWater
output$pctBOD5 = output$nbNPBOD5 / output$nbFacility
output$pctBOD5WW = output$nbNPBOD5WW / output$nbWasteWater
output$pctColi = output$nbNPColi / output$nbFacility
output$pctColiWW = output$nbNPColiWW / output$nbWasteWater
```

OLS Regression of the Rate of Inspection

```
inspintvl = subset(inspcountquart, 9 <= quarterid & quarterid <= 28)
insprate = summaryBy(nbinsp ~ perexno + Recon, data=inspintvl, FUN=mean)
names(insprate)[names(insprate)=="nbinsp.mean"] = "IR"
insprate[insprate$Recon==0,]$Recon = "C"
insprate[insprate$Recon==1,]$Recon = "R"

facIR = cast(insprate, perexno ~ Recon)
names(facIR)[names(facIR)=="C"] = "IRC0307"
names(facIR)[names(facIR)=="R"] = "IRR0307"

datacum <- merge(datacum, facIR, all.x=T)

inspintvl = subset(inspcountquart, 13 <= quarterid & quarterid <= 32)
insprate = summaryBy(nbinsp ~ perexno + Recon, data=inspintvl, FUN=mean)
names(insprate)[names(insprate)=="nbinsp.mean"] = "IR"
insprate[insprate$Recon==0,]$Recon = "C"
insprate[insprate$Recon==1,]$Recon = "R"

facIR = cast(insprate, perexno ~ Recon)
names(facIR)[names(facIR)=="C"] = "IRC0408"
names(facIR)[names(facIR)=="R"] = "IRR0408"

datacum <- merge(datacum, facIR, all.x=T)

inspintvl = subset(inspcountquart, 17 <= quarterid & quarterid <= 36)
insprate = summaryBy(nbinsp ~ perexno + Recon, data=inspintvl, FUN=mean)
names(insprate)[names(insprate)=="nbinsp.mean"] = "IR"
insprate[insprate$Recon==0,]$Recon = "C"
insprate[insprate$Recon==1,]$Recon = "R"

facIR = cast(insprate, perexno ~ Recon)
names(facIR)[names(facIR)=="C"] = "IRC0509"
names(facIR)[names(facIR)=="R"] = "IRR0509"

datacum <- merge(datacum, facIR, all.x=T)

datacum = merge(datacum, facilities, all.x=T)
datacum$impaired = 0
datacum[datacum$S303D == "Y",]$impaired = 1
datacum$wastewat = 0
datacum[datacum$SIC == 4952,]$wastewat = 1
```

```
datacum$elec = 0
datacum[datacum$SIC == 4911,]$elec = 1
datacum$plastic = 0
datacum[datacum$SIC == 2821,]$plastic = 1

temp = subset(datacum, NPR0307==0)
OLS.R3 = lm(IRR0307 ~ E01 + E02 + IE01 + IE02
         + travMin + wastewat + elec + plastic, data=temp)

temp = subset(datacum, NPR0408==0)
OLS.R4 = lm(IRR0408 ~ E02 + E03 + IE02 + IE03
         + travMin + wastewat + elec + plastic, data=temp)

temp = subset(datacum, NPR0509==0)
OLS.R5 = lm(IRR0509 ~ E03 + E04 + IE03 + IE04
          + travMin + wastewat + elec + plastic, data=temp)

temp = subset(datacum, NPC0307==0)
OLS.C3 = lm(IRC0307 ~ E01 + E02 + IE01 + IE02
         + travMin + wastewat  + elec + plastic, data=temp)

temp = subset(datacum, NPC0408==0)
OLS.C4 = lm(IRC0408 ~ E02 + E03 + IE02 + IE03
         + travMin + wastewat  + elec + plastic, data=temp)

temp = subset(datacum, NPC0509==0)
OLS.C5 = lm(IRC0509 ~ E03 + E04 + IE03 + IE04
         + travMin + wastewat + elec + plastic, data=temp)

tmp = mtable(OLS.R3, OLS.R4, OLS.R5, OLS.C3, OLS.C4, OLS.C5)
toLatex(tmp)
```

Test for Mulitcollinearity

```
vif(OLS.R3)
vif(OLS.R4)
vif(OLS.R5)
vif(OLS.C3)
vif(OLS.C4)
vif(OLS.C5)
```

Test for Heteroskedasticity

```
library(lmtests)

ncvTest(OLS.R3)
bptest(OLS.R3$model, data=temp)
ncvTest(OLS.R4)
bptest(OLS.R4$model, data=temp)
ncvTest(OLS.R5)
bptest(OLS.R5$model, data=temp)
ncvTest(OLS.C3)
bptest(OLS.C3$model, data=temp)
ncvTest(OLS.C4)
bptest(OLS.C4$model, data=temp)
ncvTest(OLS.C5)
bptest(OLS.C5$model, data=temp)
```

Probit Models for the Probability of Non-Random Inspections

```
probit.R5  <- glm(NPR0509 ~ IE01 + IE02   + IE03 + IE04
                  + V_bot_2001 + V_bot_2002
                  + V_bot_2003 + V_bot_2004
                  + V_clo_2001 + V_clo_2002
                  + V_clo_2003 + V_clo_2004
                  + V_col_2001 + V_col_2002
                  + V_col_2003 + V_col_2004
                  + V_ddl_2001 + V_ddl_2002
                  + V_ddl_2003 + V_ddl_2004
                  + V_tss_2001 + V_tss_2002
                  + V_tss_2003 + V_tss_2004
                  + travtime
                  + impaired  + noTSS + noChlo + noBOD5
                  , family=binomial(link="probit"), data=datacum)

probit.R4  <- glm(NPR0408 ~ IE01 +  IE02 + IE03
                  + V_bot_2001
                  + V_bot_2002 + V_bot_2003
                  + V_clo_2001
                  + V_clo_2002 + V_clo_2003
                  + V_col_2001 +
                  + V_col_2002 + V_col_2003
```

```
                 + V_ddl_2001
                 + V_ddl_2002 + V_ddl_2003
                 + V_tss_2001
                 + V_tss_2002 + V_tss_2003
                 + travtime
                 + impaired  + noTSS + noChlo + noBOD5
                 , family=binomial(link="probit"), data=datacum)

probit.R3  <- glm(NPR0307 ~ IE01 + IE02
                 + V_bot_2001 + V_bot_2002
                 + V_clo_2001 + V_clo_2002
                 + V_col_2001 + V_col_2002
                 + V_ddl_2001 + V_ddl_2002
                 + V_tss_2001 + V_tss_2002
                 + travtime
                 + impaired  + noTSS + noChlo + noBOD5
                 ,family=binomial(link="probit"), data=datacum)

tmp = mtable(probit.R3, probit.R4, probit.R5)


temp = subset(datacum, wastewat == 1)
probit.C5  <- glm(NPC0509 ~    impaired + perdflw + nonpotw
                 +  pretreat + critical
                 + IE01 + IE02 + IE03 + IE04,
                 family=binomial(link="probit"), data=temp)

probit.C4  <- glm(NPC0408 ~   impaired + perdflw + nonpotw
                 + pretreat + critical
                 + IE01 + IE02 + IE03,
                 family=binomial(link="probit"), data=temp)

probit.C3  <- glm(NPC0307 ~   impaired + perdflw +  nonpotw
                 + pretreat + critical
                 + IE01 + IE02,
                 family=binomial(link="probit"), data=temp)
tmp = mtable(probit.C5, probit.C4, probit.C3)

temp = subset(datacum, wastewat == 0)
probit.C5  <- glm(NPC0509 ~   impaired +  pretreat + noBOD5
                 + IE03 + IE04
                 + MV_2003 + MV_2004
                 + MaxV_2003 + MaxV_2004
                 , family=binomial(link="probit"), data=temp)
```

```
probit.C4  <- glm(NPC0408 ~   impaired +   pretreat + noBOD5
                  + IE02 + IE03
                  + MV_2002 + MV_2003
                  + MaxV_2002 + MaxV_2003
                  , family=binomial(link="probit"), data=temp)

probit.C3  <- glm(NPC0307 ~   impaired +  pretreat   + noBOD5
                  + IE01 + IE02
                  + MV_2001 + MV_2002
                  + MaxV_2001 + MaxV_2002
                  , family=binomial(link="probit"), data=temp)
tmp = mtable(probit.C5, probit.C4, probit.C3)
```

Independence Test for Inspection Types

```
# Independence Test for Reconnaissance / Compliance
library(MASS)
t = table(datacum$NPC0509, datacum$NPR0509)
chisq.test(t)
```

References

Cochran, W. G. (1954). Some methods for strengthening the common χ^2 tests. *Biometrics*, 10(4):pp. 417–451.

Earnhart, D. (2004). Regulatory factors shaping environmental performance at publicly-owned treatment plants. *Journal of Environmental Economics and Management*, 48(1):655 – 681.

Eckert, H. (2004). Inspections, warnings, and compliance: the case of petroleum storage regulation. *Journal of Environmental Economics and Management*, 47(2):232–259.

Eckert, H. and Eckert, A. (2010). The geographic distribution of environmental inspections. *Journal of Regulatory Economics*, 37:1–22. 10.1007/s11149-009-9101-8.

Gray, W. B. and Deily, M. E. (1996). Compliance and enforcement: Air pollution regulation in the U.S. steel industry. *Journal of Environmental Economics and Management*, 31(1):96–111.

Harrington, W. (1988). Enforcement leverage when penalties are restricted. *Journal of Public Economics*, 37(1):29–53.

Helland, E. (1998). The enforcement of pollution control laws: Inspections, violations, and self-reporting. *The Review of Economics and Statistics*, 80(1):141–153.

Laplante, B. and Rilstone, P. (1996). Environmental inspections and emissions of the pulp and paper industry in Quebec. *Journal of Environmental Economics and Management*, 31(1):19–36.

Magat, W. A. and Viscusi, W. K. (1990). Effectiveness of the EPA's regulatory enforcement: The case of industrial effluent standards. *Journal of Law and Economics*, 33(2):331–60.

Nadeau, L. W. (1997). EPA effectiveness at reducing the duration of plant-level noncompliance. *Journal of Environmental Economics and Management*, 34(1):54–78.

Rousseau, S. (2007). Timing of environmental inspections: survival of the compliant. *Journal of Regulatory Economics*, 32:17–36. 10.1007/s11149-006-9013-9.

Shimshack, J. P. and Ward, M. B. (2008). Enforcement and over-compliance. *Journal of Environmental Economics and Management*, 55(1):90–105.

Wortman, M. A. and Klutke, G.-A. (1994). On maintained systems operating in a random environment. *Journal of Applied Probability*, 31(2):pp. 589–594.