UNIVERSITÀ DEGLI STUDI DI SIENA

QUADERNI DEL DIPARTIMENTO
DI ECONOMIA POLITICA E STATISTICA

Marcello Basili
Filippo Belloc
Simona Benedettini
Antonio Nicita

Warning, Learning and Compliance: Evidence from Micro-data on Driving Behavior

n. 639 – Maggio 2012

Abstract - In many contexts, warning systems of law enforcement are used to let uninformed individuals learn what is illegal, while sanctions are applied only after a number of repeated violations. Surprisingly no em- pirical evidence is available so far, over the learning impact of warnings. This paper is a first attempt to empirically investigate the warning's effect on individuals' behavior employing a unique database on a traf- fic law enforcement system, which constitutes an extraordinary nat- ural laboratory to test whether experience warning induces learning. Specifically, we use six-year longitudinal data on about 50000 drivers under the Italian point-record system of traffic law. Our statistical re- sults show that warned drivers become more compliant. To the extent individuals learn through their repeated behavior, a warning system makes it possible to apply sanctions only to (presumably) informed violators.

Keywords: warning, law enforcement, mixture models.
JEL classification: K42, C14.

Acknowledgments: We would like to thank Bruno Deffains, Nuno Garoupa, Fernando Gomez, Nicola Persico, and Mitchell Polinsky for helpful comments. Usual disclaimers apply.

Marcello Basili, DEPFID, Università degli Studi di Siena, marcello.basili@unisi.it
Filippo Belloc, Department of Economics, University "G.d'Annunzio" of Chieti-Pescara, f.belloc@unich.it
Simona Benedettini, IEFE, Bocconi University, simona.benedettini@unibocconi.it
Antonio Nicita, DEPS, Università degli Studi di Siena, antonio.nicita@unisi.it

1 Introduction

Illegal behaviors are likely to differ across individuals depending on hidden characteristics of the violator, such as his information about the law or the wealth he has available to pay fines (Kaplow, 1990; Levitt, 1997; Polinsky, 2006). In these cases, it is efficient for the authority to implement enforcement devices that distinguish the types of offenders (Kaplow, 1990). In particular, when individuals have different information on the illegality of their possible actions, warning systems of law enforcement are often used in order to differentiate informed from uninformed offenders, and to reduce over-compliance of uniformed ones. A first-time offender is warned that his action was illegal without being sanctioned, while the sanction is applied after a certain number of repeated violations. Hence, the uninformed individual is allowed to learn what is illegal and he is punished only once (presumably) he is informed.

However, there is not a strong theoretical prediction that under a warning system individuals do learn. First, individuals might ignore the warning signal depending of unobservable personal factors. Second, violators might actually have perfect information ex-ante on the illegality of their actions, so that simply there is nothing to learn. If learning is absent, then a warning system only introduces under-deterrence, without reducing at the same time the social costs of over-compliance by imperfectly informed individuals. The absence of theoretical predictions on the learning effect of warning makes the study of warning-induced compliance an empirical issue. In this paper we study empirically how issuing warning affects individuals' behavior using six-year longitudinal data on 48154 drivers under the Demerit Point System (DPS) of traffic law enforcement introduced in Italy in 2003.

A point-record system in traffic law enforcement is an extraordinary natural laboratory to test whether experiencing warning induces learning. Indeed, it is unlikely that drivers have complete information on which are all possible violations of traffic law and warning can actually provide additional information to uninformed drivers. For instance, in Italy 39 percent of total violations are infringements of articles 141 and 142 of the Italian traffic law code concerning speed. In particular, article 141 says, among other things, that the driver must adjust the speed to the visibility conditions, "*especially when the road is steep, when the pedestrians on the road show uncertainty, and when animals on the road appear to get scared*". This prescription is very vague and people are unlikely to know its exact meaning in practice ex-ante (i.e. before being sanctioned). Similarly, drivers are unlikely to be able to precisely measure ex-ante what is a safety distance "*that allows stopping the vehicle timely*" (see article 149 of the Italian traffic law code). Again, drivers are unlikely to be able to precisely understand ex-ante if they have drunk alcohol for more than 0.5 gram of alcohol per liter of blood, that is the limit permitted by law (see article 186 of the Italian traffic law code).

The Italian DPS assigns an initial given amount of points to drivers (twenty points). Violations of traffic law are admonished with a points curtailment, that ranges from one to ten depending on the seriousness of the violation (points curtailments are generally coupled also with a financial penalty). Therefore, each points curtailment can be considered as a seriousness-weighted warning. Driving license suspension applies only once the endowment of points is exhausted, not before, so that drivers are allowed to infringe traffic law some times before incurring the non-monetary sanction. For example, on average the non-monetary sanction is delayed after twenty violations for improper use of lights, after ten violations for infringement of traffic signals, and after two violations for drunk driving.

We employ a semi-parametric latent variable model for dichotomous responses and find that, on average, the higher the amount of residual points a driver holds - i.e. the lower the number of (weighted) warnings he received -, the higher his individual-specific probability of committing an additional traffic violation. We interpret this result arguing that the number of warnings an uninformed driver receives through his repeated behavior induces learning. Otherwise, if drivers were ex-ante perfectly informed, we should have observed that a driver's probability to commit an infraction is unaffected by the number of points curtailments the driver has previously experienced. Furthermore, in a robustness check of our empirical study we rule out the possibility that deterrence (i.e. drivers who lose points become increasingly more compliant because are deterred by the threat of losing their license), if present, drives our findings.

To the best of our knowledge, this paper is the first providing empirical evidence on a continuous reduction of law violation probability in the number of warnings. The only two available papers on warning (Nyborg and Telle, 2004; Rousseau, 2009) discuss a theoretical model in which the practice of issuing warning is used by a law enforcement authority in the presence of imperfect information. In particular, Nyborg and Telle (2004) explain that a warning system may be used to avoid sanctioning accidental violators and to punish only recidivists. However, whether warnings actually induce increasing compliance of (presumably) accidental violators so far has remained an empirically unanswered question.

The remaining of the paper is as follows: in section 2 we briefly present the related literature, in section 3 we describe data and empirical strategy, in section 4 we discuss the estimation results obtained both from a basic model specification and from group-specific regressions, section 5 concludes.

2 Related literature

The literature dealing with the economic rationale for tolerating non-compliance of first time violators is broad. Nevertheless, until recently, this literature has ignored the practice of issuing warning before prosecution.

Bose (1995) shows that the optimal penalty should be non-maximal when the authority may mistakenly prosecute compliant individuals. Similarly, Polinsky and Shavell (2000) argue that the presence of type-II errors justifies lower sanctions for first time offenders, while repeated offenders should be punished more harshly.[1] Heyes and Rickman (1999) suggest that tolerating non-compliance may be rational when the enforcement authority and the regulated agent interact in more than one domain, and non-compliance in one domain can be accepted by the enforcer in exchange for compliance in another domain. Garvie and Keeler (1994), finally, argue that incomplete enforcement can be admitted when the authority faces a trade-off between monitoring and sanctioning. All these studies never mention the possibility for the enforcement authority to issue warning.

Notable exceptions in this body of literature are Nyborg and Telle (2004) and Rousseau (2009). Rousseau (2009) argues that an economic rationale for issuing warning is to reduce over-compliance of economic agents. In the model of Rousseau, the enforcement authority makes mistakes in evaluating agents' compliance. These errors are associated to social costs, because incorrect sanctioning decreases the trust people put in the legal system. In addition, detection errors induce over-compliance to the extent that economic agents fear the risk of being erroneously sanctioned. Therefore, to issue warning reduces both the social costs and over-compliance, at the price of introducing some under-deterrence. If the costs of convicting compliant individuals are sufficiently large, then the use of warning can increase welfare. In a previous paper, Nyborg and Telle (2004) propose a more general model in which economic agents may make mistakes and, even if they decide to comply, may violate due to errors. Their model focuses on pollution permit regulation, but it can be applied also to other cases. The crux argument of Nyborg and Telle is twofold. First, the enforcement authority does not make mistakes (or, if it does, mistakes are negligible). Second, prosecution is costly and the law enforcement authority has a limited budget, hence - when the number of violators exceeds a critical level - the threat of punishment is not credible and becomes insufficient to effectively deter violation. As a result, to prosecute only those who fail to comply after receipt of the warning reduces the number of prosecuted violators and therefore the probability that the authority loses control (i.e. that the system shifts to a low compliance Nash equilibrium). A crucial condition in the model of Nyborg and Telle is that prosecution costs rather monitoring costs are the most significant. Thus, after first time violators are detected, issuing warning without sanctioning allows the authority to respect its budget constraint and to make a credible threat against recidivists.

In the Nyborg and Telle's (2004) model, the practice of issuing warnings

[1] See also Polinsky and Rubinfield (1991).

prior to prosecution is interpreted as a device to increase compliance, or, phrased differently, to discriminate systematic violators from the rest. An implicit assumption is that, after being warned, people reduce their accidental violations. Our research objective is to empirically investigate if warning does induce lower law infringements actually. In many regulated environments, such as traffic regulation, individuals have an imperfect ability to precisely evaluate whether their actions are conform with what the law prescribes, typically because they have imperfect information about every law prescription or because their ability to drive is imperfect. How do people react to warning in an environment of such type?

3 Design of the empirical study

3.1 Hypothesis to be tested

Consider a system of law enforcement in a N-periods setting. In each period, an individual commits a fixed amount of actions that, for the sake of simplicity, we normalize to 1. Each action can be performed in a legal or illegal manner. Let us call u the benefit the individual gets from committing a legal act, and U the benefit the individual gets from committing an illegal act (with U and u both positive, and $U > u$). In each t_n period (with $n = 1, ..., N$) a detected law violation is punished with a monetary sanction S_m (with $S_m > 0$). In the t_N period, the illegal act is punished with the monetary sanction plus a non-monetary sanction S_{nm} (with $S_{nm} > 0$).[2] δ is the detection probability. We assume that $\delta \in [0, \overline{\delta}]$, which implies that an individual can be detected with a probability no greater than some $\overline{\delta} < 1$.[3] Thus, δS_m is the expected monetary sanction (ES_m) and δS_{nm} is the expected non-monetary sanction (ES_{nm}). We normalize the population to 1 and suppose that it is composed by three groups ε, λ and $1-\varepsilon-\lambda$, such that:

(i) For the population fraction ε:

$$\begin{cases} (U_\varepsilon - u_\varepsilon) > ES_m \\ (U_\varepsilon - u_\varepsilon) > (ES_m + ES_{nm}) \end{cases}$$

(ii) For the population fraction λ:

$$\begin{cases} (U_\lambda - u_\lambda) > ES_m \\ (U_\lambda - u_\lambda) < (ES_m + ES_{nm}) \end{cases}$$

[2] See Shavell (1987) for a theory of non-monetary sanction when courts have imperfect information.

[3] For the sake of simplicity, we assume that individuals formulate an expectation on δ that coincides with the true value of δ, over the N periods.

(iii) For the population fraction $1 - \varepsilon - \lambda$:

$$\begin{cases} (U_{1-\varepsilon-\lambda} - u_{1-\varepsilon-\lambda}) < ES_m \\ (U_{1-\varepsilon-\lambda} - u_{1-\varepsilon-\lambda}) < (ES_m + ES_{nm}) \end{cases}$$

Suppose that in the first period t_1 the average individual has imperfect information about the illegality of his possible actions, i.e. he believes that an illegal action is legal with a probability α (with $0 < \alpha < 1$). Hence, α can be considered an exogenous (initial) degree of imperfect information of the population.

We can now calculate the probability that an individual commits a law violation. For ε-type individuals, it is straightforward that the probability π_ε of committing an illegal act is equal to 1 in each time period from t_1 to t_N. Indeed, ε-type individuals always get a benefit from the illegal act that is higher than the expected sanction, whether they are informed or not informed about the illegality of their actions. Similarly, λ-type individuals show a probability π_λ of committing an illegal act that is equal to 1 in each time period from t_1 to t_{N-1}. In the last period t_N, however, they expect a sanction from committing a law violation that is higher than the benefit they get from the illegal act; thus, in t_N λ-type individuals commit law violations only to the extent they are imperfectly informed about the illegality of their action. Since individuals get informed through detection, it can be showed that in t_N the probability that an action by λ-type individuals is illegal is $(1-\delta)^{t_{N-1}}$. Finally, individuals of $(1-\varepsilon-\lambda)$-type show a probability of law violation $\pi_{1-\varepsilon-\lambda}$ that is decreasing in time. Indeed, at t_1 the $(1-\varepsilon-\lambda)$-type individual commits an illegal action with probability α, i.e. he commits an illegal act only if he believes that the given action is legal. Therefore in t_1 he experiences a probability of incurring in detection (and of getting a monetary sanction) equal to $\alpha\delta$. At the end of period t_1 the individual has learnt that a fraction $\alpha\delta$ of his actions was illegal. At t_2 the $(1-\varepsilon-\lambda)$-type individual commits an illegal acts with a probability $\alpha - \alpha\delta$, i.e. the probability of committing an illegal action faced by the individual at t_1 reduced by the learning. Now, the probability of incurring in detection becomes $(\alpha - \alpha\delta)\delta$. Analogously, at t_3 the individual commits an illegal action with probability $\alpha - \alpha\delta - [(\alpha - \alpha\delta)\delta]$ and learns by $\{\alpha - \alpha\delta - [(\alpha - \alpha\delta)\delta]\}\delta$, and so on. More generally, rearranging terms we can sum up these results as follows:

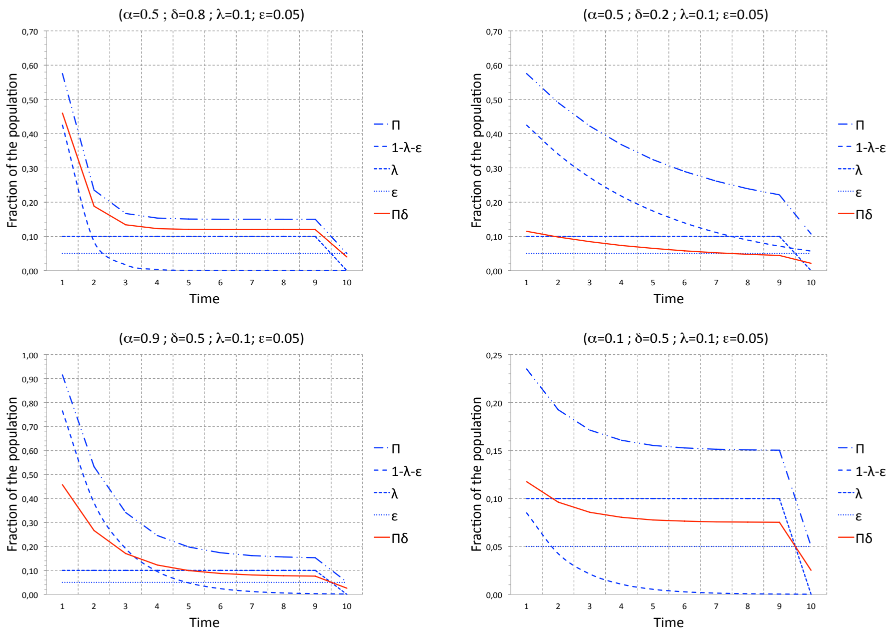
$$\begin{cases} \pi_\varepsilon = 1 & \forall t \in [t_1, t_N] \\ \pi_\lambda = 1 & \forall t \in [t_1, t_{N-1}] \\ \pi_\lambda = (1-\delta)^{t_{N-1}} & \text{if } t_n = t_N \\ \pi_{1-\varepsilon-\lambda} = \alpha(1-\delta)^{t_{N-1}} & \forall t \in [t_1, t_N] \end{cases}$$

Given that the population is normalized to 1, in each time period t_n the fraction of the population committing an illegal action coincides with the probability that a generic (average) individual commits a law violation, i.e.:

$$\begin{cases} \Pi_n = \varepsilon + \lambda + (1-\varepsilon-\lambda)\alpha(1-\delta)^{t_n - 1} & \forall t \in [t_1, t_{N-1}] \\ \Pi_n = \varepsilon + \lambda(1-\delta)^{t_n-1} + (1-\varepsilon-\lambda)\alpha(1-\delta)^{t_n-1} & \text{if } t_n = t_N \end{cases}$$

It is intuitive that the probability of observing a law violation for the average individual is $\delta\Pi_n$. The enforcer decides the number of N periods after which the individual can be reasonably deemed to be informed about the illegality of the actions he commits. That is, after N periods the enforcer can assume that Π_N is sufficiently close to ε so that the illegal actions committed (and detected) are only ε-type law violations. It is straightforward that, when $\alpha = 0$ (i.e. perfect information), in each period from t_1 to t_{N-1} all the committed (and detected) law violations are ε-type or λ-type law violations and in t_N all the committed (and detected) law violations are ε-type law violations, i.e. $\Pi_n = \epsilon + \lambda$ ($\forall n \in [1, N-1]$) and $\Pi_n = \varepsilon$ (for $n = N$). In this case, learning is absent and the probability that the average individual commits an illegal act is constant across time periods until the $N-1$ period. Hence, when $\alpha = 0$ a warning system only introduces under-deterrence. To clarify, let us show in Figure 1 a simulation that can be obtained for different values of δ and α.

Figure 1: Simulation of law violations in the population ($N = 10$).



From an empirical point of view, each t_n corresponds to a detection instant, in which one or more warnings w_n are issued. What we want to measure is whether Π_n is decreasing in the number of warnings at the average individual level. We do not directly observe Π, while we observe $\delta\Pi$. Hence, if we find that $(\partial\delta\Pi/\partial w) < 0$, then this impies that $\alpha > 0$ and warning induced learning holds.[4]

3.2 Description of the database

In order to empirically investigate the effect of warning on law violation probability, we use data on the point-record system in traffic law enforcement introduced in Italy in 2003. This is a unique natural laboratory to test whether experience warning induces learning, because it offers the opportunity to measure changes in infraction probability at an individual level due to the cumulate number of warnings the single driver has experienced in the past. In the Italian traffic law enforcement system (DPS) an initial given amount of twenty points is assigned to drivers. When the driver violates traffic law, and he is detected, he is admonished with a points curtailment coupled with a financial penalty. Since the point curtailment changes in size depending on the seriousness of the violation, each points curtailment can be considered as a 'weighted' warning (e.g. the infringement of traffic signals is admonished with a two-point curtailment, while drunk driving - that is a much more serious action - can be punished with a ten-point curtailment). The non-monetary sanction, in the form of driving license suspension, is then applied only once the driver exhausts his endowment of points.

The source of data for the empirical study is the database that the Italian Traffic Police maintains on the demerit points changes showed by all the Italian drive licenses currently holding in Italy. The data we use include the demerit points changes of a representative sample of 50000 drivers observed over the period between July 1st - 2003 and May 20th - 2009, picked from among all drivers included in the Italian Traffic Police's database on May 20th - 2009. Given that July 1st - 2003 is the date in which the DPS came into force in Italy, all the drivers included in our sample have the same initial endowment of points (i.e. twenty points). Each individual record of the dataset refers to a single points change of an individual driver, i.e. a points curtailment for infraction or a point credit for not committing infraction for at least two consecutive years. Moreover, for each single driver the data contain information on: sex, age, nationality, place of residence, type of drive license held, number and type of the infractions committed in the past, and number of residual points. Hence, the dataset allows us to observe a driver's

[4] Under moral hazard, drivers who lose points may assign increasing value to the residual points and become increasingly more compliant (i.e. drivers are deterred by the threat of losing their license). In a robustness check of our empirical study we rule out the possibility that deterrence, if present, drives our findings.

characteristics in correspondence of both an illegal behavior and normal (i.e. legal) driving, on a yearly basis.

In order to carry out the empirical analysis on drivers with fully available information, we drop out of the sample those drivers who show at least one incomplete record (due to missing data) for one or more variables. We drop out of the sample also those infractions committed while using a drive license of a type different from type A (which covers all types of motorcycle), type B (which covers all types of motor vehicle for the transport of people - up to 9 individuals - and cargo with a weight up to 3.5 tons) or type C (which covers all types of motor vehicle covered by the type B, plus cargo vehicles with a weight higher than 3.5 tons). The final sample used in our study is made up of 48154 drivers and 154346 points changes. Only 10.11 percent of the points changes are negative changes (which are due to traffic law violation). Most of the points changes are showed by Italians (about 96 percent) and males (about 58 percent). The average age on a points change is 42.76 years. The average residual points that drivers show at the moment of a points change is 19.75. About 48 percent of points changes involve drivers residing in the northern Italy, about 22 percent those residing in the central Italy, and about 30 percent those residing in the southern Italy and in the main islands. Of those who commit at least one infraction (10782 drivers in our sample) 95 percent are Italians and 70 percent are males. Most of the traffic law violations (88.7 percent) are committed by drivers using a type B drive license. Table 1 provides summary statistics and a description of the main variables used in the empirical analysis.

Table 1: Variables' description and summary statistics.

Variable	**Description**	**Mean**	**Std.Dev.**
ResidualPoints	Residual points held at the moment of infraction.	19.75	4.42
Age	Age of the driver.	42.76	15.13
Sex	Sex of the driver (1 = female, 0 = otherwise).	0.43	0.50
Nationality	Nationality of the driver (1 = Italian, 0 = otherwise).	0.96	0.19
PastOffenses	Cumulate number of infractions committed in the past.	0.16	0.57
DriveCourse	The driver attended a driving course for avoiding the suspension of the drive license after the zeroing of the residual points (1 = the last points change is due to attending the course, 0 = otherwise).	0.00	0.04
LicenseType	Set of three dummies for the type of the drive license used at the moment of infraction:		
	1 = "Type A", 0 = otherwise,	0.00	0.05
	1 = "Type B", 0 = otherwise,	0.93	0.25
	1 = "Type C", 0 = otherwise.	0.06	0.24

3.3 Econometric model

We employ a latent variable model for dichotomous responses. Our model specifies the conditional expectation of the response variable given observed covariates and latent unobserved explanatory variables. Data at an individual driver level are strongly affected by hidden heterogeneity (for instance, drivers may show a different risk aversion, may use the vehicle for different purposes, and so on). In our model, the latent variables are accounted for by a set of individual-specific random effects. The conditional distribution of the response given the observed covariates and random effects is then specified via a binomial family and a logit link function.

We use a dummy variable as the measure for committing infraction (i.e. our response variable). We construct a dummy variable which equals 1 when the driver shows a points curtailment for infraction and equals 0 otherwise. Let us call Y_{it} the binary outcome for the i-th individual at time t (with $i = 1, ..., n$ and $t = 1, ..., T$), $\mathbf{x}_{it}$ a vector of p covariates (i.e., $\mathbf{x}_{it} = x_{it1}, ..., x_{itp}$), and u_i the set of individual-specific random effects which account for cross-individual heterogeneity. The outcome variable can be specified as follows:

$$Y_{it}|\mathbf{x}_{it}, u_i \sim \text{Bin}(1,\pi_{it}), \qquad \text{logit}(\pi_{it}) = \beta_0 + \sum_{l=1}^{p} \mathbf{x}_{it}\beta_l + u_i \tag{1}$$

where $\beta = \beta_0, ..., \beta_p$ is a vector of regression parameters.

Generally, latent variable models are based on the assumption that the random effects u_i have a parametric distribution $\mathcal{G}(u_i)$ known ex-ante (e.g. Gaussian distribution), and require the adoption of numerical integration techniques in order to obtain maximum likelihood (ML) estimates (e.g. adaptive Gaussian Quadrature). In an empirical study of individuals' behavior, however, to assume a priori that unobservable effects have a specific known distribution is unnecessarily restrictive, as modern econometrics makes relaxing such an assumption computationally simple. In our model, we leave $\mathcal{G}(u_i)$ unspecified and use non-parametric maximum likelihood (NPML). Doing so, we do not make any assumption on the distribution of the unobserved factors across individuals, and obtain a flexible model structure (Heckman and Singer, 1984). We allow, moreover, the unobserved factors' distribution to be a finite mixture distribution that reflects a possible group-structure of the data, that can be due to unobservable clusters of drivers (for example on the basis of risk aversion); this makes our model a mixture model. More formally, we maximize the likelihood function by employing a discrete distribution $\mathcal{G}_K(u_i)$, with $K \leq n$ support points. Describing the discrete distribution by p_k masses (with $k = 1, ..., K$) on $\mathbf{u}_k$ locations, i.e. $p_k = \Pr(u_k)$, the likelihood function can be written as follows:

$$L(\cdot) = \prod_{i=1}^{n} \left\{ \sum_{k=1}^{K} [f(y_{it} \mid \mathbf{x}_{it}, \mathbf{u}_k)] p_k \right\} \tag{2}$$

We carry out the maximum-likelihood estimation by employing the expectation-maximization (EM) algorithm for finite mixtures, which is well suited when the likelihood function involves latent variables (Dempster et al., 1977). The EM algorithm is based on an iterative process in which an expectation step (E-step) and a maximization step (M-step) are alternated until the likelihood converges. In the E-step, the expected likelihood is obtained with respect to the computed conditional distribution of the latent variables, given the current parameters' setting (the parameters' values are randomly assigned in the first iteration) and the observed data. In the M-step, all the parameters are re-estimated by maximizing the expected likelihood obtained in the E-step. Once new parameters' values are generated in the M-step, the E-step is repeated, another expected likelihood is hence obtained, and so on. This process continues until a local maximum is reached (i.e. the likelihood converges).

Making the covariates explicit, the operative model that we estimate can be written in its baseline form as follows:

$$\begin{aligned} g\left(\mathrm{E}\left[y_{it} \mid \mathbf{x}_{it}, u_i\right]\right) = \mathrm{logit}(\pi_{it}) = \beta_0 &+ \beta_1 ResidualPoints_{it} + \beta_2 Sex_i + \\ &+ \beta_3 Nationality_i + \beta_4 Age_{it} + \beta_5 PastInfractions_{it} + \\ &+ \beta_6 DriveCourse_{it} + \beta_{7,\ldots,9} LicenseType_{it} + \\ &+ \beta_{10} DetectionTrend_{it} + \beta_{11,\ldots,30} Region_i + \\ &+ \beta_{31,\ldots,37} Year_t + \mathcal{G}(u_i) \end{aligned} \tag{3}$$

where some explanatory variables ($LicenceType$, $Year$, and $Region$) are included as a set of dummies. Notice that some explanatory variables vary both across individuals and time (Age, $PastInfractions$, $DriveCourse$, $LicenceType$, and $ResidualPoints$), while some others vary only across individuals and are time constant (Sex, $Nationality$, and $Region$) or vary across years and are constant across individuals ($Year$). Variables' description is provided in Table 1. $Region$ refers to the place of residence of the driver, $Year$ refers to the date of the points change event. Notice that we have included a detection trend in equation (3): given that it is reasonable to assume that drivers face a positive detection probability δ lower than 1 (i.e. $0 < \delta < 1$), the probability of incurring in sequential detections decreases in the number of actual detections converging to zero. In order to account for this non-linear decreasing trend of detection probability we include a $DetectionTrend$ variable built as δ^{t_n} (where t_n indicates the cumulate number of past detected infractions of the driver). This is a crucial issue for identification. We calculate an initial detection probability δ using a proxy variable that measures the number of control units (i.e. number of traffic policemen, speed detectors, and traffic-light detectors) per squared kilometre of urban area at a province level, normalized to 1. We have calculated the size of provinces' urban area using a Geographic Information System on satellite data. Data on control units are provided by the Italian Traffic Police.

4 Empirical evidence

4.1 Basic results

Table 2 presents the results obtained from the estimation of model (3). The first column lists the variables, the second column reports the odds ratios, while the remaining columns present the estimated coefficients, standard errors, and p-values.

Our econometric model provides an estimated coefficient of the variable $ResidualPoints$ positive and statistically significant at 1 percent level. In particular, to experience one-point curtailment less than the average is associated to an increased probability of violation equal to a 1.29 odds ratio. This result is an evidence in favor of our learning argument, as drivers are shown to reduce progressively their infraction risk before reaching the last punishable violation. We also obtain further interesting results. To be female (Sex) exerts a negative and statistically significant effect on the probability of committing infraction. The driver's age (Age) has a positive and statistically significant coefficient (with p-value = 0.000). Interestingly, the cumulate number of past infractions ($PastInfractions$) is positively associated in a statistically significant way to a higher probability of infraction (with p-value = 0.000), and this suggests the existence of a recidivism effect for those individuals (presumably informed) that are not influenced by repeated warning. Consistently, we also find that the variable $DriveCourse$ is associated to a positive and statistically significant parameter (with p-value = 0.000), i.e. those recidivist drivers which attended a driving course for avoiding the suspension of the driving license after the zeroing of the residual points tend to show an increased probability of violating traffic law. Italian nationality of drivers ($Nationality$) has a positive and statistically significant effect on infraction probability. Finally, as expected, $DetectionTrend$ is shown to be associated to a negative a strongly statistically significant estimated parameter.

Table 2: Basic estimation results (event = committing infraction). Note, statistical significance: "**" 1 percent level, "*" 5 percent level. NPML estimation of model (3). EM algorithm for finite mixtures is employed. Individual effects are random. 6 iterations before convergence (log-likelihood: -33404.149).

Variable	**Odds ratio**	**Coeff.**	**Std.Err.**	***p*-value**
ResidualPoints	1.291**	0.256**	0.004	0.000
Sex (being female)	0.542**	-0.611**	0.022	0.000
Nationality (being Italian)	1.393**	0.332**	0.048	0.000
Age	1.005**	0.005**	0.001	0.000
PastInfractions	2.341**	0.851**	0.020	0.000
DriveCourse	2.980**	1.092**	0.178	0.000
License (type A)		benchmark		
License (type B)	2.363*	0.860*	0.399	0.031
License (type C)	2.823*	1.038*	0.400	0.010
DetectionTrend	3.4e-10**	-21.798**	0.841	0.000
Constant	0.006**	-4.968**	0.410	0.000
19 regional dummies		yes		
6 year dummies		yes		
Individual effects		yes		

4.2 Robustness check against deterrence

Our findings could be subject to a caveat concerning an identification issue. Suppose that drivers are perfectly informed about the illegality of their possible actions (i.e. $\alpha = 0$). In this case, while in each t_n ($\forall t \in [t_1, t_{N-1}]$) ε-type individuals always commit infractions and $(1-\varepsilon-\lambda)$-type individuals never commit infractions, the careful driving effort exerted by a rational individual of type λ is maximum only when the driver is at risk of losing his license, that is when an additional law violation implies the zeroing of the residual points. This causes a drop in the infraction probability for λ-type individuals after the $N-1$ detection instant. Formally, we will have $\Pi_n = \varepsilon + \lambda$ ($\forall t \in [t_1, t_{N-1}]$) and $\Pi_n = \varepsilon$ (if $t_n = t_N$). Hence, if we linearly estimate the population average effect of a point change between the maximum ($\overline{R}$) and minimum ($\underline{R}$) non-null endowment of points shown by drivers, even if $\alpha = 0$ and learning is absent we could still observe an increasing law violation probability in the number of points. In order to circumvent this problem and to isolate the effect of learning, if any, we restrict our estimation analysis to the behavior of drivers showing an amount of residual points ranging from $\overline{R}$ and R^*, where R^* is an amount of points such that an additional traffic law violation (whatever its seriousness is) does not imply driving license suspension. Since in the Italian DPS the largest points curtailment for a single infraction is 10 points (for example this is the case for drunk driving or for exceeding speed limit by more than 40 km/h), we set R^* to 11. Estimation results so obtained are collected in Table 3. The first column lists the variables, the second column reports the odds ratios, while the remaining columns present the estimated coefficients, standard errors, and p-values.

Table 3: Robustness check: estimation results (event = committing infraction). Note, statistical significance: "**" 1 percent level, "*" 5 percent level. NPML estimation of model (3). EM algorithm for finite mixtures is employed. Individual effects are random. 5 iterations before convergence (log-likelihood: -29969.45).

Variable	**Odds ratio**	**Coeff.**	**Std.Err.**	**p-value**
ResidualPoints $[\overline{R}, R^*]$	1.440**	0.365**	0.008	0.000
Sex (being female)	0.556**	-0.586**	0.023	0.000
Nationality (being Italian)	1.362**	0.309**	0.050	0.000
Age	1.004**	0.004**	0.001	0.000
PastInfractions	2.462**	0.901**	0.029	0.000
DriveCourse	3.434**	1.234**	0.202	0.000
License (type A)		benchmark		
License (type B)	2.778*	1.022*	0.434	0.019
License (type C)	3.346*	1.208*	0.436	0.006
DetectionTrend	7.2e-11**	-23.350**	0.887	0.000
Constant	0.001**	-9.057**	0.483	0.000
19 regional dummies		yes		
6 year dummies		yes		
Individual effects		yes		

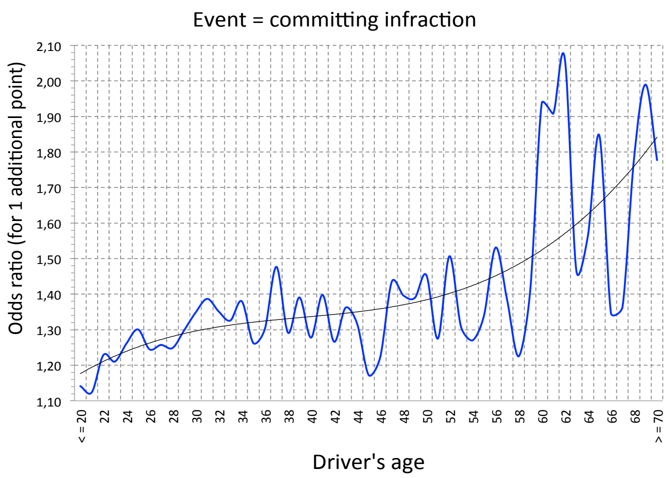
Again, we find that the estimated coefficient of the variable *ResidualPoints* (now restricted between $\overline{R}$ and R^*) is positive and statistically significant at 1 percent level. In particular, to experience one-point curtailment - between $\overline{R}$ and R^* - less than the average is associated to an increased probability of violation equal to a 1.44 odds ratio. Notice that between $\overline{R}$ and R^* the risk of driving license suspension remains null for any points curtailment due to infraction. Therefore, the reducing probability of violation in the reduction of points can be associated only to learning effects. Moreover, the estimated parameter of the variable *ResidualPoints* turns out larger between $\overline{R}$ and R^* than between $\overline{R}$ and $\underline{R}$ (as in the basic model results presented in Table 2). This is an important result that rules out the possibility that deterrence drives our findings. Under moral hazard, drivers who lose points become increasingly more compliant because they are at threat of losing their license (deterrence effect), i.e. they do not care about each lost point equally but assign increasing value to residual points as they decrease (Dionne et al., 2011). In this case, the effect of an additional point on infraction probability should be concave in the number of residual points. Empirically, then we should observe an estimated parameter of the variable *ResidualPoints* lower (larger) between $\overline{R}$ and R^* (between and R^* and $\underline{R}$) than between $\overline{R}$ and $\underline{R}$. Instead, we find evidence of a convex curve for the residual points parameter. Therefore, deterrence effects, if any, do not outweigh learning effects.

Control variables' estimated parameters remain with the same sign and statistical significance as those obtained in the basic model estimation.

4.3 Sub-sampling

We have estimated model (3) also on group-specific sub-samples picked from among the population of drivers. Specifically, we have estimated the *ResidualPoints*'s parameter at different driver's ages, and distinguishing female and male, Italian and non-Italian, and drivers located in different areas of the country (North, Center, South). The estimated equation includes all the control covariates of model (3), and excludes the variable(s) identifying the sub-sample group-specific estimation. Odds ratios (statistically significant at 1 percent level) are reported graphically in Figure 2 and Figure 3. An odds ratio measures the effect size of a one-point increase on the infraction probability. It therefore gives a measure of the intensity at which individuals react to a one-point warning, i.e. how much fast they learn.

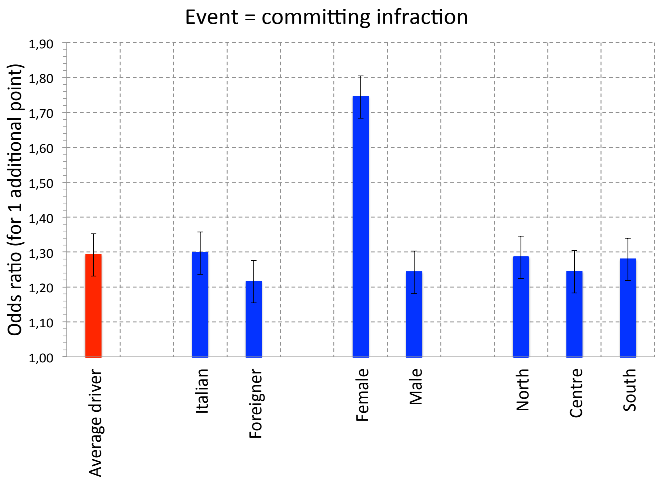
Figure 2: Odds ratios (event = committing infraction) for 1 additional point at different driver's ages.



We find that the representative average drivers react less intensively to warning when they are relatively younger. A driver with age in between 18-21 years, on average, has an odds ratio lower than 1.20 (p-value: 0.000), while a driver with age in between 22-59 years has an average odds ratio equal to 1.32, and drivers over 60 years show an odds ratio of about 1.73. Comparing the average odds ratio for the 18-21 age class with that for the average representative driver (over the whole population), we obtain that the intensity of the effect of one-point warning on infraction probability of 18-21 years old drivers is 87.68 percent of the average effect observed in the whole population. This result has an important policy implication, because it suggests a need of increasing the warning magnitude for younger drivers.

For example, traffic law could admonish young violators (e.g. 18-21 years old) with an increased-size warning with respect to adult individuals.

Figure 3: Odds ratios (event = committing infraction) for 1 additional point per different driver's characteristics.



Group-specific results also unveil that being Italian or foreigner and the geographical area of residence do not affect substantially traffic violators' behavior. Specifically the odds ratio for Italian is 0.296 (p-value: 0.000) and that of non-Italian is 0.215 (p-value: 0.000). The odds ratio for drivers in the North of Italy is 1.285 (p-value: 0.000), while the odds ratio is 1.243 (p-value: 0.000) and 1.278 (p-value: 0.000), respectively, for drivers in the Centre and in the South. Differently, sex appears a significant factor, female's odds ratios being systematically higher than male ones (1.743 with p-value 0.000 against 1.242 with p-value 0.000).[5]

4.4 Conclusions

Point-record systems in traffic law enforcement have been analyzed both theoretically and empirically from various perspectives. For instance, Bourgeon and Picard (2007) develop a theoretical model that explains how point-record licenses can be used as a device to selectively incapacitate reckless individuals who are not deterred by monetary fines. More recently, Dionne et al. (2011) use Quebec's data on point-record traffic law enforcement to calculate monetary fine equivalents for demerit point accumulation. In this paper we use data on point-record licenses and focus on the warning function

[5]The tables of results are available upon request.

of demerit point systems. Specifically, we perform a rigorous econometric investigation of six-year longitudinal data on Italian drivers, and try to address the so-far unanswered question of whether to issue warnings induces learning in the population of violators. Our statistical results show that warned traffic violations cause more compliant driving. We also find that learning is less intense for young individuals. These findings contribute to the theoretical economic literature available on warning (Nyborg and Telle, 2004; Rousseau, 2009). Moreover, they have relevant policy implications both for road safety regulation and the general law enforcement. When some individuals are imperfectly informed about the illegality of their actions and the authority lacks sufficient resources to distinguish informed and uninformed individuals, a warning system makes it possible to apply sanctions only to presumably informed violators.

References

[1] Bose, P. (1995) Regulatory errors, optimal fines and the level of compliance. *Journal of Public Economics* 56(3): 475-484.

[2] Bourgeon, J.M. and Picard, P. (2007) Point-record driving license and road safety: an economic approach. *Journal of Public Economics* 91(1-2): 235-258.

[3] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B* 39(1): 1-38.

[4] Dionne, G., Pinquet, J., Maurice, M. and Vanasse, C. (2011) Incentive mechanism for safe driving: a comparative analysis with dynamic data. *Review of Economics and Statistics* 93(1): 218-227.

[5] Garvie, D. and Keeler, N. (1994) Incomplete enforcement with endogenous regulatory choice. *Journal of Public Economics* 55(1): 141-162.

[6] Heyes, A. and Rickman, N. (1999) Regulatory dealing - Revisiting the Harrington paradox. *Journal of Public Economics* 72(3): 361-378.

[7] Heckman, J.J. and Singer, B. (1984) A method for minimizing the impact of distributional assumptions in econometric models of duration. *Econometrica* 52(2): 271-320.

[8] Kaplow, L. (1990) Optimal deterrence, uninformed individuals, and acquiring information about whether acts are subject to sanctions. *Journal of Law, Economics and Organization* 6(1): 93-128.

[9] Levitt, S.D. (1997) Incentive compatibility constraints as an explanation for the use of prison sentences instead of fines. *International Review of Law and Economics* 17(2): 179-192.

[10] Nyborg, K. and Telle, K. (2004) The role of warnings in regulation: keeping control with less punishment. *Journal of Public Economics* 88(12): 2801-2816.

[11] Polinsky, A.M. (2006) Optimal fines and auditing when wealth is costly to observe. *International Review of Law and Economics* 26(3): 323-335.

[12] Polinsky, A.M. and Rubinfield, D.L. (1991) A model of optimal fines for repeat offenders. *Journal of Public Economics* 46(3): 291-309.

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

[14] Rousseau, S. (2009) The use of warnings in the presence of errors. *International Review of Law and Economics* 29(3): 191-201.

[15] Shavell, S. (1987) The optimal use of non-monetary sanctions as a deterrent. *American Economic Review* 77(4): 584-592.