INSTITUTE OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

FOI Working Paper

Quota Enforcement in Resource Industries: Self-Reporting and Differentiated Inspections

Lars Gårn Hansen
Frank Jensen
Linda Nøstbakken

2010 / 10

FOI Working Paper 2010 / 10

Quota enforcement in resource industries:
Self-reporting and differentiated inspections

(revised version May 2011)

Authors: Lars Gårn Hansen, Frank Jensen, Linda Nøstbakken

Institute of Food and Resource Economics
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.foi.life.ku.dk

Quota Enforcement in Resource Industries: Self-Reporting and Differentiated Inspections*

Lars Gårn Hansen† Frank Jensen‡ Linda Nøstbakken§

Abstract

Quotas or permits are frequently used in the management of renewable resources and emissions. However, in many industries there is concern about the basic effectiveness of quotas due to non-compliance. We develop an enforcement model of a quota-regulated resource and focus on a situation with significant non-compliance and exogenous constraints on fines and enforcement budget. We propose a new enforcement system based on self-reporting of excess extraction and explicit differentiation of inspection rates based on compliance history. In particular, we use state-dependent enforcement to induce firms to self-report excess extraction. We show that such system increases the effectiveness of quota management by allowing the regulator to implement a wider range of aggregate extraction targets than under traditional enforcement, while ensuring an efficient allocation of aggregate extraction. In addition, inspection costs can be reduced without reductions in welfare.

JEL Classification: D61, H0, Q20, Q22, Q28
Keywords: Enforcement, Non-compliance, Self-reporting, Differentiated inspections, Quotas, Emissions standards, Resource Management

*We are grateful to Peder Andersen, Heather Eckert, Rögnvaldur Hannesson and John Livernois for helpful comments and suggestions on earlier versions of the paper. All remaining errors and omissions are our own.

†Institute of Food and Resource Economics, University of Copenhagen, Denmark.
‡Institute of Food and Resource Economics, University of Copenhagen, Denmark.
§Alberta School of Business, University of Alberta, Canada, and SNF, Norway.

1 Introduction

Common pool resources, such as clean air, fisheries, ground water and forests, are often regulated by extraction (or emission) quotas, and much research has been devoted to the optimal design of quota-based systems. However, the productivity growth in production technologies has strengthened the incentives for firms to violate quotas and increased the importance of ensuring effective enforcement. At the same time, there seems to be widespread political reluctance to respond to the challenge by increasing inspection resources and sanctions. The literature reflects these mounting non-compliance problems by showing an increasing focus on enforcement issues in resource management. Generally, increased compliance requires tougher enforcement or tougher punishment. However, both options may be politically infeasible due to budgetary and legal constraints.[1] This could leave resource managers in a situation with substantial non-compliance problems but without the ability to take further actions to reduce violations.

In this paper, we propose a reform of the traditional enforcement system that increases the effectiveness of quota regulation while satisfying budgetary and legal constraints. We analyze the welfare implications of the reformed enforcement system within a standard resource model, and investigate whether it is possible to achieve improved welfare without violating budgetary and legal constraints.

We develop a standard resource model where a given number of firms with heterogenous production costs harvest a resource that is regulated by non-tradable extraction quotas. The enforcement model contains two important extensions of the traditional quota enforcement model. First, firms may self-report resource extraction in excess of quotas. Upon doing so the firm pays a given amount per self-reported unit (a reduced "fine"). Hence, it becomes legal to exceed quotas as long as the correct production level is reported and paid for. Second, we introduce differentiated inspections based on firms' compliance history. Firms that are inspected and found to exceed their quotas without correctly self-reporting this, are moved into an inspection group with a higher inspection rate for a given period of time. In addition, detected violators are prosecuted and punished (fined). Thus, we propose a system of state-dependent enforcement with self-reporting. Note that in contrast to much of the previous work on enforcement of environmental regulations, quota violations can take on a continuum of values in our model. Consequently, self-reports may not be truthful and inspections of firms that self-report are required.

[1] Fines are typically constrained by the principle that the punishment should be proportional to the crime, which restricts the use of higher fines to combat illegal resource use. Furthermore, substantial increases in enforcement costs are often politically infeasible due to budgetary constraints.

Our results show that we can improve welfare by using state-dependent enforcement to induce firms to submit correct self-reports. When fines are constrained, the reformed enforcement system can always achieve a given aggregate production target more efficiently than under traditional enforcement. This holds irrespectively of whether we allow for differentiated inspection rates across firms under traditional enforcement. Furthermore, the proposed system increases the effectiveness of the quota instrument, since it allows us to achieve a wider range of target production levels. Hence, the proposed enforcement system increases both the effectiveness and the efficiency of enforcement. We provide an example and numerical results that demonstrate these improvements, as well as possible limitations of the proposed enforcement system.

To better understand the results, note first that it is unattractive for firms to be in the group where the inspection rate is high ("control hell"). The threat of being moved into control hell is an additional deterrent that induces higher compliance in groups where inspection rates are lower. If this threat is effective, few firms enter the group with high inspection rates, and hence, the inspection cost for this group is low. So far we are in line with the existing literature on state-dependent enforcement. However, contrary to this literature, we also use the threat of control hell together with a tight quota to induce firms to exceed their quota and truthfully self-report. If all firms self-report excess extraction, they all face the same shadow price of production, and hence, total production is allocated efficiently across firms. Consequently, it becomes more important to focus on inducing truthful self-reporting than obtaining full quota compliance. This is in contrast to recent results on enforcement of environmental regulations by Macho-Stadler & Pérez-Castrillo (2006).

Finally, the introduction of self-reporting and differentiated inspections into the management of quota-regulated resources represents another contribution of the paper. These instruments are commonly used in environmental management, but have not been analyzed as options for the management of renewable resources.

The enforcement literature has proposed both regulatory dealing and self-reporting as mechanisms that can achieve increased compliance without increasing the number of inspections.[2] First, high compliance rates combined with low sanction rates

[2]Other explanations for high compliance rates in environmental regulation with low sanctions and inspection rates have been suggested. One explanation is the risk of repercussions on financial and output markets by violation of environmental regulations, which affect firm profits (see e.g. Hamilton, 1995; Konar & Cohen, 1997; Anton et al., 2004). If consumers or investors care about the firm's environmental reputation, their reaction to disclosures of non-compliance with environmental regulations could be costly to the firm, which may explain higher compliance rates even though regulatory sanctions are small. Such effects may be important in the case of large differentiated firms that consumers and investors can identify in the market, but are presumably less important for smaller, undifferentiated firms that are not easily identified in the market, such as those operating in many resource industries. See also Helland (1998), Sandmo (2000), and Short & Toffel (2008).

can be the result of what Heyes & Rickman (1999) refer to as regulatory dealing (Harrington, 1988; Greenberg, 1984; Heyes & Rickman, 1999). The basic idea is that firms are given lenient treatment in some situations where they do not comply in exchange for increased compliance in others. The mechanism we use in our model is based on Greenberg (1984) and strengthens the incentives to compliance by defining explicit rules that govern how inspection rates are differentiated across firms. Contrary to the existing literature, we use state-dependent enforcement (differentiated inspections and sanctions) to induce firms to self-report, rather than to increase compliance rates.

Second, self-reporting is commonly used in environmental regulation (Russell, 1990) and have proven to be effective in many cases where high compliance rates are achieved even if sanctions and inspection rates are relatively low (Livernois & McKenna, 1999). Much work considers the introduction of self-reporting into a fine-based environmental regulation system, showing that this can increase compliance and efficiency (e.g. Malik, 1993; Kaplow & Shavell, 1994; Livernois & McKenna, 1999; Innes, 1999, 2001; Macho-Stadler & Pérez-Castrillo, 2006; Evans et al., 2009). The main advantage is that self-reporting allows the regulator to increase compliance by focusing control resources on agents that do not self-report violations (Kaplow & Shavell, 1994; Malik, 1993; Innes, 1999).[3] Furthermore, self-reporting may allow regulated agents to reduce their avoidance costs (Innes, 2001). The enforcement system must give agents incentives to self-report, for example by reducing the fine for self-reported relative to unreported violations (Livernois & McKenna, 1999).

We use the case of the fishery as an example throughout the paper. Most fisheries are quota regulated and illegal fishing is currently widespread worldwide. Recent estimates suggest that illegal and unreported catches constitute on average about 20% of reported catches globally, with a total value of US$5-11 billion (Agnew et al., 2009). A number of studies in the fisheries economics literature investigate optimal enforcement of a regulated fishery within the traditional enforcement system (Sutinen & Andersen, 1985; Milliman, 1986; Anderson & Lee, 1986; Furlong, 1991). Others consider the choice of regulatory instruments in the presence of non-compliance (Charles et al., 1999; Chavez & Salgado, 2005). Our approach of combining self-reporting and differentiated inspections to increase the effectiveness of a given inspection budget is a novel contribution both to this literature and the general enforcement literature. This enforcement system can potentially yield significant improvements in efficiency also in other settings such as pollution, both in terms of control efforts and emissions abatement.

[3] In the environmental literature, compliance is usually a binary choice (to comply with or violate regulations). Hence, if one self-reports, there is no reason for such report to be untruthful.

The paper is organized as follows. Section 2 presents the basic enforcement model under the traditional quota enforcement system. The model is specified for a quota regulated fishery. In section 3, we introduce our proposed enforcement system based on self-reporting and differentiated inspections and proves theoretically that the proposed system generally is both more efficient and more effective than the traditional enforcement system. Welfare effects are analyzed in section 4. Section 5 provides a numerical example to illustrate the ideas. A simulation model is used to illustrate optimal enforcement under the traditional and the self-report based enforcement systems as well as to compare the two under different specifications and requirements with regard to the regulatory intensity needed. Section 6 provides concluding remarks.

2 The Traditional Enforcement System

In this section, we develop a model of a quota-regulated fishing industry consisting of n firms that harvest a fish stock. A regulator sets a total quota that is allocated in equal shares to the n firms as non-tradable quotas.[4] The regulator can only detect quota violations through costly inspections that allow him to observe firm level catches.

The objective of the regulator is to maximize sustainable welfare. With a constant output price, this is equivalent to maximizing aggregate industry profits net of inspection costs. Under traditional enforcement, the regulator has two instruments; the size of the total quota and the inspection rate. When a firm is inspected and found to violate regulations, it can be fined. The maximum fine is assumed exogenously given, and hence, higher fines cannot necessarily be imposed to reduce illegal fishing. Furthermore, there is a budget constraint on control efforts that limits the inspection rate, since substantial increases in enforcement costs are often politically infeasible due to budgetary constraints.

2.1 The Firms and the Resource Stock

Total harvest is subject to a resource constraint $\Delta X_t = F(X_t) - Y_t$, which states that the change in the resource stock in period t equals the period's stock growth, $F(X_t)$, minus the total harvest, Y_t. To keep the analysis tractable we disregard transition dynamics and assume that the regulator compares sustainable states.[5] That is, the regulator considers sustainable catch level (Y) and stock (X) combinations that

[4] In real-world fisheries, many quota systems allow for some trade in quotas, but such trade is often highly restricted.

[5] Accounting for transition dynamics would imply that the quota and other regulatory variables also depend on the current stock size and therefore changes as the system approaches its steady state.

satisfy:

$$Y = F(X), \tag{1}$$

given the objective of maximizing aggregate sustainable profits net of inspection costs.[6]

In a sustainable equilibrium without quota regulations, each firm in the industry chooses the extraction level that maximizes its own profit conditional on the resource stock:

$$\pi_i = py_i - c(y_i, \alpha_i, X), \tag{2}$$

where p is the output price, y_i is firm i's harvest, and $c(\cdot)$ is a cost function that is increasing and convex in harvest quantity and decreasing in the size of the fish stock X. The cost parameter α_i is firm specific, indicating cost differences between the n firms. All differences between firms are captured in the cost parameter α_i. Hence, the industry is uniquely characterized by the distribution of cost parameters $g(\alpha)$.

Let $y^*(\alpha_i, X)$ denote the optimal harvest level of a firm with cost parameter α_i at a given stock level. Aggregate harvest is the sum of all firms' catches, $Y = \sum\limits_i y_i^*$. Steady-state aggregate harvest must equal stock growth in each period. This implies the following steady-state relationship between aggregate harvest and stock:

$$\sum_i y^*(\alpha_i, X) = F(X). \tag{3}$$

A large resource stock lowers marginal extraction costs ($\frac{\partial^2 c}{\partial y \partial X} < 0$). Hence, sustainable combinations of large stock and yield are preferable. Without regulations, however, firms with relatively low marginal costs harvest too much and the stock is driven down. Hence, the sustainable unregulated equilibrium is characterized by low yield and stock levels. In certain cases, the stock can even be driven to extinction. Regulations are introduced because of this externality in resource extraction. The purpose of regulation is to reduce extraction below the uncoordinated level to reach the preferred equilibrium. Next we introduce quota regulation and enforcement.

2.2 The Regulator and Enforcement

Each firm is allocated a non-transferable quota, q. The firm chooses whether to comply with its quota, knowing that quota violations come at the risk of being fined if detected. The regulator can only observe the firms' harvest levels by conducting

[6] The analysis presented in the following generalize to a dynamic setting. However, while this complicates derivations, it does not affect the general results nor does it provide additional insights into the functioning of the suggested enforcement system.

costly inspections and is constrained by an inspection budget that allows for a given number of inspections per period, $m < n$. Without differentiation between firms, this results in an inspection rate of $\gamma = \frac{m}{n} < 1$ for each firm per period of time. The cost per inspection is c_m. We assume that each firm is inspected at the most once per period, and that the inspection accurately reveals the actual harvest level of the firm in that period (no inspection error). Hence, we disregard the possibility of firms making several fishing trips per period.

A fine up to a maximum of f can be imposed per unit harvested in excess of the quota. The maximum fine is exogenously given by legislation and statutes, and is assumed to be high enough to fully deter quota violations if applied with certainty. The regulator knows the industry's cost function and the statistical distribution of cost parameters $g(\alpha)$, but does not know the individual firm's cost parameter α_i. All n firms are allocated the same resource quota $q = \frac{Q}{n}$, where Q is the total allowable harvest.[7] Firms choose harvest quantities to maximize profits net of expected fine payments (cf. equation 2), i.e.:[8]

$$y_i^*(\alpha_i, q, \gamma, X) = \arg\max_{y_i} \left[\pi\left(y_i, \alpha_i, X\right) - \gamma f\left(\max(0, y_i - q)\right)\right] \tag{4}$$

The regulator maximizes total sustainable industry profit net of enforcement costs. Assuming that the fine is set to its maximum value, f, the problem of the regulator can be stated as follows:

$$\begin{array}{ll} \max\limits_{\gamma,q} & \left(n \int\limits_{\alpha} \pi\left(y_i^*, \alpha_i, X\right) dg(\alpha) - c_m \gamma n\right) \\ \text{s.t.} & y_i^* = \arg\max\limits_{y_i} \left[\pi\left(y_i, \alpha_i, X\right) - \gamma f\left(\max\left(0, y_i - q\right)\right)\right] \ , \forall i \\ & n \int\limits_{\alpha} y_i^*\left(\alpha_i, q, \gamma, X\right) dg(\alpha) = F(X) \\ & 0 \leq \gamma \leq \frac{m}{n} \end{array} \tag{5}$$

The first line of the problem (5) is the sum of industry extraction profit, which is given as the number of firms n multiplied by the average extraction profit over all firms, minus inspection costs ($c_m \gamma n$). Industry profit depends on the distribution of the cost parameter α. The problem is to choose the quota and inspection rate that maximize extraction profits subject to three constraints: (i) firms choose profit maximizing harvest quantities (second line), (ii) aggregate harvest equals stock growth in equilibrium (third line), and (iii) the inspection rate does not cause a violation of the inspection budget (fourth line).

[7] Note that regulators typically differentiate quotas according to e.g. the type and size of the firm. This is not our focus and to keep the analysis tractable we disregard this. Extending the model accordingly is straight forward.

[8] We disregard price or cost differences between fish extracted legally and illegally. The analysis easily generalizes to the case of price and/or cost differences between legal and illegal extraction.

If it is necessary to cap aggregate catches, the solution to the problem (5) is to set quotas tight enough for all firms to catch illegally. All firms are then constrained by the expected fine on illegal catches rather than by the quota. The regulator then sets inspection rates (γ) so that the optimal catch level is achieved, given that the available inspection budget allows for it. This ensures an efficient allocation of the aggregate catch target since all firms face the same marginal shadow price of catches.[9]

The standard regulatory approach in fisheries seems to be to take the enforcement system and its costs as given (i.e., fix inspection costs at the allowed maximum) and use the resource quota q as the only policy instrument.[10] Hence, γ is fixed at its maximum level, $\bar{\gamma} = \frac{m}{n}$, and the problem becomes:

$$\begin{array}{ll} \max\limits_{q} & \left(n \int\limits_{\alpha} \pi\left(y_i^*, \alpha_i, X\right) dg(\alpha) - c_m \bar{\gamma} n \right) \\ \text{s.t.} & y_i^* = \arg\max\limits_{y_i} \left[\pi\left(y_i, \alpha_i, X\right) - \bar{\gamma} f\left(\max\left(0, y_i - q\right)\right)\right] \ , \forall i \\ & n \int\limits_{\alpha} y_i^*\left(\alpha, q, \bar{\gamma}, X\right) dg(\alpha) = F(X) \end{array} \tag{6}$$

For large values of q no firms are constrained by the quota. As q is tightened, there is a point at which some firms become constrained, and from this point onward quota reductions reduce aggregate harvest. As we continue to reduce q, more firms become constrained and the shadow cost of the quota increases for those already constrained. At some point the quota constraint is so restrictive that the most efficient firms choose to harvest illegally.[11] From this point onward, these firms do not respond to further reductions in quotas since they are restricted by the expected fine, not the quota. If we continue to reduce q, more and more firms exceed their quota and the effectiveness of the quota instrument is gradually reduced. Eventually, all firms exceed the quota and further quota reductions do not affect aggregate harvest. At this point, the quota instrument is completely ineffective.

In a situation where most firms are quota constrained, the allocation of production shares is inefficient since heterogeneous firms are constrained by a uniform quota. The standard recommendation in such situation is to make quotas tradable, which allows for an equalization of shadow prices of catches across firms. However, in many fisheries there are substantial non-compliance problems and one may be close to or at the point where all firms violate quotas. In such cases, (almost) all

[9] In contrast, if firms were constrained by the uniform catch quota or by different expected fines, aggregate catch would be allocated inefficiently because the marginal shadow price of catches would differ across firms.

[10] This implies spending the entire inspection budget, and consequently, a maximization of the inspection rate γ.

[11] This occurs when the marginal shadow cost of their quota exceeds the expected fine.

firms are exceeding their catch quotas and are thus constrained by the expected fine on illegal catches. When all firms face (and perceive) the same inspection probability, their shadow prices of catches become identical. Consequently, total catches are allocated efficiently across firms. The main problem in such situation is that further reductions in aggregate harvest cannot be achieved by tightening the quotas. It may therefore be well-founded when resource regulators seem more concerned with the lack of effectiveness of quotas than with quota tradability.

When constrained by the inspection budget, the effectiveness of enforcement increases if we differentiate inspection rates between firms. Although such differentiation typically is not part of the formal enforcement system, this may be what control agencies try to do when targeting firms that in the past have been less compliant than others.[12] This increases the enforcement effectiveness if these firms are in fact more responsive to changes in expected fines. However, such differentiation leads to differences in the expected punishment between firms, and consequently, reduces the efficiency of the aggregate catch allocation.

We must consider alternative enforcement schemes to achieve further reductions in aggregate harvest while ensuring an efficient allocation. The self-report based system introduced next aims at doing just that.

3 The Self-Report Based Enforcement System

We now propose an alternative to the traditional quota enforcement system based on self-reporting and differentiated inspection rates. Although illegal fishing is a considerable problem worldwide (Sumaila et al., 2006; Agnew et al., 2009), neither self-reporting nor differentiated inspections have been formally analyzed in the context of fisheries, nor have they been applied in fisheries regulation.[13]

We present the alternative enforcement system within the same framework as we used for traditional enforcement above. There are, however, some important differences. Instead of inspecting all firms with the same probability, firms are assigned to one of two enforcement groups that differ in inspection probabilities; group 1 with low probability of inspection, and group 2 with high probability of inspection. In the first group, firms are allowed to self-report harvest quantities in

[12] To our knowledge, differentiated inspection rates are not a formal part of the enforcement system in any fishery. However, we know that at least in some fisheries, inspectors do to some degree target firms that based on their compliance records are perceived to have a higher likelihood of violating regulations.

[13] Some regulatory systems have elements that resemble self-reporting. In many regions, such as Australia, Canada, the European Union, Iceland, Norway, and the United States, fishing vessels are required to keep logbooks with information about their catches and harvest activities. However, the key element of a self-report based enforcement system, namely that firms are given incentives to self-report violations, is to our knowledge not part of current fisheries regulation systems.

excess of quota, in which case there is a rebate on the fine paid. If inspected firms are found to have self-reported all excess catches they remain in the first group. If they have not, they must pay the full fine and are moved to the second inspection group. The threat of being moved to the second group, the so-called "control hell" with high inspection rates, is an effective deterrent that makes it possible to increase firms' perceived punishment relative to traditional enforcement, without increasing inspection costs.[14] In addition, the self-reporting scheme allows the regulator to use the self-report rebate, that is, the reduction in fine when a firm self-reports excess catches, as an additional control variable. This increases the flexibility of the enforcement system and makes it possible to increase the allocation efficiency of the system, as we show below.

The idea of using the threat of control hell to strengthen the firms' incentives to comply without increasing fines or inspection costs was originally proposed by Greenberg (1984). We use it in basically the same form but for a different purpose; to induce self-reporting of violations rather than compliance. Self-reporting of violations in the environmental enforcement literature is often seen as a way to increase efficiency by reallocating inspection resources to firms that do not self-report violations (Kaplow & Shavell, 1994; Malik, 1993; Innes, 1999). This is because violations in these models can take on only one value and a self-reported violation by a rational agent therefore must be truthful. In our setting, quota violations can take on a continuum of values and therefore require inspection to ensure truthfulness. This type of violations is considered by Macho-Stadler & Pérez-Castrillo (2006), but they find that enforcement resources should be focused on inducing compliance rather than truthful self-reporting. In our case, we enforce an inefficient allocation of quotas and it is therefore better to allow firms to exceed their quotas and instead induce truthful self-reporting of excess catches, which improves the allocative efficiency.

3.1 The Regulator and Enforcement

A firm's inspection probability depends on whether the firm is in group 1 or group 2 and is denoted $\gamma_j \in [0, 1]$, where $j = 1, 2$ refers to the group. A firm in group 1 that self-reports harvest in excess of quota must pay a fine rf per unit, where $r \in (0, 1)$ is a factor representing the fine rebate for self-reporting. In group 2, self-reporting gives no rebate, hence, a firm that self-reports must pay the full fine f per unit. Furthermore, a firm in group 1 that is inspected and found to have underreported its quota violation must pay the full fine and is moved to group 2. Once in group 2,

[14] There are several possibilities to make control hell even crueler and thereby strengthen its deterrence effect, such as to introduce quota reductions for firms in control hell.

the firm stays there until found to have self-reported correctly during u consecutive inspections after which the firm is moved back into group 1.

Table 1: Punishment Scheme

	Group 1		Group 2	
	Self-report	Violate	Self-report	Violate
Not inspected	$rf(y_i - q)$	0	$f(y_i - q)$	0
Inspected	$rf(y_i - q)$	$f(y_i - q)$	$f(y_i - q)$	$f(y_i - q)$
Inspection prob.	γ_1 *Violate: move to group 2*		$\gamma_2 > \gamma_1$ *Full self-reporting u times: move to group 1*	

The inspection probabilities in table 1 are determined by the regulator and are constrained by the inspection budget. As under the traditional enforcement system, the regulator can perform a given number of inspections per year, denoted m, which determines the inspection probability. If all firms are equally likely to be inspected (all firms are in group 1), the inspection probability is $\gamma_1 \leq \frac{m}{n}$.[15] The inspection rate is higher in group 2 than in group 1. Hence, the more firms there are in inspection group 2, the lower the inspection rate can be in group 1.[16] As before, the maximum fine is exogenously given and high enough to fully deter quota violations if applied with certainty.

The regulator seeks to maximize total industry profit net of enforcement costs $n \int_\alpha \pi\left(y^*, \alpha, X\right) dg(\alpha) - c\hat{\gamma}n$, where $\hat{\gamma}$ refers to the average inspection rate over both inspection groups (weighted average). However, now the set of policy instruments available to the regulator includes two inspection rates (γ_1 and γ_2) and the period of time a detected violator must be in control hell (group 2) before it can be moved back into group 1.

To ensure an efficient allocation of aggregate catch across firms, the regulator sets the total quota sufficiently low for the individual quota to bind for all firms, thereby inducing them to exceed the quota. As discussed above, this results in all firms having the same marginal harvest cost, and hence, an efficient catch allocation.

3.2 The Firms

Under self-report based enforcement, the firm has four main options. It can (i) comply with its quota, (ii) report the entire illegal extraction, (iii) report some of

[15]The inspection probability is assumed to be positive and strictly below one.

[16]In general, the following must hold: $1 \geq \gamma_2 > \gamma_1 > 0$. In addition, the inspection budget cannot be exceeded, which implies that $\gamma_1 \leq \frac{m - \gamma_2 n_2}{n - n_2}$, where n_2 is the number of firms in inspection group 2. This implies that if the severity of control hell is constrained, there exists a possibility that too many firms end up in group 2, thereby draining group 1 for inspection resources.

the illegal extraction, or (iv) not report any extraction in excess of the quota. With a fine structure that is linear in illegal quantity and detection probabilities being constant, it is easily shown that the firm either reports all or does not self-report any excess extraction. Thus, the relevant options for a firm are reduced from four to three, as option (iii), where one exceeds the quota and reports only part of the excess quantity, is never chosen.

This leaves us with three distinct behavioral strategies a profit-maximizing firm can use. The firm chooses the strategy that yields the highest sum of discounted future profits.[17]

Strategy A: Stay in group 1. To ensure that the firm is never moved into group 2, the firm must always comply with regulations. Consequently, the firm must self-report any excess extraction (options i or ii). Since the quota is set sufficiently low for no firm to find option (i) optimal, only option (ii) remains. In a sustainable equilibrium, optimal harvest is constant over time. Hence, $y_i^a = \arg\max_{y_i} \left[\pi\left(y_i, \alpha_i, X\right) - rf\left(y_i - q\right)\right]$, which gives a net expected profit of $\Pi_i^a = \pi\left(y_i^a, \alpha_i, X\right) - rf\left(y_i^a - q\right)$. If we let EV_i^a denote the present value of future profits for firm i when following strategy A, we have that:

$$EV_i^a = \sum_{t=0}^{\infty} \beta^t \Pi_i^a, \tag{7}$$

where β is the discount factor.

Strategy B: Alternate between groups. To alternate between groups, the firm must be willing to violate regulations while in group 1 and comply with regulations while in group 2. Thus, the behavior of a firm that follows strategy B depends on the inspection group the firm is currently in. In group 1, the firm violates quotas (option iv), while in group 2, the firm self-reports all excess extraction (option ii). Formally, when in group 1, the firm chooses $y_i^{b1} = \arg\max_{y_i} \left[\pi\left(y_i, \alpha_i, X\right) - \gamma_1 f\left(y_i - q\right)\right]$, which gives net expected profit of $\Pi_i^{b1} = \pi\left(y_i^{b1}, \alpha_i, X\right) - \gamma_1 f\left(y_i^{b1} - q\right)$. In group 2, the firm chooses $y_i^{b2} = \arg\max_{y_i} \left[\, \pi\left(y_i, \alpha_i, X\right) - f\left(y_i - q\right)\, \right]$, which gives net expected profit of $\Pi_i^{b2} = \pi\left(y_i^{b2}, \alpha_i, X\right) - f\left(y_i^{b2} - q\right)$. In the first period under this strategy, the firm is in group 1 and expected profit is Π_i^{b1}. The inspection rate γ is the probability of being moved to group 2 in the next period, and hence, expected profit in the next period is $(1-\gamma)\Pi_i^{b1} + \gamma\Pi_i^{b2}$. In every future period t the firm perceives some probability $0 \leq \nu_i(t) \leq 1$ of being in group 2 (where $\nu_i(0) = 0$, $\nu_i(1) = \gamma$, etc.). Hence, the expected profit in period t is $\left(1 - \nu_i(t)\right)\Pi_i^{b1} + \nu_i(t)\Pi_i^{b2}$. Thus, the present

[17] For more complex punishment schemes in repeated games, see Abreu (1988) and the literature that followed on optimal penal codes.

value of future profits under strategy B becomes:

$$EV_i^b = \sum_{t=0}^{\infty} \beta^t \left[(1 - \nu_i(t)) \Pi_i^{b1} + \nu_i(t) \Pi_i^{b2} \right], \tag{8}$$

where $0 \leq \nu_i(t) \leq 1$ for all t.

Strategy C: Stay in group 2. To always be in group 2, the firm must never comply with its quota nor self-report excess extraction. Thus, the firm's only option is to always violate the quota (option iv). This yields a catch quantity of $y_{it}^c = \arg\max_{y_{it}} [\pi(y_{it}, \alpha_i, X_t) - \gamma_2 f(y_{it} - q_t)]$, with corresponding net expected profit of $\Pi_{it}^c = \pi(y_{it}^c, \alpha_i, X_t) - \gamma_2 f(y_{it}^c - q_t)$. The present value of future profits for a firm following strategy C is:

$$EV_i^c = \sum_{t=0}^{\infty} \beta^t \Pi_i^c. \tag{9}$$

Since the maximum fine (f) by definition is sufficiently high to fully deter violations if applied with certainty, it is never optimal for a firm to play strategy C in group 2 when $\gamma_2 = 1$.[18] Hence, strategy C is always dominated by self-reporting for firms in group 2 (strategy B). Thus, firms choose either strategy A or strategy B.

4 Welfare implications

We now turn to developing the core results of the paper. We do this by comparing welfare under the different enforcement systems; self-report based enforcement and traditional enforcement with and without differentiation of inspection rates.

Compared to traditional enforcement, there are several additional policy instruments available under the self-report based system. Introducing more enforcement policy variables to the regulator's toolbox, cannot reduce welfare if policy variables are set optimally, since the traditional enforcement system is a possible specification. In the following, we prove two propositions showing that there generally is a welfare gain when shifting to the self-report based enforcement system.

The first proposition considers the situation where quotas under the traditional enforcement system have been tightened so much that all firms violate. In this situation, all firms are constrained by the expected fine and not the quota, hence,

[18] As long as the budget constraint allows, $\gamma_2 = 1$. If many firms end up in group 2 at the same time, it is possible that $\gamma_2 = 1$ cannot be achieved without violating the inspection budget. This is easily dealt with by introducing additional inspection groups that are even less attractive to firms, or by increasing the deterrence effect of group 2 in other ways, such as by introducing quota reductions. We therefore ignore this possibility in the following.

further quota reductions have no effect on aggregate harvest. The first proposition states that a self-report based enforcement system allows the regulator to implement further welfare increasing reductions in aggregate harvest, while ensuring an efficient distribution of these reductions across firms. The second proposition considers the situation where quotas under traditional enforcement are still effective, that is, some firms are constrained by the quota. The proposition states that in such case, the aggregate catch target can be implemented more efficiently under self-report based, state-dependent enforcement.

Proposition 1. *When all firms violate quotas so that aggregate harvest cannot be reduced further under traditional enforcement, there generally exists an enforcement system with self-reporting and differentiated inspections that reduces aggregate harvest and allocates this reduction efficiently among firms without increasing the inspection cost of the enforcement agency.*

Proof. The proof of proposition 1 begins by considering a differentiated inspection system where the inspection rates are $\gamma_1 = \gamma$ and $\gamma_2 = 1$, and where the fine rebate factor when self-reporting is $r = \gamma$. Noting that quotas are exceeded by all firms when the expected fine is γf, we have $y_i^a = \arg\max_{y_i} [\pi(y_i, \alpha_i, X) - rf(y_i - q)]$, $y_i^{b1} = \arg\max_{y_i} [\pi(y_i, \alpha_i, X) - \gamma f(y_i - q)]$, and $y_i^{b2} = \arg\max_{y_i} [\pi(y_i, \alpha_i, X) - f(y_i - q)]$. From (2) we know that $\pi(\cdot)$ is concave. In addition, we know that $rf = \gamma f < f$ and $y_i^{b2} = q$. Consequently, the net expected profits associated with the harvest levels of the different strategies are so that $\Pi^a = \Pi^{b1} > \Pi^{b2}$.

This implies that $\beta^t \Pi_i^a \geq \beta^t \left[(1 - \nu_i(t))\, \Pi_i^{b1} + \nu_i(t)\Pi_i^{b2}\right]$ for all t when $0 \leq \nu_i(t) \leq 1$. Hence, by equations (7) and (8) we have

$$\sum_{t=0}^{\infty} \beta^t \Pi_i^a \geq \sum_{t=0}^{\infty} \beta^t \left[(1 - \nu_i(t))\, \Pi_i^{b1} + \nu_i(t)\Pi_i^{b2}\right],$$

where the term on the right-hand side (RHS) is EV_i^a and the term on the left-hand side (LHS) is EV_i^b. Furthermore, with $\nu_i(1) = \gamma$ and $\Pi^{b1} > \Pi^{b2}$ it is clear that condition (4) can only hold with equality if $\beta = 0$. Hence, $EV_i^a = EV_i^b$ only occurs if the firm completely disregards the future. The expected present value of strategy A is strictly larger than that of strategy B if $\beta > 0$. It follows that for $\beta > 0$, where EV_i^a is strictly greater than EV_i^b, there exists a value of $r = \gamma + \epsilon$, where ϵ is a small positive constant, for which strategy A dominates for all firms. Thus, with self-reporting and differentiated inspection rates it is possible to reduce illegal catches slightly, without exceeding the exogenous constraint on the imposed fine or the inspection budget. Since all firms choose strategy A, no firms enter group 2, and hence total inspection costs equal $c\gamma n$. Furthermore, since all firms self-report all quantities in excess of quotas and pay rf per unit, firms' optimal harvest quantities ensure that all firms face the same marginal shadow cost of harvesting in equilibrium. Consequently, the aggregate harvest reduction is allocated efficiently across firms. □

Corollary 1. *In the situation specified in proposition 1, an enforcement system with self-reporting and differentiated inspections allocates a reduction in aggregate harvest more efficiently than what is possible with differentiation of inspection rates under traditional enforcement.*

Proof. From proposition 1 it follows that the self-report based enforcement system with differentiated inspections allocates the reduction efficiently. Hence, increased efficiency is impossible regardless of enforcement system used. Furthermore, any reduction in aggregate catch resulting from a differentiation of inspection rates under the traditional enforcement system implies a corresponding differentiation of expected fines. Since any differentiation of expected fines results in inefficient allocation of aggregate catch, such allocation must be strictly less efficient than the allocation implemented by the enforcement system with self-reporting and differentiated inspections. □

Proposition 2. *When some firms under traditional enforcement do not violate quotas, there generally exists an enforcement system based on self-reporting and differentiated inspection rates that implements the same aggregate catch target more efficiently without increasing the inspection cost of the enforcement agency.*

Proof. Consider the same differentiated inspection system as above, with inspection rates $\gamma_1 = \gamma$ and $\gamma_2 = 1$ and with a self-report rebate factor of $r = \gamma$. From the proof of proposition 1 it is clear that $EV_i^a > EV_i^b$ for all firms that violate their quotas when $\beta > 0$. If the rebate factor r is increased marginally, this is also the case for $\beta = 0$. Thus, all firms that violate their quota choose strategy A and self-report violations. Now, consider a quota reduction to the point where all firms choose to exceed their quotas. This results in aggregate harvest below the target. Next, reduce inspection rates in group 1 and increase the self-report rebate factor proportionally until aggregate harvest again reaches the target level. The proportional reductions in γ and r ensure the dominance of strategy A over strategy B, and hence, all firms continue to follow strategy A. Since all firms exceed their quota and fully self-report, they all face the same marginal shadow cost of catch in equilibrium. Hence, the aggregate catch target under self-report based enforcement is implemented efficiently. By assumption, some firms are constrained by quotas and not fines under traditional enforcement. Hence, the aggregate catch target under traditional enforcement is implemented inefficiently. Furthermore, since all firms choose strategy A under self-report based enforcement, no firms enter group 2. Consequently, inspection rates in group 1 are reduced, which implies lower total inspection costs than under traditional enforcement: $c\gamma n \leq C$. It follows that it is possible to reach the same aggregate production target more efficiently than under traditional enforcement with lower inspection costs, without exceeding the exogenous constraint on the fine. □

Irrespectively of how intensive the quota enforcement is under traditional enforcement (with uniform inspection rates), a shift to the proposed self-report based enforcement system generally allows the regulator to increase welfare. Our focus is on the situation where traditional quota regulation is no longer effective (covered by proposition 1 and its corollary). In this situation, the advantages of the proposed enforcement system arise from the combination of differentiated inspection rates and the possibility to self-report excess harvest. First, with two inspection groups, the risk of being moved to control hell increases expected punishment relative to the traditional compliance system, but without increasing inspection costs or exceeding the maximum fine. Second, self-reporting allows the regulator to use the self-report rebate rather than the harvest quota as the control variable when implementing the aggregate harvest target. This ensures an efficient allocation of the total harvest quantity across heterogenous firms. Hence, reducing aggregate catch by shifting to an enforcement system with self-reporting and differentiated inspections results in a strictly greater welfare gain than what would result from any differentiation of inspection rates within the traditional enforcement system.

5 Quantifying welfare effects: An Example

We have shown that a shift to the proposed self-report based enforcement system generally increases welfare and that the welfare gain is strictly greater than what could result from a differentiation of inspection rates within the traditional system. However, if such a reform is to be an attractive option in practice, the welfare gain must be substantial. If the effectiveness problem under traditional enforcement is small or if there are political or legal constraints on the amount of time violators can be assigned to control hell this may limit the welfare advantage of introducing the proposed enforcement system. Furthermore, if differentiation of inspections within the traditional system is a possibility, this might be an attractive second-best enforcement strategy.

In the following, we investigate this using numerical simulations within a standard parametrization of the model developed above. We calculate and compare the welfare of the model fishery under (i) traditional enforcement, (ii) traditional enforcement with differentiated inspections, and (iii) self-report based enforcement with and without constraints on the severity of control hell. We do this for different parameterizations of the model, which reflect different levels of the effectiveness problem under traditional enforcement (see appendix A for more details).

5.1 Parametrization

We assume functional forms that are standard in the natural resource economics literature. The resource constraint (1) is specified using the logistic growth function (see e.g. Clark, 1976):

$$Y = F(X) = hX\left(1 - \frac{X}{K}\right), \tag{10}$$

where h and K, respectively, denote the intrinsic growth rate and the carrying capacity of the resource stock. Extraction costs in (2) have the quadratic functional form (see e.g. Smith, 1969)

$$\pi(y_i, \alpha_i, X) = py_i - \frac{\alpha_i y_i^2}{2X}, \tag{11}$$

where the firm specific cost parameter α is uniformly distributed: $g(\alpha) = \frac{1}{\bar{\alpha} - \underline{\alpha}}$ for $\bar{\alpha} \geq \alpha \geq \underline{\alpha}$. Except for the sensitivity of harvest costs to changes in stock size, which depends on the so-called stock-output elasticity, the chosen parameter values do not significantly affect relative performance of the enforcement systems. Parameter values are therefore normalized. Our model parametrization implies a stock-output elasticity of 0.5 which is in the insensitive tail of the distribution of empirical estimates of this parameter.[19] The derivation of individual and aggregate catch levels and an overview of parameter values used in the simulation model can be found in appendix A.

Finally, we assume that the regulator seeks to implement the maximum sustainable yield (MSY) of the fishery and that he is only concerned with welfare in long-run sustainable states.[20] With this target and parametrization, the effectiveness problem of traditional enforcement can be given a precise and intuitive form.

We wish to compare welfare under alternative enforcement systems at different levels of the aforementioned effectiveness problem under traditional enforcement. To do this we define an indicator of how challenging it is for the regulator to reach

[19] The cost function specified in (11) is equivalent to the production function $y = a_0 E^{0.5} X^{0.5}$, where E is fishing effort and a_0 is a productivity parameter. Hence, the implied stock-output elasticity is 0.5. This is close to parameter estimates for the most cost insensitive types of fisheries; schooling fisheries. See e.g. Bjørndal (1987), who estimates production functions for herring.

[20] The MSY target and the focus on sustainable states facilitate parsimonious comparisons of enforcement systems. However, the results presented are in fact simulated using a dynamic model and generalize to the dynamic setting and to other policy objectives than the MSY target, including that of maximizing economic yield (MEY). Note that the MEY differs from the MSY when costs are as in (11). However, when the objective is MSY, the target stock level is independent of the intrinsic growth rate, which we vary in the following. The target stock level under an MSY objective is half of the pristine stock level.

the management objective: the required regulatory intensity (RRI). For the fishery model specified above, the regulator's challenge increases as the intrinsic growth rate is reduced. The lower the growth rate, all else equal, the stronger the need for regulation and enforcement in order to maintain a certain stock level. The stock regenerates more slowly at low values of h, and hence, equilibrium catches are lower. Firms' incentives to harvest are, however, unchanged. This calls for tougher enforcement to ensure that the target stock level is maintained, and hence, the RRI is higher. If the intrinsic growth rate is sufficiently high ($h \geq \bar{h}$), no regulation of catches is necessary and we say that the required regulatory intensity (RRI) is zero. At the other end of the scale, the toughest challenge the regulator can face is a fishery with an intrinsic growth rate close to zero. In this case, enforcement and regulations must ensure an almost complete elimination of fishing effort.

On this basis, we can formally define the RRI. First, let $\bar{h}$ denote the intrinsic growth rate where the unregulated fishery would result in MSY.[21] Next, we can define the required regulatory intensity as:

$$RRI = 1 - \frac{h}{\bar{h}}.$$

The RRI is a function of the growth rate of the stock. The RRI is normalized to lie between zero and one. RRI is zero when the growth rate of the stock is at its upper bound and one when the growth rate is at its lower bound (arbitrarily close to zero). As the growth rate is gradually reduced from its upper toward its lower bound, the RRI gradually increases toward one.

In the following we make welfare comparisons for different levels of RRI (corresponding to different levels of the intrinsic growth parameter h in our model). We start at the RRI value for which the traditional inspection system has just become ineffective and compare welfare for three alternative enforcement systems at increasing levels of RRI. Under the first enforcement system, the regulator uses traditional enforcement with an undifferentiated inspection rate. The regulator sets the inspection rate to its maximum, as given by the inspection budget, and quotas low enough to ensure that all firms produce illegally. As RRI increases, the regulator cannot do anything to tighten regulations or enforcement. The simulations reflect the welfare loss from the suboptimal stock reduction occurring when landings are too high to maintain MSY. As RRI increases, the distance between the target stock and the implemented stock increases and so does the associated welfare loss.

The second enforcement system we consider is traditional enforcement with differentiated inspection rates. The advantage of differentiation is that target stocks can be implemented at higher RRI values than without differentiation. The disad-

[21] For a given parametrization of the model, $\bar{h}$ is a constant. See appendix A.1 for details.

vantage is that the catch allocation across firms becomes inefficient because firms do not face the same expected fines. The simulations capture the sum of these two effects. We assume that inspection rates are differentiated optimally across firms. This represents the upper bound on regulatory performance in terms of quota implementation under traditional enforcement for a given inspection budget.

The third enforcement system we consider is our proposed self-report based system. As we have shown, this system is capable of attaining the target stock and efficient allocation at the initial point, i.e., the point where traditional enforcement is marginally ineffective. The simulations reflect how much of the potential welfare is captured at increasing values of RRI, both in the unconstrained and constrained cases. We assume that inspection rates are set at the maximum allowed by the budget, that quotas are not reduced for firms in group 2 and that the self-report rebate is always set low enough to ensure that no firm plays strategy B where there is a risk of entering group 2. These restrictions are introduced for technical tractability but are not generally efficient. Thus, we find a lower bound on regulatory performance under this system.

5.2 Comparison of enforcement systems

In the following we compare welfare net of inspection costs for each of the three enforcement systems.[22]

Results are summarized in table 2. Traditional enforcement with a uniform inspection rate is ineffective at RRI values above 0.40. Beyond this RRI value, all firms exceed their quotas and quota reductions no longer affect aggregate catch. Maximal welfare is achieved at precisely this RRI value under traditional enforcement, as indicated in table 2. When increasing the RRI beyond 0.40, catch allocations continue to be efficient but the traditional system can no longer implement the aggregate catch target because the inspection budget constraint has been reached and the enforcement effort cannot be increased any further. As a result, the equilibrium stock level falls further and further below the target, which reduces welfare. In contrast, the unconstrained self-report based system is both effective and efficient over the full range of RRIs and implements the first-best solution. This generates considerably higher welfare over a large interval of RRIs.[23] For an RRI of 0.5, the traditional system only achieves 89.5% of the potential welfare while the equilibrium aggregate catch is 4.1% below the target level. As the RRI increases, the

[22] We assume that the full inspection budget is used under all cases considered. Hence, it is appropriate to compare welfare before deduction of inspection costs.

[23] We do not allow the regulator to adjust inspection costs (i.e., the number of inspections). For the self-report based system there is a trade-off between inspection rates and time spent in control hell. We therefore want to focus on welfare from other sources since it may not be feasible in real-world situations to have very high values of u.

gap between the outcomes under the traditional system and the self-report based system increases. For high values of RRI, that is, for slow growing species, traditional enforcement cannot prevent extinction. In our example, this happens at RRI ≥ 0.74 (cf. table 2).

Table 2: Equilibrium welfare[a] and yield for different RRIs by enforcement system. *Scores relative to first-best solution (100 = optimal).*

RRI	0.40		0.50		0.60		0.74	
	Welfare	Yield	Welfare	Yield	Welfare	Yield	Welfare	Yield
First-best solution	26.83	76.66	23.96	63.88	20.44	51.11	14.45	33.22
Trad., uniform insp.	100	100	89.5	95.9	65.7	75.1	5.8	7.3
Trad., diff. insp.	100	100	66.3	99.4	55.2	88.3	0	0
Self-rep., unconstr.	100	100	100	100	100	100	100	100
Self-rep., $u = 2$	100	100	95.53	98.91	77.86	85.97	21.26	25.58

[a] Welfare before deduction of inspection costs, which are identical for all cases considered.

As long as the RRI is low, it is optimal under traditional enforcement to let the inspection rate be the same for all firms in order to promote efficient allocation of aggregate catch. However, when the objective of MSY can no longer be achieved with an undifferentiated inspection rate ($\gamma_i = 0.2$, $\forall i$), illegal fishing can be reduced by increasing the inspection rates of the most cost efficient firms, while reducing the inspection rates facing the least efficient firms. We explore this possibility by introducing perfect differentiation of inspection rates under the traditional system.[24]

The numerical analysis shows that with perfectly differentiated inspection rates, the MSY catch target can be achieved at RRIs below 0.47, compared to 0.41 with a uniform inspection rate. By targeting those firms that have the highest sensitivity to changes in expected punishment (i.e., the most cost efficient firms), the regulator can reach the catch target for a wider range of RRIs. However, this reduces the cost efficiency of the industry because it causes inefficient allocation of catches across firms. According to our simulation results, this causes a loss of welfare relative to undifferentiated enforcement, even when differentiation achieves aggregate catch levels close to or at the target level. This is illustrated in table 2, where aggregate catch levels under traditional enforcement for RRIs of 0.4 and 0.5 are higher when inspection rates are perfectly differentiated, while welfare levels are considerably lower. For an RRI of 0.5, the regulator almost achieves the catch target by per-

[24] As noted, this represents the best possible outcome in terms of achieving the MSY target by use of differentiated inspection rates. In real industries, regulators do not have perfect information on firm-level costs and must settle with imperfect differentiation of inspection rates.

fectly differentiating inspection rates (99.4%). However, the inefficient allocation of catches causes a significant reduction in welfare (33.7% reduction compared to first-best solution). Hence, with perfect differentiation of inspection rates under traditional enforcement, the loss in welfare from inefficient allocation of catches across firms exceeds the welfare gain from increased aggregate yield.

Consistent with the formal results derived above, we see that self-report based enforcement is always at least as efficient as the traditional system. When the RRI is only slightly higher than the value where traditional enforcement becomes ineffective, little is gained from introducing the self-report based system. If, however, the need for enforcement is high relative to available enforcement resources, the potential gains from introducing the self-report based system can be considerable. We have thus far assumed that u, the number of periods a detected violator must spend in group 2 (control hell), can be chosen freely. In real-world resource management, u may be constrained for legal/political reasons or as a safeguard against imprecise inspection results and inadvertent harvester errors in self-reports. We therefore conclude this section by investigating the implications of imposing an upper limit on the enforcement parameter u. The results for the self-report based system with the number of periods in control hell constrained to $u = 2$ are shown in table 2.[25]

Our numerical results show that self-report based enforcement is considerably less flexible when the number of periods a violater is confined to control hell (u) is constrained. The lower the upper limit on u, the smaller the interval of RRIs over which the enforcement system is capable of reaching the target harvest level. The constrained self-report based enforcement system can maintain the target equilibrium level for RRIs below 0.41 ($u = 1$), 0.42 ($u = 2$) or 0.43 ($u = 3$).[26] Thus, the RRI at which a severely constrained enforcement system no longer can achieve the harvest target is only extended slightly compared to the traditional system and not nearly as much as under the traditional system with perfect differentiation. Nonetheless, the constrained self-report based system is still significantly more efficient. As is evident from table 2, this system generates significant welfare gains compared to traditional enforcement. The higher the RRI, the larger the gain compared to traditional enforcement since the marginal value of improved enforcement increases. Finally, in our simulation, the constrained self-report based system achieves a higher welfare level than the perfectly differentiated traditional system

[25] As noted, when u is constrained, it is no longer necessarily optimal to set policy parameters so that all firms are induced to choose strategy A (self-reporting). To keep things tractable, we assume in the simulations that the self-report rebate factor r is set just low enough to induce all firms to choose strategy A. The r value can be calculated from equation (A.9) in the appendix, by setting $\alpha_0 = \underline{\alpha}$. Thus, the simulations in the table reflect a lower bound on this system's performance.

[26] To improve the performance of a constrained self-report based system, we can introduce quota reductions for firms in group 2. In general, any strategy can be used that makes control hell more hellish and thereby further deters firms from choosing strategy B.

can achieve even though that system achieves a higher equilibrium stock (yield). The higher welfare is due to the welfare loss from inefficient allocation under the traditional system with differentiated inspections.

As noted above, the relative performance of traditional enforcement with differentiated inspections increases if the sensitivity of industry costs to stock size is reduced. When the stock-sensitivity of costs is low, ineffective regulation leads to a larger reduction in the resource stock, all else equal. This implies that the welfare gain from increasing the effectiveness of regulation by differentiating inspections increases. The welfare loss from suboptimal catch allocation is, however, not significantly affected by changes in the stock-output elasticity. Hence, differentiation of inspections within the traditional system yields the largest potential gain when harvest costs are insensitive to changes in stock size. Thus, by further decreasing the stock-output elasticity implied by our model, traditional differentiated inspections would outperform a highly constrained self-report based system at RRI values where the self-report based system is no longer effective. However, the sensitivity of industry costs to stock size implied by our simulation model is already in the lower tail of the range of empirical estimates. In addition, the implementation of perfectly differentiated inspections requires that regulators can identify and focus inspections on the most cost efficient harvesters. This is difficult in the real world, where cost parameters are typically private information. Thus, while we cannot rule out the possibility that traditional enforcement with differentiated inspections in certain situations outperforms a highly constrained version of the self-report based system, this is generally not the case. Furthermore, as long as the constraints on the self-report based system are not too tight, this will never be the case.

6 Concluding Remarks

In this paper, we present an alternative enforcement system for quota-regulated industries. The system is based on self-reports of quota violations and differentiation of inspection rates based on firms' compliance records. Firms that exceed their quotas and self-report pay a reduced fine. We address a situation with significant non-compliance problems, and where both the punishment for quota violations and the inspection budget are constrained. Under traditional enforcement, once these constraints are binding, further quota reductions are ineffective as they cannot be enforced (all or most firms violate their quotas). Inspection agencies may try to address this issue by target inspections on firms with poor compliance records to increase enforcement effectiveness. This, however, comes at the cost of reduced allocative efficiency.

The enforcement system we propose is based on explicit and well-defined differentiation of inspection rates contingent on correctly self-reported quota violations. Rather than targeting firms that are perceived to be more responsive to changes in expected fines, we introduce the threat of a control hell to all firms. Any firm that is detected exceeding its quota without having correctly reported this, faces higher inspection rates for a certain period of time. This threat strengthens deterrence. Furthermore, by relying on self-selection through the self-reporting component, our system increases the effectiveness of inspections, or the range of total extraction targets that can be reached, without prior knowledge about individual firms' responsiveness to incentives. Finally, correct self-reporting increases the allocative efficiency compared to quota compliance, a result that is independent of the initial distribution of quotas. Hence, we achieve increased effectiveness in enforcement without reducing the allocative efficiency.

We use a numerical example to demonstrate these improvements, as well as possible limitations of the proposed enforcement system. The main limitation is that the system's ability to increase the enforcement effectiveness depends on possible constraints on the severity of punishment in control hell. However, our results show that even with tight constraints on control hell the self-report based system generates significant welfare gains relative to traditional enforcement when the inspection budget is constrained. Differentiating inspections across firms under traditional enforcement does generally not increase welfare compared to our proposed system and this represents an unlikely special case.

The use of state-dependent enforcement to induce firms to self-report is a novel contribution to the general enforcement literature. Although we have used the case of the fishery as an example in our analysis, the proposed enforcement system can generate significant welfare gains if applied to other industries facing regulatory non-compliance problems. One example is the enforcement of emissions standards. Note in that respect that the system we propose resembles a system with a combination of non-tradable quotas and a tax on production when firms are induced to always self-report. In addition, we introduce the risk of control hell to make sure firms do not deviate.

Our results imply a shift in focus away from inducing quota compliance *per se*, toward correct self-reporting of violations. This implies a number of additional advantages not captured by our analysis. First, as pointed out by Innes (2001), once regulated firms correctly self-report, they no longer have an incentive to avoid inspections. In many industries there may be significant avoidance opportunities, and consequently, the costs of avoidance and combating avoidance may be substantial (Anderson & Lee, 1986; Milliman, 1986). The welfare effect of not incurring such costs may be substantial, which further increases the relative efficiency of the proposed enforcement system. A second advantage is the reduced risk for firms.

As noted by Kaplow & Shavell (1994), risk-bearing costs are eliminated under self-reporting, which is relevant if firms are risk averse. A third advantage is the possibility of increased precision in extraction and stock estimates when firms report actual extraction. The value of decreased measurement error depends on the characteristics of the resource but can be significant. Hence, in addition to the advantages we have focused on in this paper, the proposed self-report based enforcement system has several other advantages that further increase the potential welfare gain relative to traditional enforcement.

There are several possibilities for extending this work. One possibility is to analysis welfare effects of introducing the proposed self-report based system to other quota regulated industries, such as pollution, water and forest management. Another relevant extension is to relax the assumption that inspections perfectly reveal actual production levels. Finally, we have ruled out the possibility that firms self-report only part of the production level that exceeds their quota by assuming a linear punishment function. Hence, allowing for a more complex punishment function may yield additional results.

References

Abreu, D. (1988). On the theory of infinitely repeated games with discounting. *Econometrica*, (pp. 383–396).

Agnew, D., Pearce, J., Pramod, G., Peatman, T., Watson, R., Beddington, J., & Pitcher, T. (2009). Estimating the worldwide extent of illegal fishing. *PLoS ONE*, 4(2), e4570.

Anderson, L. & Lee, D. (1986). Optimal governing instrument, operation level, and enforcement in natural resource regulation: the case of the fishery. *American Journal of Agricultural Economics*, 68(3), 678–690.

Anton, W., Deltas, G., & Khanna, M. (2004). Incentives for environmental self-regulation and implications for environmental performance. *Journal of Environmental Economics and Management*, 48(1), 632–654.

Bjørndal, T. (1987). Production economics and optimal stock size in a North Atlantic fishery. *Scandinavian Journal of Economics*, 89(2), 145–164.

Charles, A. T., Mazany, R. L., & Cross, M. L. (1999). The economics of illegal fishing: A behavioural model. *Marine Resource Ecnomics*, 14, 95–110.

Chavez, C. & Salgado, H. (2005). Individual transferable quota markets under illegal fishing. *Environmental and Resource Economics*, 31, 303–324.

Clark, C. W. (1976). *Mathematical Bioeconomics*. New York: Wiley & Sons.

Evans, M., Gilpatric, S., & Liu, L. (2009). Regulation with direct benefits of information disclosure and imperfect monitoring. *Journal of Environmental Economics and Management*, 57(3), 284–292.

Furlong, W. (1991). The deterrent effect of regulatory enforcement in the fishery. *Land Economics*, (pp. 116–129).

Greenberg, J. (1984). Avoiding tax avoidance: A(repeated) game-theoretic approach. *Journal of Economic Theory*, 32(1), 1–13.

Hamilton, J. T. (1995). Pollution as news: Media and stock market reactions to the toxics release inventory data. *Journal of Environmental Economics and Management*, 28(1), 98–113.

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

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

Heyes, A. & Rickman, N. (1999). Regulatory dealing – revisiting the [h]arrington paradox. *Journal of Public Economics*, 72(3), 361–378.

Innes, R. (1999). Remediation and self-reporting in optimal law enforcement. *Journal of Public Economics*, 72(3), 379–393.

Innes, R. (2001). Violator avoidance activities and self-reporting in optimal law enforcement. *Journal of Law, Economics, and Organization*, 17(1), 239.

Kaplow, L. & Shavell, S. (1994). Optimal law enforcement with self-reporting of behavior. *The Journal of Political Economy*, 102(3), 583–606.

Konar, S. & Cohen, M. (1997). Information as regulation: The effect of community right to know laws on toxic emissions. *Journal of Environmental Economics and Management*, 32(1), 109–124.

Livernois, J. & McKenna, C. (1999). Truth or consequences-enforcing pollution standards with self-reporting. *Journal of Public Economics*, 71(3), 415–440.

Macho-Stadler, I. & Pérez-Castrillo, D. (2006). Optimal enforcement policy and firms' emissions and compliance with environmental taxes. *Journal of Environmental Economics and Management*, 51(1), 110–131.

Malik, A. (1993). Self-reporting and the design of policies for regulating stochastic pollution. *Journal of Environmental Economics and Management*, 24(3), 241–257.

Milliman, S. (1986). Optimal fishery management in the presence of illegal activity. *Journal of Environmental Economics and Management*, 13(4), 363–381.

Russell, C. (1990). Monitoring and enforcement. In P. Portney (Ed.), *Public policies for environmental protection* (pp. 243–274). Washington DC.

Sandmo, A. (2000). *The public economics of the environment.* Oxford: Oxford University Press.

Short, J. & Toffel, M. (2008). Coerced confessions: Self-policing in the shadow of the regulator. *Journal of Law, Economics, and Organization*, 24(1), 45–71.

Smith, V. (1969). On models of commercial fishing. *Journal of Political Economy*, 77(2), 181–198.

Sumaila, U. R., Alder, J., & Keith, H. (2006). Global scope and economics of illegal fishing. *Marine Policy*, 30, 696–703.

Sutinen, J. G. & Andersen, P. (1985). The economics of fisheries law enforcement. *Land Economics*, 61(4), 387–397.

APPENDIX

A Deriving aggregate catch levels

We start out by presenting the parametrization of the theoretical model used in the simulations. In the following subsections, we show how aggregate catch levels are calculated under traditional and self-report based enforcement.

A.1 Model parametrization

The parameter values used in the simulations are given in table A.1.

Table A.1: Parameter values

Parameter	Value	Description
p	0.5	Price (per unit)
f	1	Fine (per unit)
$[\underline{\alpha}, \bar{\alpha}]$	$[75, 125]$	Interval, cost parameter α
n	100	Number of fishing firms
m	20	Total number of inspections given by budget
h	$[0, 1.0217]$	Interval, intrinsic growth rate of resource stock
K	500	Carrying capacity of fish stock

Recall that the cost parameter α is uniformly distributed over the interval $[\underline{\alpha}, \bar{\alpha}]$. Based on the parameter values from table A.1, there are $n = 100$ agents with cost parameters ranging from $\underline{\alpha} = 75$ to $\bar{\alpha} = 125$.

The interval for the intrinsic growth rate h given in table A.1, represents the range of growth rates we analyze. Note that there is no variation in the growth rate, instead we evaluate the performance of the self-report based system over a range of intrinsic growth rates representing different types of fisheries. At low growth rates, there is a high need for enforcement to meet the aggregate catch objective. As we increase the growth rate, the need for enforcement declines until $h = \bar{h} = 1.0217$, when no enforcement is needed. In this case, the aggregate catch target is reached under open access.

A.2 Traditional enforcement system

Profits for compliant and non-compliant firms are given by equations (4) and (11). By solving the profit maximization problem of the firm for any value of α_i, it can be shown that firm level harvest is:

$$y_i^* = \begin{cases} \min\left(\frac{pX}{\alpha_i}, q\right) & \text{for } \alpha_i \geq \hat{\alpha} \\ \frac{X}{\alpha_i}(p - \gamma f) & \text{for } \alpha_i < \hat{\alpha}, \end{cases} \tag{A.1}$$

where $\hat{\alpha}$ is the value of the firm-specific cost parameter α for which a firm would be indifferent between compliance and non-compliance.

A.3 Self-report based enforcement system

To calculate aggregate harvest as a function of the self-reporting rebate when firms choose strategies A and B, we start out by analyzing optimal firm-level behavior. By maximizing the profit functions given above, we find that optimal catches are as follows:

$$y_a^* = \frac{X}{\alpha}(p - rf) \tag{A.2}$$

$$y_{b1}^* = \frac{X}{\alpha}(p - \gamma_1 f) \tag{A.3}$$

$$y_{b2}^* = q, \tag{A.4}$$

where subscripts a, $b1$, and $b2$ denote a firm choosing strategy A (in group 1), a firm choosing strategy B currently in group 1, and a firm choosing strategy B currently in group 2, respectively. By substituting catch response functions from equations (A.2-A.4) into equation (3) and adjusting for the long-run shares of strategy B firms that are in groups 1 and 2, an expression for aggregate catch can be found.

We can now calculate the value of α for which a firm is indifferent between strategies A and B, which we denote α_0. Strategy B, is relatively more attractive to more productive firms (low α_i) because their gains from not self-reporting excess catches in group 1 are greater than for less productive firms (with high α_i). Thus, if some firms prefer strategy B to strategy A it must be firms with low values of α_i.

We now derive the value of α that makes a firm indifferent between strategies A and B, which we denote α_0. The present value of all future payoffs for a firm following strategy A is:

$$EV_a = \sum_{t=0}^{\infty} \beta^t \pi_a^*(\alpha_i, X), \tag{A.5}$$

which can be rewritten:

$$EV_a = \frac{\pi_a^*(\alpha_i, X)}{1 - \beta}. \tag{A.6}$$

Correspondingly, the expected present value of all future payoffs for a firm following strategy B is:

$$EV_b = \sum_{t=0}^{\infty} \beta^t \pi_b^*(\alpha_i, X). \tag{A.7}$$

This can be rewritten as follows:[27]

$$
\begin{aligned}
EV_b &= \pi_{b1}^*(\alpha_i, X) + (1-\gamma_1)\beta EV_b + \gamma_1 \left(\sum_{t=0}^{u} \beta^t \pi_{b2}^*(\alpha_i, X) + \beta^{u+1} EV_b \right) \\
EV_b &= \frac{\pi_{b1}^*(\alpha_i, X) + \gamma_1 \left(\sum_{t=0}^{u} \beta^t \pi_{b2}^*(\alpha_i, X) \right)}{1-(1-\gamma_1)\beta - \gamma_1 \beta^{u+1}}. \qquad \text{(A.8)}
\end{aligned}
$$

The value of α_i that separates firms choosing strategy A from firms choosing strategy B can be identified by equating the present values of the two strategies ($EV_a = EV_b$) and is denoted α_0. We substitute in for the maximized profit functions, $\pi_a^* = \frac{X}{2\alpha}(p - rf)^2 + rfq$ and $\pi_{b1}^* = \frac{X}{2\alpha}(p - \gamma_1 f)^2 + \gamma_1 fq$, and obtain:

$$
\frac{\frac{X}{2\alpha_0}(p-rf)^2 + rfq}{1-\beta} = \frac{\frac{X}{2\alpha_0}(p-\gamma_1 f)^2 + \gamma_1 fq + \gamma_1 \sum_{t=0}^{u} \beta^t \left(pq - \frac{\alpha_0 q^2}{2X} \right)}{1-(1-\gamma_1)\beta - \gamma \beta^{u+1}} \qquad \text{(A.9)}
$$

Rearranging the expression yields the following second order equation in α_0:

$$
\begin{aligned}
&\alpha_0^2 (1-\beta)\gamma_1 \sum_{t=0}^{u} \left(\frac{\beta^t q^2}{X} \right) - 2\alpha_0 \left[\gamma_1 (1-\beta) \left(fq + \sum_{t=0}^{\infty} \beta^t pq \right) - rfq \left(1-(1-\gamma_1)\beta - \gamma_1 \beta^{u+1} \right) \right] \\
&+ X(p-rf)^2 \left(1-(1-\gamma_1)\beta - \gamma_1 \beta^{u+1} \right) - X(p-\gamma_1 f)^2 (1-\beta) = 0
\end{aligned} \qquad \text{(A.10)}
$$

Solving equation (A.10) gives the following:

$$
\alpha_0 = \frac{-B + \sqrt{B^2 - 4AD}}{2A}, \qquad \text{(A.11)}
$$

where A , B and D are defined as follows:

$$
\begin{aligned}
A &= (1-\beta)\gamma_1 \sum_{t=0}^{u} \left(\frac{\beta^t q^2}{X} \right), \\
B &= 2 \left[rfq \left(1-(1-\gamma_1)\beta - \gamma_1 \beta^{u+1} \right) - \gamma_1 (1-\beta) \left(fq + \sum_{t=0}^{\infty} \beta^t pq \right) \right], \\
D &= X(p-rf)^2 \left(1-(1-\gamma_1)\beta - \gamma_1 \beta^{u+1} \right) - X(p-\gamma_1 f)^2 (1-\beta).
\end{aligned}
$$

Firms with $\alpha_i \geq \alpha_0$ find it optimal to use strategy A, while firms with lower production costs (α_i) choose strategy B.

Finally, we calculate the aggregate catch response function under the assumption that u can be set high enough to ensure that all firms chose strategy A. This requires that u is set high enough for the inequality $\underline{\alpha} \geq \alpha_0(u)$ to hold, where $\alpha_0(u)$ is given by equation (A.11). We use the reaction function of strategy A firms

[27] We assume that firms take the current level of the stock, as well as all policy variables, as given when considering future operations and profits.

from equation (A.2). In addition we know the probability density function of the uniformly distributed variable α, which is $\frac{1}{\bar{\alpha}-\underline{\alpha}}$ (for $\underline{\alpha} \leq \alpha \leq \hat{\alpha}$). Given that there is a continuum of firms, total catches can be expressed as:

$$Y = \frac{nX\left(p - rf\right)}{\bar{\alpha} - \underline{\alpha}} \ln\left(\frac{\bar{\alpha}}{\underline{\alpha}}\right). \tag{A.12}$$