Let the Punishment Fit the Crime: Enforcement with Error

Indranil Chakraborty
Department of Economics
National University of Singapore
indro@nus.edu.sg

and

R. Preston McAfee
Yahoo! Research
preston@mcafee.cc

1. Motivation

Divergence of social and private interests is a standard feature of many economic situations. Market failures may be minimized or avoided by an appropriate choice of a tax or subsidy schedule that induces the individual to internalize the external costs and benefits of his action. Consider, for instance, the external costs generated as a car travels at different speeds on an expressway. Suppose that as the car drives at a speed x it imposes a net social externality $s(x)$. The external effect includes the danger to other drivers in the event of an accident as well as the possibility of an accident itself, both of which vary with speed. If the driver has a net benefit $B(x)$ from driving at a speed x then he can be induced to drive at the socially optimal speed that maximizes $B(x)$-$s(x)$ by a *penalty* or *Pigouvian tax* $p(x)$, which, if speed is observed, satisfies $p(x)$=$s(x)$ regardless of the specific functional form of $B(x)$.

The tax achieves its purpose when the speed x of the car is perfectly measured. If, however, the technology allows only an imperfect measurement of the speed of the car then the driver of the car will not generally choose the socially optimal speed. As an example, suppose $s(x)$ is convex, and Y is the unbiased but imperfect measure of speed. Then $E[s(Y)\,|\,x] > s(x)$. Thus, a penalty function $p(y)$=$s(y)$ based on the observed speed y will usually result in a choice of speed which is different from the socially optimal speed.

The natural question then is whether a penalty function based on the imperfectly observed speed y can force the agent to internalize the cost $s(x)$ for driving at speed x regardless of his benefit function. We consider the following problem: Suppose a tax or subsidy function $s(x)$ associated with an action level x that is measured perfectly achieves a certain objective.[1] For ease of reference, let us call $s(x)$ a (*social*) *externality function*. The action x is only imperfectly observed as a random variable Y whose distribution depends on the true action x.

[1] The function $p(x) = s(x)$ could be a Pigouvian tax, or alternately, $s(x)$ could be viewed as a price/penalty function that can help the market to co-ordinate towards the optimum described by Coase (1960). Generally, we will treat $s(x)$ to be a given target tax function.

We examine how and when a tax or subsidy function $p(y)$ of the observed signal gives rise to the same choice of action by the agent as the tax or subsidy function $s(x)$ regardless of the nature of his benefit function. In this case, there is a Pigouvian solution to the problem of externalities even when the behavior is observed with error.[1]

We will show that in a broad class of circumstances, there is indeed a solution to the problem for risk neutral agents. What is needed is that there is enough information in the signal to separate distributions of behaviors. If two distinct distributions over behaviors produce the identical distribution of signals, then it is not possible to distinguish these distributions with a penalty function. If the function s separates the distributions, then no solution can possibly exist, because the signal does not distinguish distributions that the penalty function separates. Therefore separating distributions is necessary. The remarkable fact is that it is also sufficient. Moreover, the same condition conveniently works for both finite and continuous state spaces.

Interest in the implementability of desired outcomes by penalty functions is not new. Pigou (1952) suggested that forcing an agent to internalize the damage he causes by taxing him the amount of the damage would take the market toward efficiency. Coase (1960) critiqued that the Pigouvian scheme for providing the wrong incentives. However, interpreted appropriately (so that the price suggested by Coase is the tax) the Pigouvian principle continues to hold in his examples. In fact, Sandmo (1975) showed that in the absence of government revenue requirements the Pigouvian tax implements the first best, and when there is a revenue requirement the Pigouvian principle extends appropriately. In all these cases, of course, the agent's action is perfectly observable and there is no uncertainty about any element of the model.

[1] Note that the role of y is purely that of a signal on the true action x that actually gives rise to the externality. If the observable y completely determines the level of externality then we are back in the perfect observability case.

Kwerel (1977), Dasgupta, Hammond and Maskin (1980), Duggan and Roberts (2002), among others, assume that there are a finite number of polluting firms with costs not observable to the regulator. They show that the first best outcome (that arises when costs are perfectly observable) are implementable in equilibrium. In these models, the emission level by each firm is perfectly observable. In reality, the total pollution is often not observable. While the rate at which a car pollutes can be observed through some tests, the amount of gasoline burnt or the number of city miles are not easy to observe. Montero (2005) assumes that the emission level is not observable but the emission rate is. The first best outcome cannot be implemented in this framework. In contrast, we consider a situation where the agent cannot perfectly control what the principal observes, that is, his action gives rise to a stochastic signal.

The principal-agent formulation of our problem demands a few words about its relationship with the agency literature. What we consider here is different from the standard agency problems in that we introduce externality to the standard problem and require that a single penalty function implement the target externality function for all relevant action levels. As a result, even if the principal and the agent are risk-neutral a non-linear externality function makes it a different type of problem. The difference in the nature of the problems is most easily seen by observing that in the moral hazard model (cf. Holmstrom, 1979), the first best is always implementable when the agent is risk neutral which is not true in our setting.[2] Thus, our results sit nicely between the standard agency models and the literature on implementation of tax functions.

Most mechanism design solutions entail complicated mechanisms that are very sensitive to the underlying description of the environment. This sensitivity is especially extreme when correlation is exploited to mitigate incentive constraints. In contrast to most related literature, we prove an approximation to the solution that has quite modest informational requirements (means and variances of the error function), which in many applications are knowable.

[2] We will discuss our result in the agency context in greater detail below.

Consequently, our approach is plausibly applicable in real world settings, such as the speeding example discussed above.

2. The Model

Let x denote the action chosen by a *risk neutral* agent and Y be the associated (publicly observable) signal. Y is distributed according to density $f(y|x)$ conditional on x, $x \in \mathbb{X}, y \in \mathbb{Y}$. The function f completely describes the relevant *(stochastic) environment* of the situation. In what follows, we present our results for the case where x and y take the same set of values, i.e., $\mathbb{Y}=\mathbb{X}$. However, with a little extra work it can be checked that the relevant results, i.e., the results and discussions of section 4, continue to hold when $\mathbb{Y} \neq \mathbb{X}$. The action x generates a private return $B(x,\theta)$ for the agent with a privately known type θ.[3] The agent has quasi-linear utility

$$U(B(x,\theta),t) = B(x,\theta) - t.$$

from taking action x and making a payment t, if any. We assume that transfers are non-distortionary.

Let us denote by $s(x)$ the externality (i.e., tax) function that the regulator wants to implement. We are interested in studying when and how a penalty function $p(y)$ makes the agent face the exact same problem that he would with perfect observation regardless of $B(x,\theta)$.[4] If the action is measured imperfectly then a function $p(y)$ which we refer to as the *penalty function* of the imperfect observation y will be said to *implement* $s(x)$ if

$$U(B(x,\theta), s(x)) = E[U(B(x,\theta), p(Y)) \mid x]$$

for all x.

[3] In what follows we do not need to directly use θ in the analysis. The notation is to highlight the fact that agents are assumed to have private information on how their utilities depend on the action. The notation also allows convenient exposition if one extends our analysis to other market structure or to the agency problems.

[4] Note, however, that the function $s(x)$ that we seen to implement may itself depend on $B(x,\theta)$.

Quasi-linearity of utility implies that $p(y)$ implements $s(x)$ if

$$B(x,\theta) - s(x) = B(x,\theta) - E[p(Y)\,|\,x],$$

that is,

$$s(x) = E[p(Y)\,|\,x].$$

If $s(x)$ is first-best and $p(y)$ implements $s(x)$, then $p(y)$ is first-best in spite of imperfect observability of action.

For expositional clarity, we consider the case where x takes finite values and the case where x takes continuous set of values over an interval. Our results will hold more generally. In the *finite actions* case we assume x takes values $1, 2, \ldots, n$ so that $f(i\,|\,j)$ is the probability that action i is observed when action j is undertaken. In this case where the agent is risk-neutral we have that $p(y)$ implements $s(x)$ if

$$\sum_{i=1}^{n} p(i) f(i\,|\,j) = s(j)$$

for all j. With *continuous actions* we have under risk-neutrality that $x, y \in [a,b]$, and $p(y)$ implements $s(x)$ if

$$\int_a^b p(y) f(y\,|\,x) dy = s(x)$$

for all x. The reason for considering both cases is that a finite dimensional approach with finite matrices makes the analysis more manageable. Conclusions can be drawn more readily. However, the intuition from the finite dimensional case does not extend to the infinite dimensional analysis where a continuous set of actions is undertaken. Also, the standard literature on both externality and the basic agency problems primarily deal with the continuous models. Thus, it is necessary to treat the continuous actions case separately. The infinite dimensional analysis does not permit the ease of working with matrices, still, we are able to verify some of the key findings from the finite actions case.

Some Notation. Throughout this paper we will denote by μ the uniform probability measure on the interval $[a,b]$. We denote by A the expectation operator on p defined by $\left(Ap\right)(x) = E[p(Y)\,|\,x]$. In the finite action case, A will denote the $n \times n$ matrix which is a standard linear transformation from R^n to R^n. When the actions are continuous it is the integral operator (a continuous linear transformation) from $L_2(\mu)$ to $L_2(\mu)$.

In light of the fact, that we work with a $L_2(\mu)$ space, the equalities and other similar relationships must be interpreted as *almost everywhere* μ wherever appropriate.

3. Can the Crime Fit the Punishment?

When does using the social externality function as a penalty work even in the presence of error? Our first result shows that in the infinite dimensional space of all possible externality functions only a "negligible" sub-collection of externality functions $s(x)$ can be implemented by $p(y) = s(y)$.

Proposition 1. Consider the continuous action case. In any given environment there are at most a finite number of linearly independent penalty functions that can be implemented straightforwardly by $p(y) = s(y)$. In other words, given $f(y\,|\,x)$ the collection of $L_2(\mu)$ functions that can be implemented straightforwardly belong to a finite dimensional space.

Proof. See Appendix.

Of course, if there is no error, the penalty function $p \equiv s$ implements the externality function *s*. However, not only is zero error sufficient, it is also necessary for all *s* to implement itself.

Proposition 2. In a given environment all penalty functions $s(x)$ can be implemented by a penalty $p(y)=s(y)$ if and only if actions are observable without errors in that environment.

Proof. See Appendix.

We now turn to positive results.

4. Existence

Most externality functions $s(x)$ cannot be implemented directly by penalty functions $p(y)=s(y)$ when action y can only be observed imperfectly. We next examine environments where an externality function can be implemented by *some* penalty function. In the continuous actions case we need to also examine the extent to which it is possible to come arbitrarily close to implementing the externality function in order to state the necessary and sufficient condition.

Definitions. (i) We say that externality function $s(\cdot)$ is *approximately implementable* if there is a sequence of penalty functions $\{p_n(\cdot)\}$ such that $E\left[p_n(Y)\,|\,X=x\right]\to s(x)$ in $L_2(\mu)$.[5]

(ii) f *separates* distributions G_1 and G_2 of *X* if

$$\int_x f(\bullet\,|\,x)dG_1(x)\neq\int_x f(\bullet\,|\,x)dG_2(x)$$

in $L_2(\mu)$.[6]

(iii) s *separates* distributions G_1 and G_2 of *X* if

$$E[s(X)\,|\,X\sim G_1]\neq E[s(X)\,|\,X\sim G_2].$$

[5] Note that $L_2(\mu)$ convergence implies convergence in probability. Therefore, if there is a sequence $\{p_n(\bullet)\}$ such that $E[p_n(Y)\,|\,X=x]\to s(x)$ in $L_2(\mu)$ then there exists a subsequence $\{p_{n_k}(\bullet)\}$ of penalty functions with the property that $E[p_{n_k}(Y)\,|\,X=x]\to s(x)$ almost surely. In other words, our results will not change if we adopt the stronger notion of almost sure convergence for approximate implementability.

[6] When *x* is discrete the integral is substituted by the summation.

Proposition 3 (Existence – infinite actions). A penalty function p in $L_2(\mu)$ implements or approximately implements s if and only if f separates distributions on X that are separated by s.

Proof. See Appendix.

The only if portion of proposition 3 is quite straightforward. If *f* doesn't separate two distributions of *x*, say *G* and *H*, then for any *p*, $E[p(X)\,|\,X \sim G_1] = E[p(X)\,|\,X \sim G_2]$ because both G_1 and G_2 give rise to the same distribution of *Y*. The remarkable fact is that the condition is sufficient – if *f* separates any distribution that *s* separates, then there exists a function *p* that approximately implements *s* in the stochastic environment.

When the set of actions is finite, existence is much more straightforward.

Proposition 4 (Existence – finite actions). Suppose that the set of possible actions is finite. Then $s(x)$ is implementable if and only if A separates distributions on X that are separated by s.[7]

Proof. First observe that the range of a finite dimensional linear transformation is closed. Hence by the Fredholm alternative theorem (see Appendix) the necessary and sufficient condition for the equation $Ap = s$ to have a solution is that.

$$s^T z = 0 \text{ whenever } A^T z = 0$$

By a construction similar to the infinite-dimensional case (see Appendix) we can assume w.l.o.g. that z defined above can be written as a difference of two non-identical probability mass functions, i.e. $z = q - \tilde{q}$ for some probability distributions q and $\tilde{q}$. Then we can restate the above necessary and sufficient conditions as follows: For any pair of distributions q and $\tilde{q}$

[7] We will refer to this necessary and sufficient condition for implementability as the *separation condition* later in this section.

$$\text{whenever } s^T q \neq s^T \tilde{q} \text{ we have } A^T q \neq A^T \tilde{q}$$

i.e., A separates distribution on X that give rise to distinct expectations $E_q[s(X)] \neq E_{\tilde{q}}[s(X)]$ ■

Proposition 4 is the finite dimensional analog of Proposition 3 and has the same interpretation. Given the above result, it is generally straightforward to describe situations in the finite-actions case where an externality function cannot be implemented. The infinite-dimensional case is less straightforward. Situations where an exact penalty function cannot exist in the infinite-action case can arise very naturally. If the conditional distribution $f(y|x)$ is continuous in x and y discontinuous penalty functions are not exactly implementable. Of course, discontinuous functions may be obtained as limits of continuous functions.

The implementable externality functions are dense in $L_2(\mu)$ only under some stronger conditions on the environment. By the modified Fredholm alternative theorem (see Appendix) the implementable functions are dense if and only if

$$\int_0^1 f(y|x)g(x)dx = 0 \ \& \ g(x) \in L_2(\mu) \Rightarrow g(x) = 0.$$

By a construction similar to that in the proof of Proposition 3 (see Appendix) it follows that the implementable function are dense in $L_2(\mu)$ if and only if for two densities $g_1(x)$ and $g_2(x)$

$$\int_0^1 f(y|x)g_1(x)dx = \int_0^1 f(y|x)g_2(x)dx$$

implies $g_1(x) = g_2(x)$, i.e. *f* separates every pair of distinct distributions.[8]

[8] This is, in fact, an infinite dimensional counterpart of the individual full rank condition of Fudenberg, Levine and Maskin (1994).

Existence of transfer functions in agency problems

Formulated in the principal-agent framework, our results can be placed naturally in the context of the standard agency literature. The problem of implementing a target expected payment function arises in Cremer and McLean (1988) for the discrete problem, and McAfee and Reny (1992). The main distinction is that both of these papers utilitize a menu of contracts, and hence can use selection by the agent as an indication of the agent's type. In contrast, we do not anticipate using a menu for the externalities problem because that requires contracting in advance. Contracting in advance is implausible in the case of speeding and absurd in the case of intoxicated drivers.

Models of pure moral hazard (cf. Holmstrom, 1979) are closer to our model in spirit than the rent extraction literature; however these models typically aim to induce the appropriate behavior at a single point; in our notation, it would be as if the function B were given and identical for all types of agents. This is a substantially easier problem to solve, because the goal is not to eliminate the randomness but to induce the agent to choose the targeted action. A second difference is that, in the agency problem, the principal cares about the outcome (i.e. the observed signal), not the (hidden) effort. In contrast, in the externalities problem, the natural assumption is that the principal cares about the hidden action, not about the signal. Positing that the principal cares about the outcome makes sense when the outcome is, say, encyclopedia sales. In contrast, when the observed outcome is performance on an exam and the hidden action is teacher effort, our model is the more sensible one. Moreover, in the pure moral hazard problem, when the agent is risk neutral the solution is to 'sell the agency' to the agent. That is not possible in the problem of externalities because the value created depends on an unobservable.

The need to implement a function at multiple points arises when the agent has multiple types. Such a situation arises when the moral hazard model also involves private agent types, e.g. Laffont and Tirole (1986) and the single agent special case of McAfee and McMillan (1987). These models have both moral hazard and adverse selection, and quasi-linear utilities. The models are distinguished by whether the agent's costs are separable in type and action. As in the rent extraction problem, the agency problems approach the solution with a menu of contracts, that is, they contract in advance. Here, too, the principle cares about the (observable) outcome. In our problem, the signal is just a proxy. In this alternate environment, we answer the question of when the incentive-efficient solution can be implemented without using a menu of contracts, but with a single contract offered to all. To our knowledge, the alternate environment (where the principal cares about the hidden action, not the signal) has not been studied, nor the solution characterized.

The separation condition

Conditions for solving systems of equations and inequations are widely used in economics and econometrics. The conditions that come closest to the finite action version of our model, in the sense that these conditions have also been used in the context of separation of public signals, are those in Fudenberg, Levine and Maskin (1994), FLM, henceforth, in the context of Folk Theorem.[9] The *individual full rank* condition requires that rows of conditional probability distributions over signals given a player's actions be linearly independent. The *pairwise full rank* is a similar full row rank condition on stacked matrices for pairs of players.[10] These conditions are applicable only when the number of signals is no less than the number of actions to each player. A weaker condition is given by the *pairwise identifiability* condition that, when used with implementability, is also a type of full row rank condition.[11] Thus, this condition also requires the signal space to be sufficiently rich.

[9] We thank Larry Samuelson and the referees for pointing out these and other related conditions.

[10] One of the rows of the matrix of conditional probabilities drops out due to linear dependence, leaving the rest of the rows linearly independent.

[11] The implementability condition may cause some of the incentive constraints that must be solved in their problem hold with strict inequality, and thus drop out of the system of equations whose solutions they look for. In that case, what the system is left with is a system of linearly independent equations.

The biggest difference between the above conditions and our separation condition comes from the nature of applications. The Folk Theorem is derived for multi-player games, whereas our problem is to implement payment functions in a single agent model. As a result, FLM have to work with beliefs spanning over pairs of players unlike our problem where we have to focus on a single agent's beliefs. The steps for deriving the Folk Theorem involve incentive (inequality) constraints and the problem is to show the existence of continuation payoffs that satisfy those constraints. Our problem is to check the existence of transfer schedule whose expectation is *equal* to a given target transfer function of actions. Moreover, we consider an environment where an agent has private types where the Folk Theorem does not hold.

The conditions also differ because of the approaches taken. The FLM conditions try to guarantee the *decomposition* of payoff profiles along *all regular hyperplanes*. Our objective, however, is to describe the complete set of transfer functions that are implementable in a given environment. Therefore, while FLM have to look for some sufficiency conditions, we must look for necessary *and* sufficient conditions. In short, the problem that we solve is completely different from that solved by FLM, which makes the machinery used to solve the problems quite different. The "full row rank" type conditions of FLM are weaker than invertibility of the coefficient matrix for a system of linear equations. Under such a condition, the system always produces a solution. Our separation condition is equivalent to a consistency condition for a system of linear equations. The coefficient matrix need not have linearly independent rows, so the existence of the solution depends on whether or not the intercept terms of the equations are consistent with the coefficients. One of the important consequences of this difference is that our conditions are applicable even if the signal space is not rich enough. The FLM conditions, however, require that the signal space be sufficiently rich relative to the action space. For instance, the individual full rank condition requires that the number of signals be at least as many as the number of actions. The pairwise identifiability condition weakens this requirement to some extent. Still, depending on the situation it may not allow the signal space to be much smaller than the action space.

If the Folk Theorem has to hold under consistency conditions like ours on FLM's signal and payoff generating environment then one needs to obtain a result that shows that for each discount level there exists a convenient set (possibly a hyperplane) that contains the continuation payoffs. It is, however, not clear whether that is a tractable exercise. One could also ask whether the Folk Theorem with imperfect monitoring could be extended to the continuous action case as done here. The individual full rank condition for the continuum case is precisely the following: *f* separates every pair of distributions on the action space that differ on a measure non-zero set (see footnote 8). It is, however, not obvious that the other conditions of FLM have a natural extension for continuous actions that would be useful for proving the Folk Theorem.

5. Construction of a Penalty Function

Existence theorems are often cold comfort to someone who needs to use a construct. In this section we provide a method of constructing the penalty function in the case of multiplicative errors.

Suppose that $Y = x\varepsilon$, $\varepsilon \geq 0$ and that the externality function s is analytic, so that it can be expressed by a Taylor series:

$$s(x) = \sum_i a_i x^i .$$

Let

$$p(y) = \sum_i \frac{a_i y^i}{E\varepsilon^i}$$

Then

$$E[p(Y) \mid x] = E\sum\left[\frac{a_i (x\varepsilon)^i}{E\varepsilon^i} \mid x\right] = \sum a_i x^i = s(x)$$

whenever the series converges absolutely at each point in the support (so that the expectation can be taken inside the summation). This is a full solution for analytic s for the multiplicative case. In addition, for any continuous s on a compact set of x, we can approximate arbitrarily closely by first approximating s with an analytic function and then using the p for the analytic function.

How does the penalty compare to the externality? Suppose that $Y=x\varepsilon$, for $\varepsilon\geq 0$, and the error is either unbiased or upward biased, i.e. $E\varepsilon \geq 1$ and that all the derivatives of $s(x)$ are nonnegative, i.e. $a_i \geq 0$, as arises with the exponential function. Then the penalty is less than the externality, i.e. $p(y) \leq s(y)$, because $E\varepsilon^i \geq (E\varepsilon)^i \geq 1$. Thus $p(y) = \sum \frac{a_i y^i}{E\varepsilon^i} \leq \sum a_i y^i = s(y)$.

The multiplicative error model provides a construction for additive errors for many externality functions. Consider the additive error model $Y = x + \varepsilon$ and assume that $s(\log(z))$ is an analytic function of z, and that $Ee^{j\varepsilon} < \infty$ for all $j < \infty$. Because $s(log(\bullet))$ is analytic, we can express $s(\log(z)) = \sum_{j=0}^{\infty} a_j z^j$. Consequently, setting $x = \log(z)$, $s(x) = \sum a_j (e^x)^j = \sum a_j e^{jx}$. It is readily verified that $p(y) = \sum \frac{a_j e^{jy}}{Ee^{j\varepsilon}}$ satisfies $E\left[p(Y)\right] = s(x)$.

6. Small Errors

So far, our main results – existence and a construction under multiplicative or additive errors – tell us little about the actual nature of the penalty function. Under the hypothesis that errors are small, we can provide a sharper characterization of the penalty functions. Imagine that a car going at 100 mph generates an externality s, would the corresponding penalty function charge more than s or less upon observing a speed of 100 mph? A 10 mph increase in speed at 100 mph could presumably do much more damage than the same increase would do at 70 mph. Would the corresponding penalty function reflect that?

When the error is sufficiently "local" in nature, the Taylor expansion allows close approximation of the penalty function given the externality function. Moreover, the higher order terms can be ignored for broad classes of situations and the penalty function compared with the externality function based on its lower order derivatives. Small errors are reasonable in many settings where large errors would prohibit legal remedies.

In what follows we let Y be the observation by the regulator and provide a formula for approximating the penalty function that implements an externality function $s(x)$ when the associated observational error is small. Let $\mu(x) = E[Y \mid x]$ and $\sigma^2(x) = E[(Y - \mu(x))^2 \mid x]$.

Proposition 5. Suppose that the family of random variables $\tilde{Y}_b \equiv \mu(x) + b(Y - \mu(x))$ with distribution functions $G_b(\tilde{y} \mid x) = F((\tilde{y} - (1-b)\mu(x))/b \mid x)$ separate any pair of distributions on X for all small $b > 0$ that are separated by $s(\bullet)$. Assume also that $\mu(x)$ is twice differentiable and monotonic.[12] If the σ²(x) is small, the penalty function is approximated by[13]

$$p(z) \approx s(\mu^{-1}(z)) - \frac{1}{2}\sigma^2 \frac{d^2}{dz^2} s(\mu^{-1}(z))$$

Proof. See Appendix.

In particular, when the signal distribution is unbiased, we have

$$p(x) \approx s(x) - ½s''(x)\,\sigma^2$$

[12] The two assumptions are necessary to apply the Taylor approximation result at $\mu(x)$, i.e. to apply the approximation on a function at the point it is necessary for the function to exist in a small neighborhood. The assumptions guarantee precisely that.

[13] Obviously, the smaller the error the better the approximation.

It is now much easier to relate the penalty to the externality function. For instance, when the observation is unbiased ($\mu(x) = x$) and the error ($\sigma^2(x)$) is small, the penalty function is smaller or larger than the externality function depending on whether s is convex or concave.

7. Conclusion

Given a penalty function which implements a social objective, this paper examines the possibility of implementing the social objective when the action is observed with error. Provided that the signal is informative in the sense that it separates distributions of actions and agents are risk neutral, the social objective remains implementable even with observational error. In addition, when errors are small, there is a closed form second-order approximation for the penalty function that depends only on first and second moments and two derivatives of the externality function. The formula is applicable when activity is measured reasonably accurately, which is necessary for a fair implementation. This formula is simple enough to lend itself to actual implementation.

In our formulation of the problem we have kept the model as context-free as possible insofar as the market structure is concerned. The principal-agent framework in which we posed our problem immediately fits the case of an externality producing monopoly firm. The analysis holds for other market structures, as well, even if with some modification.

The analysis does not apply in the form of first-best implementation when the agent is risk averse. The difficulty arises due to the welfare effect of the redistributive role of tax function when the agent is risk averse. If the agent is risk averse and the penalty is a function of a stochastic signal, the socially optimal penalty function depends on the conditional distribution of the signal. Implementing a function that is first-best when the action is observed perfectly may not be optimal in the stochastic environment due to the *risk cost* to the society.[14]

Implementation of a target function $s(x)$ in itself may, however, be possible. For instance, when the agent has constant absolute risk aversion utility of the form

$$\frac{1}{\lambda}\left(1-e^{-\lambda(B(x)-s(x))}\right).$$

Steps similar to those under risk neutrality show that the penalty function of y that forces the agent to behave the same as the externality function $s(x)$ is approximated by

$$p(y)\approx s(y)-\sigma^2\frac{1}{2}\left(\lambda s'(y)^2+s''(y)\right)$$

for small observational error $\varepsilon = Y-\mu(x)$ satisfying $E\varepsilon = 0$. Of course, $p(y)$ is generally not optimal any more.

[14] We thank Ilya Segal for pointing out this difficulty in analyzing risk aversion.

Appendix: Proofs

Proof of Proposition 1. First note that the integral operator A is compact. Therefore, if there is a non-zero function $s(x)$ such that $As = s$ then the operator has an eigenvalue 1. The result then follows upon applying Proposition II.4.13 of Conway (1990) and observing that the space of functions s for which $(A - I)s = 0$ is at most finite dimensional. ■

Proof of Proposition 2. The "if" part follows straightforwardly. To show the "only if" part, let us consider I , the identity operator $Is = s \ \forall s \in V$ where V is the relevant (finite or infinite dimensional) space of penalty functions. Now, $As = s$ for all $s \in V$ implies that $(A - I)s = 0$ for all $s \in V$. This implies that $\|A - I\| = 0$ where $\|\cdot\|$ is the norm for the space $B(V)$ of bounded linear operators on V. Thus we have $A = I$ which completes the proof. ■

The proof of propositions 3 and 4 will use the following result:

Fredholm Alternative Theorem (cf. Keener, 1988). If A is a bounded linear operator in Hilbert space H with a closed range, the equation $Ap = s$ has a solution if and only if $\langle s, u \rangle = 0$ for every u in the null space of the adjoint operator A^*.

Definition. We say that the equation $Ax = b$ has an *approximate solution* if there exists a sequence $\{x_n\}_{n=1}^{\infty}$ in $L_2(\mu)$ with $Ax_n \to b$.[15]

A Modified Fredholm Alternative Theorem. Let A be a compact linear operator. Then $Ax = b$ has either a solution or an arbitrarily close approximate solution if and only if

$$\langle v, b \rangle = 0 \text{ for all } v \text{ satisfying } A^* v = 0 .$$

[15] Note that $\{x_n\}_{n=1}^{\infty}$ may or may not be convergent or even have an accumulation point.

Moreover, all solutions in the equation $Ax = b$ are exact if and only if $\operatorname{ran} A$ is finite dimensional.

Proof. *Only if part*. Suppose that $Ax = b$ is at least approximately solvable. Then there exists a sequence $\{x_n\}_{n=1}^{\infty}$ possibly all identical such that $b_n \equiv Ax_n \to b$.
This implies that for v satisfying $A^* v = 0$

$$\langle v, b \rangle = \lim_n \langle v, b_n \rangle = \lim_n \langle v, Ax_n \rangle = \lim_n \langle A^* v, x_n \rangle = \lim_n 0 = 0$$

If part. Suppose $\langle v, b \rangle = 0$ for all v satisfying $A^* v = 0,$ but $Ax = b$ does not have even an approximate solution. Then $b = b^r + b^o$ where b^r is in the closure of the range of A and b^o is in its orthogonal subspace. Therefore, $\langle b^o, Ax \rangle = 0\ \forall x$ so that $\langle A^* b^o, x \rangle = 0\ \forall x$ which implies that $A^* b^o = 0$. Now using the hypothesis of this part we have $\langle b^o, b \rangle = 0$ which implies $\langle b^o, b^o + b^r \rangle = 0$, that is $\langle b^o, b^o \rangle + \langle b^o, b^r \rangle = 0$. Since b^o is orthogonal to b^r this implies that $\langle b^o, b^o \rangle = 0$ or $b^o = 0$.
Hence, $b \in \operatorname{cl}(\operatorname{ran} A)$, i.e. $Ax = b$ either has an exact solution or an approximate solution.

To prove the second part of the result observe that by Problem 7.1.1 of Abrahamovich and Aliprantis, a compact operator has a closed range if and only if its range is finite dimensional. Next we show that $Ax = b$ has only exact solutions if and only if $\operatorname{ran} A$ is closed. The if part follows from the original Fredholm Alternative Theorem (see above). To see the second part, suppose $\operatorname{ran} A$ is not closed. Then there exists a sequence $\{x_n\}_{n=1}^{\infty}$ such that $b \equiv \lim_n Ax_n$ is well defined and $b \notin \operatorname{ran} A$. Thus $Ax = b$ is only approximately solvable. ■

Proof of Proposition 3.

The modified Fredholm alternative theorem implies that a necessary and sufficient condition for at least an approximate solution to the equation

$$\int_a^b p(y)\, f(y \mid x)dy = s(x)$$

to exist is that

$$\int_a^b s(x)u(x)dy = 0 \text{ whenever } \int_a^b f(y \mid x)u(x)dy = 0$$

Now suppose a $L_2(\mu)$ function $u(x)$ satisfies $\int_a^b f(y \mid x)u(x)dx = 0$ and does not vanish over some positive μ measure subset. Define

$$v(x) = \left|u(x)\right| + u(x) \text{ and } w(x) = \left|u(x)\right|.$$

Then v and w are non-negative functions satisfying $u(x) = v(x) - w(x)$. Also, $\mu(x : u(x) = 0) < 1$ implies that $v(x) \neq w(x)$ over a set with positive measure.

Next $\int_{\underline{a}}^{\bar{a}} f(y \mid x)u(x)dx = 0$ implies that

$$\int_a^b \int_a^b f(y \mid x)u(x)dxdy = 0 \;\Rightarrow\; \int_a^b u(x) \int_a^b f(y \mid x)dydx = 0 \;\Rightarrow\; \int_a^b u(x)dx = 0$$

i.e.

$$\int_a^b v(x)dx = \int_a^b w(x)dx = K \text{ (say)}$$

Since v and w are non-negative functions satisfying $v(x) \neq w(x)$ on a set with positive measure, $K > 0$. Thus

$$\tilde{v}(x) = \frac{v(x)}{K} \text{ and } \tilde{w}(x) = \frac{w(x)}{K}$$

are probability density functions satisfying

$$\int_a^b f(y\,|\,x)\tilde{v}(x)dx = \int_a^b f(y\,|\,x)\widetilde{w}(x)dx.$$

Thus the necessary and sufficient condition above can be restated as that for two densities $\tilde{v}(x)$ and $\widetilde{w}(x)$

$$\int_a^b s(x)\tilde{v}(x)dy = \int_a^b s(x)\widetilde{w}(x)dy \text{ whenever } \int_a^b f(y\,|\,x)\tilde{v}(x)dy = \int_a^b f(y\,|\,x)\widetilde{w}(x)dy$$

In other words, $f(y\,|\,x)$ separates any pair of densities $\tilde{v}(x)$ and $\widetilde{w}(x)$, i.e.

$$\int_a^b f(y\,|\,x)\tilde{v}(x)dy \neq \int_a^b f(y\,|\,x)\widetilde{w}(x)dy,$$

whenever $\tilde{v}(x)$ and $\widetilde{w}(x)$ give rise to separate expectations for $s,$ i.e.

$$\int_a^b s(x)\tilde{v}(x)dx \neq \int_a^b s(x)\widetilde{w}(x)dx. \blacksquare$$

Proof of Proposition 5

Let Y be the observation with a conditional distribution $F(y\,|\,x)$. Define $\varepsilon = Y - \mu(x)$, which has a distribution $H(\varepsilon\,|\,x) = F(\varepsilon + \mu(x)\,|\,x).$ The corresponding density is $h(\varepsilon\,|\,x).$ Recall that $\mu(x)$=E[Y|x] and $\sigma^2(x)$=E[(Y—$\mu(x))^2$|x] so that $E[\varepsilon\,|\,x] = 0$ and $E[\varepsilon^2\,|\,x] = \sigma^2(x).$

For any $b > 0$ but small let $\tilde{p}(\bullet, b)$ solve (suppressing the limits in the integrations)

$$\int \tilde{p}(z + b\varepsilon, b)h(\varepsilon\,|\,x)d\varepsilon = s(\mu^{-1}(z)) \qquad \text{(P5.1)}$$

Our hypothesis guarantees that the functions $\tilde{p}(\bullet, b)$ exist for all small $b > 0.$ Existence in the case of $b = 0$ is, of course, immediate from the monotonicity of $\mu(x).$

Our target is the solution at $b=1$. Note that $p(z,0)=s(\mu^{-1}(z))$, so that $p_1(z,0)=\frac{d}{dz}s(\mu^{-1}(z))$ and $p_{11}(z,0)=\frac{d^2}{dz^2}s(\mu^{-1}(z))$.

Taking the derivative with respect to b of both sides of equation (P5.1) above we have

$$\int\left[\tilde{p}_1(z+b\varepsilon,b)\varepsilon+\tilde{p}_2(z+b\varepsilon,b)\right]h(\varepsilon\,|\,x)d\varepsilon=0\,,$$

which at $b=0$ gives

$$\tilde{p}_1(z,0)\int\varepsilon h(\varepsilon\,|\,x)d\varepsilon+\tilde{p}_2(z,0)\int h(\varepsilon\,|\,x)d\varepsilon=0$$

or, $\tilde{p}_2(z,0)=0$. This also implies $p_{12}(z,0)=0$.

Taking the second derivative with respect to b of both sides of equation (P5.1), we have

$$\int\left[\tilde{p}_{11}(z+b\varepsilon,b)\varepsilon^2+2\tilde{p}_{12}(z+b\varepsilon,b)\varepsilon+\tilde{p}_{22}(z+b\varepsilon,b)\right]h(\varepsilon\,|\,x)d\varepsilon=0\,.$$

Setting $b=0$,

$$\int\left[\tilde{p}_{11}(z,0)\varepsilon^2+2\tilde{p}_{12}(z,0)\varepsilon+\tilde{p}_{22}(z,0)\right]h(\varepsilon\,|\,x)d\varepsilon=0$$

or,

$$0=\tilde{p}_{11}(z,0)E[\varepsilon^2\,|\,x]+\tilde{p}_{22}(z,0)$$

Hence

$$\begin{aligned}\tilde{p}_{22}(z,0)&=-p_{11}(z,0)E[\varepsilon^2\,|\,x]\\&=-\sigma^2\frac{d^2}{dz^2}s(\mu^{-1}(z))\end{aligned}$$

Now we use the second order approximation on the first argument of $p(x,b)$:

$$\begin{aligned}\tilde{p}(z,b)&\approx s(\mu^{-1}(z))+b\tilde{p}_2(z,0)+\frac{1}{2}b^2\tilde{p}_{22}(z,0)\\&=s(\mu^{-1}(z))-\frac{1}{2}b^2\sigma^2\frac{d^2}{dz^2}s(\mu^{-1}(z))\end{aligned}$$

At b=1,

$$\tilde{p}(z,1) \approx s(\mu^{-1}(z)) - \frac{1}{2}\sigma^2 \frac{d^2}{dz^2} s(\mu^{-1}(z))$$

It is straightforward at this point to see that scaling the error down and scaling b up in the same amount keeps the entire calculation the same. Hence, $b=1$ is without loss of generality and we have the result. ■

References

Coase, R.H., “The problem of social cost,” *The Journal of Law and Economics*, vol. 3, 1-44, 1960.

Conway, J.B., *A Course in Functional Analysis*, second edition, Springer-Verlag New York, Inc., 1990.

Cremer, K., and R. McLean, “Full extraction of the surplus in Bayesian and dominant strategy auctions,” *Econometrica* (56), 1988, 1247-1257.

Dasgupta, P., P. Hammond, and E. Maskin, “On imperfect information and optimal pollution control,” *Review of Economic Studies*, (XLVII), 1980, 857-860.

Duggan, J. and J. Roberts, “Implementing the efficient allocation of pollution,” *The American Economic Review*, 2002.

Fudenberg, D., D. Levine, and E. Maskin, “The Folk Theorem with imperfect public information,” *Econometrica* (62), 1994, 997-1039.

Holmstrom, B., “Moral hazard and observability,” Bell *Journal of Economics*, 1979.

Keener, J.P., *Principles of Applied Mathematics*, Addison-Wesley Publishing Company, 1988.

Kwerel, E, “To tell the truth: Imperfect information and optimal pollution control,” *Review of Economic Studies*, 44(3), 1977, 595-601.

McAfee, R.P., and P. Reny, “Correlated information and mechanism design,” *Econometrica*, 60(2), 1992, 395-421.

McAfee, R.P., and P. Reny, “Competition for agency contracts,” RAND Journal of Economics, 18(2), 1987, 296-307.

Montero, J-P, “Pollution markets with imperfectly observed emissions,” *RAND Journal of Economics*, 36(3), 2005, 645-660.

Pigou, A.C., *The Economics of Welfare*, Macmillan, London, 1952.

Sandmo, A., “Optimal taxation in the presence of externalities,” *Swedish Journal of Economics*, 1975.