

Optimal Audit Policy and Heterogenous Agents[□]

Marisa Ratto[∧]

Thibaud Vergé[∧]

December 2002

Abstract

Frauds can be explained not only in terms of individual willingness to cheat, but may also be driven by opportunities to behave dishonestly. The audit policy should therefore be different for different categories of agents.

This paper focuses on the optimal audit policy when there are two categories of agents and shows that the auditor adopts different policies depending on its budget. When resources are quite limited, the auditor sets identical audit probabilities for both types. On the other hand, in most of the cases, the authority should first ensure that people with lower opportunities choose not to commit an offence.

Keywords: fraud, audit, opportunities, budget allocation.

JEL: D81, G22, H26, K42.

[□]We would like to thank David Alary, David de Meza, Paul Groot, Simon Vicary and Deborah Wilson for helpful discussions. We also benefited from comments from participants at the IFS Public Economics Working Group (Warwick, May 2002), the 3th International Conference on Public Economic Theory (Paris, July 2002) and the CMPO Workshop (Bristol). All remaining errors are ours.

[∧]Leverhulme Centre for Market and Public Organisation, University of Bristol.

[∧]University of Southampton and CMPO.

1 Introduction

A major concern in various fields of activity is the fight against fraud. This is obviously an important issue for society when it is related to major illegal activities like drug dealing, smuggling or other organised criminal activities. However, it is also a major concern for the tax authority to fight against tax evasion and the "underground economy", for insurance companies to identify false claims, for environmental organisations to prevent pollution, or for commercial businesses to protect transactions against internet frauds. Estimating the amount of fraud is impossible, given the nature of the activities involved. However there is a common view that cheating is a pervasive phenomenon in many fields of activities.¹ In the insurance market for example, estimating the exact figure of fraudulent claims² is impossible, but Weisberg and Derrig (1991) have found that, in automobile bodily injuries, the level of suspected fraud was about 10%; and according to Mooney and Salvatore (1990), 13% of automobile insurance claims in Florida are fraudulent.³ The main obstacle for combating fraud seems to be the high cost attached to the activity of auditing and monitoring and the fact that only limited resources can be devoted to deter cheaters. This usually implies that the rate of detection is very low. Evidence for tax evasion suggests that, in the US in 1995, only 1.7% of all individuals tax returns were audited.⁴

Cheating is costly for the offender. There is the cost of punishment, borne by the

¹The National Audit Office estimates that H.M. Customs & Excise are losing more than \$7bn a year to fraud on indirect taxes such as VAT and tobacco duty. In 2001, 2770m of smuggled cigarettes were seized but it is thought this accounts for just 15% of the actual smuggling.

²Fraudulent claims can be of different natures and include false or deliberate claims (claims for a loss that never occurred) or build-up (claiming higher damages than the actual loss, often in order to obtain a payment that covers the contractual excess).

³In the US, the National Insurance Crime Bureau estimates that the annual cost of insurance fraud between \$20 and \$30 billion, while fraud in health insurance is estimated to be larger than \$50 billion according to the National Health Care Association.

⁴See Andreoni et al. (1998).

opportunity cost in case of detection, but also a cost intrinsic in the activity of committing a given fraud, in that the offender usually needs to arrange his affairs in a particular way. In the case of insurance fraud, for example, the cheater has to provide false statements and proofs for his/her fraudulent claim, which could be costly. In general this intrinsic cost depends on the type of fraud. Moreover, for a given type of fraud, the cost will also depend on the opportunities individuals have to cheat: some individuals may find it easier to commit a fraud, thanks to their occupation or other personal characteristics. In case of tax evasion, for example, hiding one's income is certainly more difficult for a university lecturer than for a self-employed. For other types of frauds, like internet frauds or industrial pollution, different technologies involving different costs attached to the activity of defrauding may exist.

We can think of different opportunities to commit an offence in terms of the different costs that will be incurred: for a given fraud, the greater the opportunity to cheat, the lower the cost incurred in doing so. Individuals with greater opportunities to commit an offence will thus more easily engage in fraudulent or criminal behaviour.⁵ Examples cover a whole range of possible situations: insurance fraud, tax evasion, pollution, or any other type of criminal activities.

The monitoring or auditing strategy to deter cheaters should therefore be designed accordingly. Evidence suggests that this is actually the tendency. For tax evasion the audit policy differs across categories of tax payers: different investigation units within the revenue departments are devoted to small and large businesses. For example, among

⁵According to a survey by the of the U.S. IRS (Taxpayer Compliance Measurement Program, 1983), tax compliance is on average 87.2% of net income, but differs a lot across sources of income (93.9% for wages - more difficult to hide as the employer pays taxes or national insurance / pension contributions related to it, but only 47.0% for small business or 18% for farmers). [See also Macho-Stadler and Perez-Castrillo (1997)].

smaller businesses some are more likely to receive an audit: in particular restaurants, take-aways, pubs, building trades, care homes and hotels are more targeted by the tax inspectors. For large business units, auditing efforts are concentrated in more complex businesses such as businesses with related overseas companies and pharmaceuticals. For insurance fraud, insurers appoint experts only when they consider that there are more opportunities to fraud.⁶

Our objective is to analyse a situation in which individuals have different opportunities to commit a fraud, where we model these opportunities in terms of costs incurred to defraud. We focus on costs intrinsic in the activity of cheating. Our view is that there may exist different technologies to commit a given fraud, involving different costs. In a context in which individuals incur different monetary fixed costs, we analyse the problem of setting the optimal audit policy for an anti-fraud authority in order to see how the probability of audit should differ across types. To make this issue relevant, we consider the case of costly audits and assume that the enforcement agency is endowed with limited resources. We show that the optimal audit policy depends on the available budget. If the budget is too low, the anti-fraud authority should not try to distinguish between the two types of agents, but should set equal probabilities of audit for both types. Above a certain threshold, it is optimal to first ensure that individuals with lower opportunities to evade are fully deterred from cheating and devote the rest of the resources to the others. This is because individuals who have to incur higher costs for cheating are more easily deterred from committing a fraud and hence auditing is more effective for this group of cheaters. If the budget is not large enough such a policy implies that those individuals

⁶See Brockett, Xia and Derrig (1998) for an audit algorithm about automobile bodily injury claims.

with greater opportunities to commit a fraud face a lower probability of being detected.

The issue of the consequences of different opportunities to cheat on the optimal audit policy does not seem to be fully addressed in the wide literature on tax evasion or insurance. Two papers can however be closely related to ours. Scotchmer (1987) analyses the optimal audit policy for a tax authority facing two classes of agents, where classes are defined with respect to the distributions of income across individuals. Our approach is quite different in that we address the issue of classes of agents who differ with respect to their opportunities to defraud, and we do not consider individuals with different income and asymmetric information (adverse selection) between individuals and the authorities. In our model individuals with the same income choose different levels of offence, once they face a different intrinsic cost. This distinguishes our approach from the literature on optimal enforcement policy.

More recently and more closely related to our paper, Macho-Stadler and Pérez-Castrillo (1997) analyse the optimal audit policy with heterogeneous taxpayers, where taxpayers differ in the income source and have either easy-to-hide or hard-to-hide income. They obtain the standard result that, in each class, only individuals who reported an income below a given threshold are audited. More interestingly, they also show that hard-to-hide sources of income should be intensively monitored. Once again, the authors consider asymmetric information between the taxpayers and the Internal Revenue Services and base their approach on signalling problems in that the IRS can base its audit policy on the self-reported income. We assume that the enforcement agency can only base its policy on the type of the agent but does not receive any additional signal about the potential fraud. Our findings are nevertheless similar to theirs in that, when the budget is very limited, resources are allocated to deter fraud from the group of individuals who incur a

higher cost of defrauding.

The paper is organised as follows: Section 2 describes the general framework. In section 3; we analyse the agents' behaviour and show that agents with lower opportunities to cheat are deterred more easily from committing an offence. Section 4 characterizes the optimal audit policy under a limited budget. Section 5 extends the analysis to two heterogenous groups of agents, both of them consisting of the two types of individuals, but in different proportions. Section 6 concludes.

2 The Framework

In this section, we present the features of the basic model. In particular, we consider the interactions between a legal authority (or auditors) and agents who have the opportunity to commit an illegal action (offence).

2.1 Agents

The individuals are divided in two categories, N and C: The size of the whole population is, without loss of generality, normalised to 1; and the proportions of type N and type C individuals are λ and $1 - \lambda$, respectively. Our objective in this paper is to isolate the impact of the effect of different opportunities to cheat (that is, different intrinsic costs) on the agents' behaviour. Therefore, we do not want income or attitude towards risk (risk aversion) to play any role and thus assume that both types of individuals have the same gross income Y and the same Von Neumann - Morgenstern utility function u : For the sake of simplicity and to eliminate all income effects, we assume that the coefficient of absolute risk aversion is constant and equal to a , so that the utility function is $u(y) = -\frac{1}{a} e^{-ay}$. Since

we assume constant absolute risk aversion, income does not matter and, without loss of generality, we can normalise it to $Y = 0$.

The agents have an opportunity to commit an offence and derive a benefit B from the offence. B can take any positive value and is a strategic variable chosen by the agent. The authority cannot directly observe whether an offence has been committed and therefore needs to investigate. If an agent is being investigated and caught, the agent has to pay a fine proportional to the monetary value of the offence, $(1 + F)B$, where $F > 0$: The agent thus pays back the benefit of the offence (B) plus a fine $F B$, so that his net income in case of detection is $-F B$:

The two types of individuals differ only in their opportunity to commit a given offence. Type C individuals incur a fixed monetary cost $c > 0$, if they decide to commit an offence. Type N individuals don't face this monetary cost.

Internet fraud is a possible example of such situations. The offender can cheat the internet users in a variety of ways. In general the scam consists of setting a web page and offering certain services which are never delivered. Examples could be bidding for items which will never be delivered, or falsely promising credit cards to people with bad credit histories on payment of up-front fees or making requests for up-front fees to claim winnings that will never be awarded. The technology required to be a fraudulent promoter differs, in that in some cases the technology is more complex or may involve more complex skills than in others. For these scams the cheater has to incur higher costs for defrauding.

Another example could be claiming benefits to which you are not entitled to. Frauds committed by income support (IS) and jobseekers allowance (JSA) claimants account for around 57% of total losses for the Benefits Agency in the UK. A report by the Analyti-

cal Services Directorate at the Department of Work and Pensions estimates that in the period April 2000 - March 2001, \$573m were overpaid for IS fraud and \$201m for JSA fraud.⁷ The offender can either submit a false claim for income support or for jobseekers allowance. In the case of jobseekers allowance the offender has also to sign on at a job centre and frequently visit it. Moreover, if the person involved in the fraud is already working, this fraud requires the collusion of the employer, who does not include the employee in his financial accounts. Hence cheating the Benefits Agency may imply different costs, depending on the "technology" adopted: claiming for jobseekers allowance may involve a higher cost relative to claiming for income support. The enforcement agency can distinguish between the two "technologies" by looking at the type of claim.

2.2 Law Enforcement Agency

In the line of Cremer et al. (1990) who make the distinction between the tax authorities (government who sets the tax rate t and the fine rate F) and the tax administration (tax inspectors who set the probability of detection), we assume that the penalty rate F is set by a different authority (e.g. the government) and focus on the role of the law enforcement agency.

The law enforcement agency (police, tax inspectors, insurance auditors, ...) is endowed with a limited budget to investigate. The agency perfectly knows or can costlessly observe the type of an agent (N or C), but cannot observe whether an individual has cheated or not without investigating. Performing an audit or an investigation is costly and we assume that the budget of the enforcement agency is sufficiently low so that it can only audit a proportion $n < 1$ of the population. If an investigation takes place, it is successful and

⁷The report is available at <http://www.dwp.gov.uk/publications/dwp/2002/fraud-error>.

the enforcement agency discovers whether an individual has cheated or not, and, when relevant, observes the actual value of B :⁸ If an individual is caught, the enforcement agency imposes a fine proportional to the damage: the individual has then to pay back $(F + 1)B$. The fine rate F is exogenous and, for simplicity, we assume that $F > 1$:⁹

The enforcement agency is able to distinguish type N from type C agents and can therefore set two different probabilities of audits, one for each class of individuals. The objective of the enforcement agency is therefore to set the probabilities p_N and p_C in order to minimise the expected monetary value of the offences, given the available budget (or the proportion of the population that can be audited) n . An important debate in the tax evasion literature is the choice of the enforcement agency's objective function. The question is usually whether the authority should incorporate the fines collected in her objective function or not. In most of the models on optimal enforcement for tax evasion, it is assumed that the objective of the enforcement agency is to maximise net revenue. In line with Greenberg (1984), we assume that the objective function of the anti-fraud authority is to minimise the level of committed frauds. However, we could integrate the expected amount of fines into the objective function without qualitatively affecting our results.

2.3 Timing of the Game

In line with most of the economic literature, we assume that the enforcement agency can perfectly commit to an audit policy before the agents choose their strategy. We

⁸This assumption is not important and we could have considered an imperfect audit. In our situation, a probability of audit (and detection) of p is equivalent to the case in which individuals are audited with probability p_A and the audit is successful with probability p_S ; where $p = p_A p_S$.

⁹This is a purely technical assumption. Our main result remains strictly identical for $0 < F < 1$: However the analysis is more complicated and we made the choice to present here the simplest possible version of the model. Details are available from the authors upon request.

are therefore looking for an ex ante optimal policy. As pointed out by Picard (1996), the outcome of interactions between the agents (potential cheaters) and the authority (auditors) critically depends on whether the authority can credibly pre-commit or not to her auditing strategy:¹⁰ "in a situation where fraud is pervasive, it is essential for the companies to credibly announce that a tough monitoring policy will be enforced. (...) However, since monitoring is costly to the insurer, what is desirable ex ante may be suboptimal ex post and a commitment to subject claims to close scrutiny may not be credible." Even though some papers have analysed situations in which the tax authority or the insurers cannot pre-commit and thus choose their optimal audit policy given the agents' signals,¹¹ this is not the most common case. Most of the existing literature assumes that the authority can announce and credibly pre-commit to an audit strategy before the agents make their report (or claim). This could be seen as a short cut for reputation: the enforcement agency will always implement its policy even though it might not be optimal ex post in order to deter offending individuals in the forthcoming periods. We do not want to try to model these dynamic interactions here and take the pre-commitment hypothesis for granted.

Moreover, in contrast with the standard literature where the issue of pre-commitment determines whether you consider a screening or a signalling game, we consider the case of a hidden action (moral hazard) which is neither verifiable (i.e. contractible) nor observable, directly or indirectly. Assuming that the agency cannot pre-commit on a given audit policy would therefore lead to a game where the agents and the enforcement agency choose their

¹⁰Similarly, Reinganum and Wilde (1985, 1986) analyse the commitment and no-commitment equilibrium audit policies in the context of tax evasion.

¹¹See for example Picard (1996) for the insurance case, and Graetz et al. (1986) or Hindrinks (1999) for tax evasion.

strategies simultaneously (amount of fraud and audit policy respectively). In this case, there would never exist an equilibrium in which at least one type of agent does not cheat.

The game we analyse is thus the following:

1. The enforcement agency allocates the investigation budget n between the two types of individuals, that is, pre-commits on the probabilities of investigation p_N and p_C :
2. Each individual ($i = N; C$) decides whether to commit an offence ($B_i > 0$) or not ($B_i = 0$). If he decides to offend, he chooses the level of benefit (and damage) B_i :
3. The investigation takes place according to the policy set in the first stage.

3 Agents' Behavior

We now analyse the agents' behaviour when they face a probability of audit p (in which case they are caught and fined at rate $F > 1$):

3.1 Type N Agents (Zero Cost)

When the agent N decides the level of fraud, B_N ; he maximises his expected utility

$$U_N(B) = (1 - p)u(B) + pu(j - FB) = \int_0^B (1 - p) e^{-aB} + p e^{-aFB}$$

Lemma 1 A type N agent commits an offence as long as the probability of being audited is less than $\frac{1}{1+F}$, and, when caught with probability p , this level of fraud is $B_N^a(p) = \mathcal{B}(p)$,

where:

$$\mathfrak{B}(p) = \frac{1}{a(1+F)} \ln \frac{1-p}{pF}.$$

Proof. The objective function is strictly concave, since u is strictly concave. The first-order condition is then:

$$\frac{dU_N}{dB} = (1-p)ae^{aB} - paFe^{aFB} = 0, \text{ with equality if } B > 0.$$

If the amount of fraud is strictly positive, this simplifies to

$$B = \mathfrak{B}(p) = \frac{1}{a(F+1)} \ln \frac{1-p}{pF},$$

and this remains valid as long as $\mathfrak{B}(p) > 0$; that is as long as:

$$\frac{dU_N}{dB} \Big|_{B=0} = a(1-p(1+F)) > 0.$$

This easily leads to the result. ■

Notice that this function $\mathfrak{B}(p)$ is strictly decreasing in p and strictly convex for any $p \in \left(0, \frac{1}{1+F}\right)$.

3.2 Type C Agents (Positive Cost)

When a type C agent decides whether to commit an illegal action, and chooses the amount of fraud (private benefit, B), he compares the utility he can get if he does not offend, that is $u(0) = 1$, with the expected utility he gets if the benefit derived from the offence is

$B > 0$, that is:

$$U_C(B) = (1 - p)u(B - c) + pu(j - FB - c) = e^{i ac}U_N(B).$$

Under our CARA assumption, if a type C agent decides to offend, the level of the offence is identical to that of a type N agent facing the same probability of audit, that is: if $B_C^*(p) > 0$; then $B_C^*(p) = B_N^*(p) = B^*(p)$. The only difference is that it is now more costly to engage in a fraud and the probability of detection above which this agent stops cheating is lower than for a type N agent. A type C agent offends and chooses $B_N^*(p)$ if and only if:¹²

$$e^{i ac}U_N(B_N^*(p)) > j - 1, \quad j - 1 > U_N(B_N^*(p)) < e^{ac}.$$

As the level of offence is a strictly decreasing function in p , $U_N(B_N^*(p))$ is a strictly decreasing function of p : For any cost c ; it thus exists a unique probability $\bar{p}_C(c)$ such that $U_N(B_N^*(\bar{p}_C(c))) = j - 1$ and this probability is a strictly decreasing function of the cost of evasion c , such that $\bar{p}_C(0) = \frac{1}{1+F}$ and $\lim_{c \rightarrow 1} \bar{p}_C(c) = 0$: The type C agent offends only when the probability of audit is lower than this threshold. Type C agents can therefore be characterized by this threshold \bar{p}_C rather than by the monetary cost of illegal actions.

Lemma 2 For any cost $c > 0$; there exists a critical value $\bar{p}_C(c) < \frac{1}{1+F}$; above which a type C individual decides not to offend. For any probability of detection p lower than this threshold, this agent chooses $B_N^*(p) = B^*(p)$:

The results of lemma 1 and 2 are summarized in Figure 1.

¹²Here we assume that, when the agent is indifferent between $B = 0$ and $B = B_N^*(p) > 0$; he chooses not to offend. This assumption does not modify the analysis.

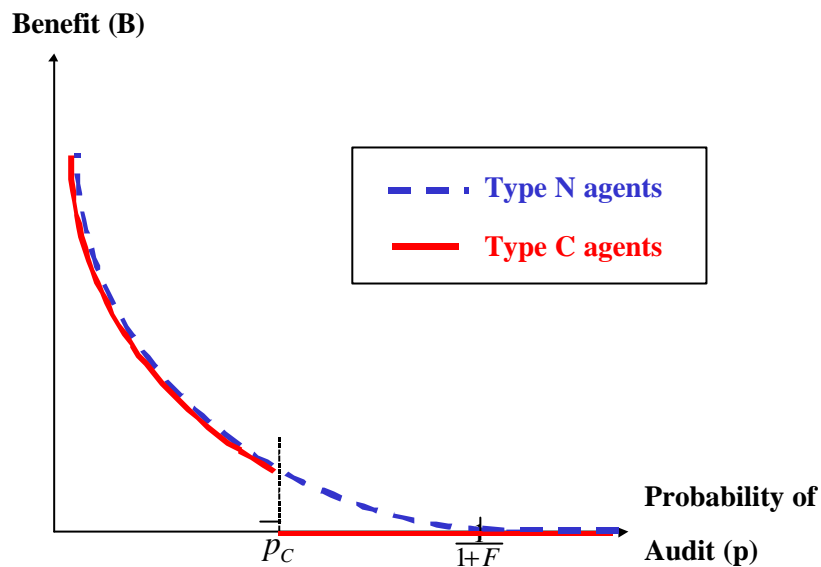


Figure 1: Level of the Offence Committed by Type N and Type C Agents

4 Optimal Audit Policy

Let us now consider the optimal investigation policy. The objective of the enforcement agency is to minimise the expected level of offence subject to its budget constraint. This minimisation program writes as:

$$\begin{aligned} \min_{(p_N; p_C)} EB(p_N; p_C) &= \alpha B_N^a(p_N) + (1 - \alpha) B_C^a(p_C) \\ \text{s.t.} \quad &: p_N = \frac{n_N}{1 + F}; p_C = \frac{n_C}{1 + F} \text{ and } n_N + n_C = n \end{aligned}$$

If the available budget is sufficiently large $\alpha n \geq \frac{n}{F+1} + (1 - \alpha) \bar{p}_C$, the enforcement agency can set probabilities of audit such that both types of agent choose not to offend and this is obviously the optimal situation. However, this case is not interesting as it implies that resources are not scarce.

Let us now suppose that resources are limited. It is now no longer possible to deter all individuals and the budget constraint will therefore be binding at the optimum. The agency's minimisation program can therefore be rewritten:

$$\min_{p \in [0; \min(1; \frac{p}{1-i})]} EB(p) = i B_N^a(p) + (1-i) B_C^a \left(\frac{n-i}{1-i} p \right)$$

Solving this programme leads to the following proposition, the results of which are summarised in Figure 2:

Proposition 3 For any probability $0 < \bar{p}_C < \frac{1}{F+1}$ and any fine rate $F > 1$; there exists a threshold n^a ; with $(1-i)\bar{p}_C < n^a \cdot \bar{p}_C$; such that:

- 2 If the available budget is too low ($n < n^a$); the enforcement agency sets equal probabilities for both types of agents, $p_C^a = p_N^a = n$: The share of the budget allocated to a specific type of agents is thus equal to the relative size of this group (i.e: $n_N^a = i \cdot n$ and $n_C^a = (1-i) \cdot n$);
- 2 If the available budget is sufficiently large ($n > n^a$); the enforcement agency first ensures that individuals with lower opportunities (type C) do not offend ($p_C^a = \bar{p}_C$) and allocates the remaining share of the budget for type N agents.

Proof. See Appendix A. ■

The intuition of these results is as follows. If the available budget is too small ($n < (1-i)\bar{p}_C$, below line Φ_1 in Figure 2), it is not possible to set a probability of audit equal to \bar{p}_C for type C agents, who will therefore commit an offence whatever the allocation of the budget (i.e. probabilities of audit) decided by the enforcement agency. In this case both types will always engage in fraud. In deciding how to allocate the budget between

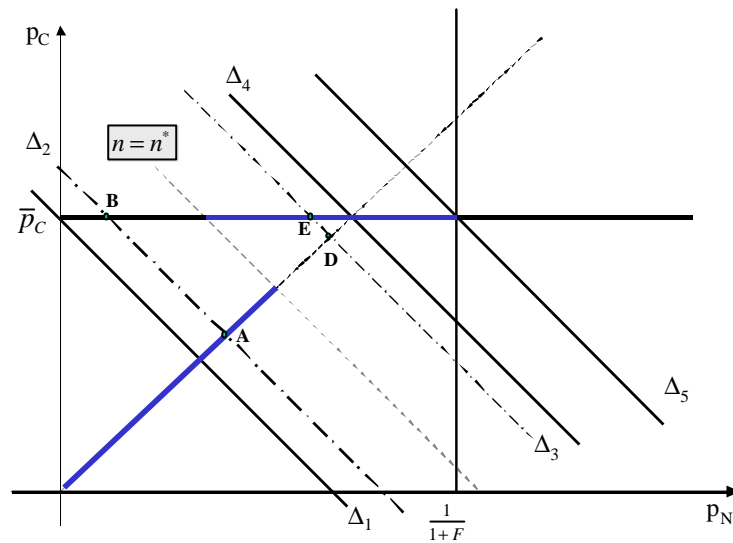


Figure 2: Optimal Audit Policy

the two types, the enforcement agency compares the marginal benefit of increasing the probability of detection for one type, e.g. type N individuals, (that is, the decrease in the expected value of the offence committed by these agents) and the marginal cost of such a change (that is, the increase in the expected offence by type C agents). For a given budget allocation $(n_N; n_C)$, probabilities are inversely related to the size of the relevant group $p_N = \frac{n_N}{n}$ and $p_C = \frac{n_C}{n}$ and the enforcement agency's objective function (expected offence) is proportional to the size of the group ($EB = p_N B_N + (1 - p_N) B_C$): For a marginal change of the probability of detection p_N ; marginal benefit equals marginal cost when the probabilities are identical, that is, when $p_N = p_C$: This is because both types will choose the same amount of fraud once they face the same probability of detection.

When the available budget is large enough ($\bar{p}_C < n \cdot \bar{n}$, between lines Φ_4 and Φ_5), setting equal probabilities is no longer optimal as this would imply $p_C > \bar{p}_C$: By doing this, the enforcement agency would devote more resources than necessary to ensure that type C individuals do not commit any offence. It is then optimal to reduce the share of

the budget devoted to type C agents to $p_C = \bar{p}_C$ and reallocate it to increase the share devoted to type N agents.

Finally, for intermediate values of the budget ($(1 - \lambda)\bar{p}_C < n \cdot \bar{p}_C$, between lines Φ_1 and Φ_4), the authority has to compare these two policies. Switching from equal probabilities to a policy in which the enforcement agency first tackles type C individuals and fully deters them from committing an offence implies to increase the audit probability for type C individuals. This has two opposite effects: on the one hand, increasing the probability of audit for type C from $p_C = n$ to $p_C = \bar{p}_C$ has a negative impact, in that it increases the expected offence by type N individuals. This negative effect is larger when the budget is more limited (lower n) and tends to 0 when n is closer to \bar{p}_C : On the other hand, by setting $p_C = \bar{p}_C$ in order to prevent type C from offending (that is moving from $p_C = \bar{p}_C$ to $p_C = \bar{p}_C$) the enforcement authority benefits from the discontinuity in the offending pattern of type C agents (they "jump" down to zero fraud).

If the budget is low (n close to $(1 - \lambda)\bar{p}_C$, line Φ_2); the negative effect is very large and overweighs the positive impact of the discontinuity. Switching from equal probabilities (point A) to focusing on type C individuals (point B) requires to lower the probability of detection for type N individuals by a substantial amount. The increase in the expected offence by type N individuals will not be compensated by the jump to zero fraud by the C types and the enforcement agency therefore chooses to set equal probabilities.

However, if the budget is large enough (n close to \bar{p}_C , line Φ_3), the decrease in the audit probability for type N individuals required to focus on type C individuals is much smaller: the negative effect of switching from D to E is more than compensated by the jump to zero fraud by the C types. The enforcement agency thus prefers to fully deter

type C agents. In between these two cases, there exists a particular value of the budget (n^*) for which the enforcement agency is indifferent between the two policies, as switching from one to another has no impact on the total expected fraud. Above n^* it will always be optimal to focus on type C individuals and below n^* it will be optimal to set equal probabilities

Finally, it is worth noting that ensuring that type C agents do not commit an offence does not necessarily mean that the enforcement agency sets a higher probability of audit for this group (or allocate a relative share of its budget higher than the relative size of the group). This will depend on the size of the available budget, n : If $n^* \cdot n < \bar{p}_C$, then we indeed have $p_C^* > p_N^*$ (that is, $n_C^* > (1 - \beta) n$), but for $\bar{p}_C < n \cdot \bar{n}$; then $p_C^* < p_N^*$: Hence, when it is optimal to first target the group of offenders with lower opportunities, the smaller the available budget for auditing, the more likely the case that individuals with greater opportunities to evade face a lower probability of being audited. In this case the optimal audit policy would further increase the disparity in opportunities to evade.

5 Heterogeneous Groups of Agents

In the previous section, we have assumed that the law enforcement agency was costlessly and perfectly able to distinguish the two types of agents. We now relax this assumption and suppose that the enforcement agency has only an imperfect knowledge of the agent's type. More precisely, we assume that the economy consists of two groups of agents of equal size (and we normalize the size of each group to be $\frac{1}{2}$; so that the size of the whole population of agents remains equal to 1), and that each group is composed of agents of types C and N: The two groups are assumed to differ with respect to the proportion of

agents who have greater opportunities to commit a fraud, namely type N agents: we thus assume that the proportion of type N agents in groups 1 and 2 are respectively θ_1 and θ_2 ; with $\theta_1 > \theta_2$. Group 1 can thus be considered as riskier by the enforcement agency, in that people are more likely to commit a fraud. The problem for the enforcement agency is to allocate the resources between the two groups, i.e. to set the audit probability for each group. We assume that the agency can observe which group an individual belongs to, but cannot distinguish between types.

An example for this setting could be tobacco, alcohol or drugs smuggling from continental Europe to the UK. In this case the Customs have to decide how many officers to allocate to each harbour, or within the Dover ferry terminal how many cars/passengers to stop for searching, knowing that some of them are more likely to smuggle because they come from countries where getting the merchandise is less costly: cars coming to the UK from France are more likely to be loaded with alcohol, for example, or passengers coming from Holland are more likely to carry cannabis. The search probability should therefore depend on they are arriving from.

The enforcement agency now needs to decide whether to focus on the "riskier" group (namely group 1) and how to allocate its resources between the two groups. Within the same group individuals will face the same probability of detection, regardless of their type, and their behaviour is thus given by lemma 1 and 2. Once again, we make the assumption that resources are scarce ($n < \frac{1}{F+1}$), so that the enforcement agency cannot fully deter

fraud, i.e. its budget constraint is binding. The enforcement agency's program is then:

$$\min_{(p_1, p_2)} \theta_1 B_N^a(p_1) + (1 - \theta_1) B_C^a(p_1) + \theta_2 B_N^a(p_2) + (1 - \theta_2) B_C^a(p_2) \quad \text{s.t.} \quad p_1 + p_2 = 2n:$$

This leads to the following proposition, the results of which are summarised in Figure 3.

Proposition 4 For any probability $0 < \bar{p}_C < \frac{1}{F+1}$, any fine rate $F > 1$ and any proportions $\theta_1 > \theta_2$, there exist three threshold values n^a , n^{aa} and n^{aaa} ; with:

$$\frac{1}{2}\bar{p}_C < n^a < \bar{p}_C < n^{aa} \cdot \frac{1}{2} \left(\bar{p}_C + \frac{1}{1+F} \right) \cdot n^{aaa} < \frac{1}{1+F};$$

such that:

1. If $n < n^a$; the enforcement agency sets equal probabilities for both groups of agents, $p_C^a = p_N^a = n$ and all agents fraud.
2. If $n^a < n < n^{aa}$; the enforcement agency first ensures that the type C agents of the "less risky" group do not fraud, that is sets $p_2^a = \bar{p}_C$ and thus $p_1 = 2n - \bar{p}_C$:
3. If $n^{aa} < n < n^{aaa}$; the optimal policy is such that all type C agents do not fraud but all type N do, and the enforcement agency allocates more resources to audit the "riskier" group, that is $\bar{p}_C < p_2^a < n < p_1^a < \frac{1}{1+F}$:
4. Finally, if $n^{aaa} < n < \frac{1}{1+F}$; the enforcement agency mainly focuses on the "riskier" group to make sure that these agents do not fraud setting $p_1^a = \frac{1}{1+F}$: In that case, only the type N agents of the "less risky" group fraud.

Proof. See Appendix B. ■

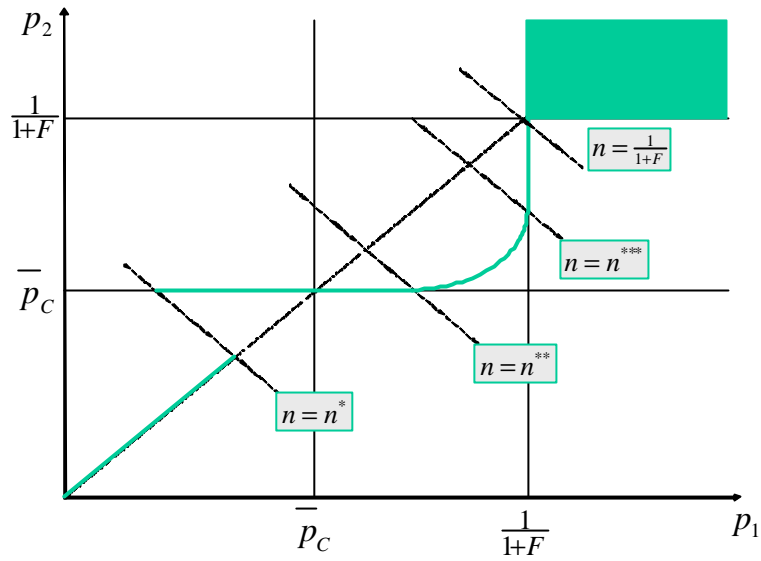


Figure 3: Optimal Audit Policy with Heterogeneous Groups of Agents

For low values of the budget ($n < \bar{p}_C$); the situation is very similar to the case in which the enforcement agency is able to observe the type of the agents. When resources are very limited $n < \frac{1}{2}\bar{p}_C$; all agents (both types in both groups) will engage in fraud because with those resources the enforcement agency is unable to set a probability of detection high enough to deter even type C individuals from cheating. In this situation, like in the case for which the enforcement agency was able to distinguish the agents (see proposition 3), the expected amount of fraud is minimised when the probabilities of detection are equal.

When resources are higher but still too low to deter all type C agents $\frac{1}{2}\bar{p}_C < n < \bar{p}_C$; the enforcement agency has the choice between deterring some of the type C agents or setting equal probabilities with all agents defrauding. If the agency has decided to deter fraud from the type C agents of a given group (let say 1), it is optimal to do it with $p_1 = \bar{p}_C$. Suppose on the contrary that $p_1 > \bar{p}_C$: since all agents in group 2 defraud

whereas only type N agents do so in group 1; decreasing slightly the probability of audit p_1 would reduce the total amount of fraud. Since it is optimal to tackle the type C agents of the chosen targeted group with a probability of audit equal to \bar{p}_C ; the amount of fraud for non-targeted group does not depend on the identity of the targeted group

$$\frac{1}{2} \mathbb{E}(2n_i - \bar{p}_C) : \text{As the proportion of type C agents is larger in the less risky group (namely group 2), it is better to deter fraud from type C agents in this group rather than in the riskier group.}$$

For these values of the agency's budget, the choice is then between equal probabilities of audit and deterring type C defrauders in the less risky group. The argument is now very similar to the "perfect information" case: switching from equal probabilities to $p_2 = \bar{p}_C$ has two opposite effects: on the one hand, increasing the probability of audit for type 2 from $p_2 = n$ to $p_2 = \bar{p}_C$ has a negative impact, while moving from $p_2 = \bar{p}_C$ to $p_2 = \bar{p}_C$ allows the enforcement authority to benefit from the discontinuity in the offending pattern of the type C agents in group 2: Since the negative effect is larger for lower values of n (infinite for $n = \frac{1}{2}\bar{p}_C$ and equal to 0 for $n = \bar{p}_C$), there exists a critical budget n^* for which the enforcement agency switches between these two regimes.

When the available budget increases and becomes larger than \bar{p}_C ; the enforcement agency can set the two probabilities of audit above \bar{p}_C in order to deter all type C agents. The choice will now depend on the relative proportions of type N individuals in the two groups. Starting from $p_1 = p_2$; the marginal benefit from an increase in p_1 is indeed proportional to θ_1 vs θ_2 . Because this proportion is higher in group 1; the enforcement agency now prefers to set a higher probability of audit for the riskier group. However, the choice is still constrained by the need to deter all type C agents, i.e. keep the audit

probability for the less risky group above \bar{p}_C : Therefore for lower values of the budget ($n < n^{***}$) this constraint is binding and the enforcement agency sets $p_2^a = \bar{p}_C$: For higher values of n ; the enforcement agency can increase the audit probabilities for both groups above this level and maintain a higher probability for the riskier group. Finally, when the budget is large enough ($n > n^{***}$), the enforcement agency can target the type N agents in the riskier group.

6 Concluding Remarks

In this paper, we have considered the effect of different opportunities to commit a fraud on the optimal audit policy with scarce resources. We first analysed the case where the enforcement agency observes the agents' types and have shown that she should usually set different probabilities of audit for agents with different opportunities to cheat. If the budget of the enforcement agency is too small, it would be too costly (or even impossible) to deter agents with lower opportunities to cheat, and the agency sets equal probabilities, thereby allocating to each group a share of the budget equal to its relative size. However, as the budget increases, it becomes optimal to first ensure that agents with lower opportunities are deterred from cheating. This might lead to situations where the agents who incur a higher cost for cheating are also the agents facing a higher probability of detection.

This result relies on some specific assumptions, but is more robust than it may first appear. Relaxing some assumptions would modify the fact that the enforcement agency sets equal probabilities (i.e. considers that there is only one class of individuals) when its budget is very limited. However, the existence of a zone for which the authority focuses

First on the agents with the lowest opportunity to evade would remain. This would, for example, be the case if we relax the CARA assumption or consider a cost of fraud of the form $C(B) = c + g(B)$. We also assumed that the enforcement agency does not take the fines into account in its objective function. Relaxing this hypothesis would not alter our results but only modify the value of the threshold n^* :

We then considered the case where the enforcement agency does not observe if an agent has a high or low opportunity to commit a fraud, but knows whether the agent belongs to a more or less risky group. Although the enforcement agency cannot distinguish whether individuals have to incur a cost for cheating, the audit rule is very similar to the case where there is perfect information about the agents type. If the budget is too small it would be too costly to set different probabilities of audit for the two groups. When the budget increases, the criterion is still to tackle first the individuals with lower opportunities to evade (although the enforcement agency cannot distinguish them) : if the resources are not enough to deter type C in both groups, they should be tackled in the group where they are more numerous, that is the less risky group.

Hence, we can generalise our results and conclude that when there are different technologies to commit a given fraud, implying different costs and hence different opportunities to engage in illegal behaviour, the audit strategy should take this into account. When the budget of the enforcement agency is limited, the enforcement agency should first tackle individuals who are more easily deterred from cheating, i.e. those incurring higher costs to defraud, or using a more complicated technology. In terms of the cases we considered above, this implies for example that frauds for claiming job seekers allowance should be more targeted than frauds from income support claims. Or, again, in the case of internet frauds, if resources to devote to auditing are tight (which we can assume is the case),

the anti-fraud authority should first tackle the fraudsters using the more complicated and hence costly technology. How this relates to actual policy is hard to tell, as it is quite difficult to access information on the audit strategy, which is considered highly confidential. For tax evasion, the audit policy seems to rely on the results of past audits. Evidence for the U.S. suggests that the IRS informs its audit strategy on the basis of the results of a program of intensive audits conducted on a stratified random sample of returns (Tax Compliance Measurement Program (TCMP)).¹³ In particular, the IRS calculates a so called "discriminant function" (DIF) from the results of the TCMP, which it uses to assign each return a score (DIF score) for the likelihood that it contains some irregularities or evasion. Apparently over one-half of all audit selections are based at least partly on the DIF score. Hence it seems to be the case that the audit policy is adjusted on the basis of previously detected cases of tax evasion. In contrast with our results (which imply focusing on frauds which are easier to deter), the criterion underlying the enforcement strategy seems to consist of first tackling frauds that are easier to detect. Distinguishing over the type of technology used or the range of different opportunities to commit a certain fraud could therefore be an additional criterion to adjust the audit strategy.

¹³See Andreoni (1998).

References

- [1] Andreoni, James, Brian Erard and Jonathan Feinstein, (1998), "Tax Compliance", *Journal of Economic Literature*, 36, 818-860.
- [2] Brockett, Patrick, Xiaohua Xia and Richard Derrig (1998), "Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud", *Journal of Risk and Insurance*, 65(2), 245-274.
- [3] Cremer, Helmuth, Michel Marchand and Pierre Pestiau (1990), "Evading, Auditing and Taxing: The Equity-Compliance Trade-off", *Journal of Public Economics*, 43; 67-92.
- [4] Graetz, Michael, Jennifer Reinganum and Louis Wilde (1986), "The Tax Compliance Game: Toward an Interactive Theory of Law Enforcement", *Journal of Law, Economics and Organization*, 2, 1-32.
- [5] Greenberg, Joseph (1984), "Avoiding Tax Avoidance: A (Repeated) Game-Theoretic Approach", *Journal of Economic Theory*, 32, 1-13.
- [6] Hindriks, Jean (1999), "Income Tax Compliance: the No-Commitment Game", *University of Exeter Discussion Papers 99* ; 19:
- [7] Macho-Stadler, Inés and David Pérez-Castrillo (1997), "Optimal Audit Policy with Heterogenous Incomes Sources", *International Economic Review*, 38(4), 951-968.
- [8] Mooney, S.F. and J.M. Salvatore (1990), *Insurance Fraud Project: Report on Research*, Insurance Information Institute.

- [9] Picard, Pierre (1996), "Auditing Claims in Insurance Market with Fraud: The Credibility Issues", *Journal of Public Economics*, 63; 27-56.
- [10] Reinganum, Jennifer and Louis Wilde (1985), "Income Tax Compliance in a Principal-Agent Framework", *Journal of Public Economics*, 26, 1-8.
- [11] Reinganum, Jennifer and Louis Wilde (1986), "Equilibrium Verification and Reporting Policies in a Model of Tax Compliance", *International Economic Review*, 27, 739-760.
- [12] Scotchmer, Suzanne (1987), "Audit Classes and Tax Enforcement", *The American Economic Review*, 77(2), 229-233.
- [13] Weisberg, Herbert and Richard Derrig (1991), "Fraud and Automobile Insurance: A Report on the Baseline Study of Bodily Injury Claims in Massachusetts", *Journal of Insurance Regulation*, 9, 427-451.

A Proof of Proposition 3

Using the results of lemma 1 and 2, we can divide the space $(p_N; p_C)$ in three zones, according to the following figure:

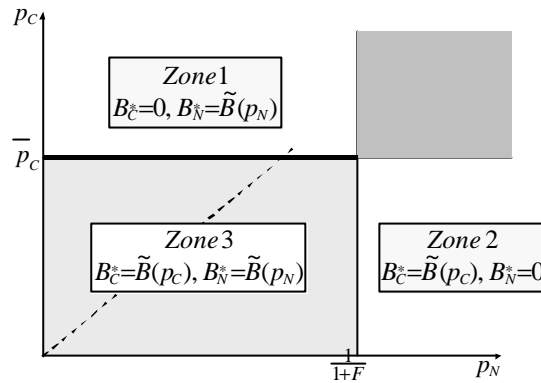


Figure 4: Agents' Behaviour

A.1 Zone by Zone Analysis

- 2 In zone 1 $p_C > \bar{p}_C$ and $p_N < \frac{1}{1+F}$: Type C individual do not commit any offence ($B_C^a = 0$) whereas type N agents frauds ($B_N^a > 0$): An increase in p_N does not affect type C's behaviour and decreases the expected offence for the type N agents. The enforcement agency therefore wants to move towards $p_C = \bar{p}_C$; that is, $p_N = \frac{n_i (1 - \alpha) \bar{p}_C}{\alpha}$:
- 2 In zone 2 $p_C < \bar{p}_C$ and $p_N > \frac{1}{1+F}$: This situation is similar to zone 1 inverting the roles played by the two types of agents. The enforcement agency therefore wants to move towards $p_N = \frac{1}{1+F}$:
- 2 In zone 3 $p_C < \bar{p}_C$ and $p_N < \frac{1}{1+F}$: In that case, both types of agents commit

an offence and the first order condition of the enforcement agency's minimisation program writes as:

$$\frac{\partial \mathcal{B}^0(p)}{\partial p} = \mathcal{B}^0 \frac{n_i - p}{1 - p} = 0, \quad p = n, \quad p_N = p_C = n$$

Notice that this first order condition is necessary and sufficient (since $\mathcal{B}^0 > 0$): In this zone 3; the enforcement agency sets equal probabilities for both types whenever it is possible, that is, whenever $n < \bar{p}_C$: For $n > \bar{p}_C$; the optimum (in zone 3) is to set a probability of audit for type C as close as possible to \bar{p}_C .

We can now summarise the results obtained in the different zones. We should first notice that the expected amount of offence is continuous when we move from zone 2 to zone 3 at $p_N = \frac{1}{1+F}$; however we have a discontinuity between zones 1 and 3 (at $p_C = \bar{p}_C$) due to the discontinuity in the level of benefit chosen by type C agents, the optimal choice for the enforcement agency being then to choose $p_C = \bar{p}_C$ (rather than \bar{p}_C):

A.2 Optimal Audit Policy

Summarising all these results we now have: the enforcement agency's optimal choice is such that:

$$^2 \text{ if } n < (1 - \beta) \bar{p}_C : p^* = n, \quad p_C^* = p_N^* = n;$$

$$^2 \text{ if } \bar{p}_C < n < \bar{n} : p^* = \frac{n - (1 - \beta) \bar{p}_C}{\beta}, \quad p_C^* = \bar{p}_C;$$

$$^2 \text{ if } (1 - \beta) \bar{p}_C < n < \bar{p}_C : p^* = n \text{ or } p^* = \frac{n - (1 - \beta) \bar{p}_C}{\beta}; \text{ Expected offence is then,}$$

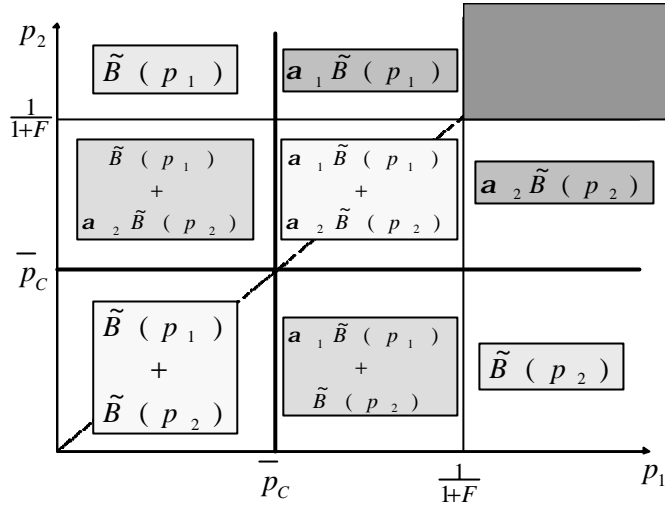


Figure 5: Expected fraud in the case of heterogeneous groups of agents

B.1 Zone by Zone Analysis

Keeping in mind that $p_1 + p_2 = 2n$, it is easy to see that, provided that $p_1 \leq \frac{1}{1+F}$; the enforcement agency wants to maximise p_2 and will thus never set $p_1 > \frac{1}{1+F}$: For the same reason, the enforcement agency will never set $p_2 > \frac{1}{1+F}$:

² Let us now consider the case $\bar{p}_C < p_1 \cdot \frac{1}{1+F}$ and $p_2 < \bar{p}_C$:

In that case, the expected fraud (as a function of p_1) is

$$EB(p_1) = \frac{1}{2} \left[\mathbb{E}^1 \mathbb{B}(p_1) + \mathbb{B}(2n - p_1) \right]$$

and

$$EB^0(p_1) = \frac{1}{2} \left[\mathbb{E}^0 \mathbb{B}^0(p_1) + \mathbb{B}^0(2n - p_1) \right] > 0:$$

The enforcement agency will therefore try to minimise p_1 in that zone and thus set either $p_1 = \bar{p}_C$ (and thus $p_2 = 2n - \bar{p}_C$) if $n \cdot \bar{p}_C$, or $p_1 = 2n - \bar{p}_C$ (and thus

$p_2 = \bar{p}_C$) if $n > \bar{p}_C$. The analysis is very similar for the "symmetric" case: $p_1 = \bar{p}_C$ and $\bar{p}_C < p_2 < \frac{1}{1+F}$:

Finally, we can easily check that it is always preferable to choose $p_2 = \bar{p}_C$ and $p_1 = 2n - \bar{p}_C$ rather than $p_1 = \bar{p}_C$ and $p_2 = 2n - \bar{p}_C$: If, for example, $n > \bar{p}_C$; we have to compare:

$$^2 \text{ for } p_1 = \bar{p}_C : EB_1 = \alpha_2 \mathcal{B}(2n - \bar{p}_C) + \alpha_1 \mathcal{B}(\bar{p}_C);$$

$$^2 \text{ for } p_2 = \bar{p}_C : EB_2 = \alpha_1 \mathcal{B}(2n - \bar{p}_C) + \alpha_2 \mathcal{B}(\bar{p}_C);$$

and we have: $EB_1 - EB_2 = (\alpha_1 - \alpha_2) \mathcal{B}(\bar{p}_C) - \mathcal{B}(2n - \bar{p}_C) > 0$:

If now $n < \bar{p}_C$; we have to compare:

$$^2 \text{ for } p_1 = \bar{p}_C : EB_1 = \mathcal{B}(2n - \bar{p}_C) + \alpha_1 \mathcal{B}(\bar{p}_C);$$

$$^2 \text{ for } p_2 = \bar{p}_C : EB_2 = \mathcal{B}(2n - \bar{p}_C) + \alpha_2 \mathcal{B}(\bar{p}_C);$$

and we have: $EB_1 - EB_2 = (\alpha_1 - \alpha_2) \mathcal{B}(\bar{p}_C) > 0$:

² Let us now consider the case $p_1, p_2 < \bar{p}_C$:

In this case, the situation is identical to the zone 3 of Figure 4 (see Appendix A - Proof of Proposition 3) and the enforcement agency's (constrained) optimal choice is to set equal probabilities for the two groups of agents, that is, $p_1 = p_2 = n$:

² The final case we need to consider is $\bar{p}_C < p_1, p_2 < \frac{1}{1+F}$:

In that case, the expected fraud (as a function of p_1) is

$$EB(p_1) = \frac{1}{2} \left(\mathbb{E}_1^3(p_1) + \mathbb{E}_2^3(2n - p_1) \right)$$

and

$$EB^0(p_1) = \frac{1}{2} \left(\mathbb{E}_1^0(p_1) + \mathbb{E}_2^0(2n - p_1) \right)$$

We can furthermore notice that $\mathbb{E}^0(p_1)$ being strictly convex, $EB(p_1)$ is also strictly convex and $EB^0(n) < 0$: Let us now define the two following functions:

$$\Phi_1(p) = \mathbb{E}_1^0\left(\frac{1}{1+F}\right) - \mathbb{E}_2^0(p) \quad \text{and} \quad \Phi_2(p) = \mathbb{E}_1^0(p) - \mathbb{E}_2^0(\bar{p}_C^+)$$

Notice that $\Phi_1\left(\frac{1}{1+F}\right) < 0$; $\Phi_2(\bar{p}_C^+) < 0$; Φ_1 is a strictly decreasing function of p , while Φ_2 is strictly increasing in p , for $\bar{p}_C^+ < p < \frac{1}{1+F}$; and

$$\Phi_1(\bar{p}_C^+) = \Phi_2\left(\frac{1}{1+F}\right) = \mathbb{E}_1^0\left(\frac{1}{1+F}\right) - \mathbb{E}_2^0(\bar{p}_C^+)$$

We therefore have to consider two possible situations:

{ if $\mathbb{E}_1^0\left(\frac{1}{1+F}\right) - \mathbb{E}_2^0(\bar{p}_C^+) > 0$; then there exist two critical values n^{min} and n^{max} , such that:

$$\bar{p}_C < n^{\text{min}} < \frac{1}{2} \bar{p}_C + \frac{1}{1+F} < n^{\text{max}} < \frac{1}{1+F};$$

and such that the minimum expected fraud in the zone $\bar{p}_C \cdot p_1; p_2 \cdot \frac{1}{1+F}$ is

attained at:

$$\text{for } \bar{p}_C < n < n^{\text{max}} : p_2 = \bar{p}_C \text{ and } p_1 = 2n - \bar{p}_C; \quad (1)$$

$$\text{for } n^{\text{max}} < n < \frac{1}{1+F} : p_1 = \frac{1}{1+F} \text{ and } p_2 = 2n - \frac{1}{1+F}; \quad (2)$$

$$\text{for } n^{\text{min}} < n < n^{\text{max}} : p_1 = p(n) \text{ and } p_2 = 2n - p(n); \quad (3)$$

where $p(n)$ is the unique solution of the equation $\mathbb{B}_1(p) - \mathbb{B}_2(2n - p)$ such

that $\bar{p}_C < p(n) < 2n - \bar{p}_C < \frac{1}{1+F}$:

{ if $\mathbb{B}_1(\frac{1}{1+F}) - \mathbb{B}_2(2n - \frac{1}{1+F}) < 0$: in this particular case we have

$$n^{\text{min}} = n^{\text{max}} = \frac{1}{2} \bar{p}_C + \frac{1}{1+F};$$

B.2 Optimal Audit Policy

Looking for the optimal audit policy is now straightforward:

2 For $n > \bar{p}_C$; the enforcement agency applies the policy given by (1); (2) and (3):

2 For $n < \bar{p}_C$; the enforcement agency has to choose between $p_1 = p_2 = n$ and $p_2 = \bar{p}_C$:

The situation is very similar to the proof of proposition 3. Therefore, there exists a

unique $\frac{1}{2}\bar{p}_C < n^* < \bar{p}_C$; solution of the equation

$$\mathbb{B}(2n - \bar{p}_C) + \mathbb{B}_2(\bar{p}_C) = 2\mathbb{B}(n);$$

such that the optimal choice is given by:

$$\text{For } n \leq n^* : p_1^* = p_2^* = n;$$

$$\text{and for } n > n^* : p_2^* = \bar{p}_C \text{ and } p_1^* = 2n - \bar{p}_C;$$