Efficient Black Markets

by

Carl Davidson, Lawrence Martin, and John D. Wilson

Department of Economics
Michigan State University
East Lansing, MI 48824

November, 2002
Revised February 2003
1. Introduction

Since Becker (1968), a substantial body of research has emerged to explain the existence of illegal activity and the optimal punishment of this activity. Perpetrators of illegal activity, such as tax evaders, face some probability of apprehension, conviction, and punishment. Becker argued that fines represent a superior method of punishment. Since the probability of detecting perpetrators involves socially costly activities, whereas a fine simply transfers income, it makes sense to increase the fine and reduce the detection probability. At the optimum, the fine must equal its feasible maximum. In a recent survey, Polinsky and Shavell (2000) provide a proof that the optimal fine is indeed set at its upper limit when offenders are risk neutral. Comparing this result with actual practice, they observe, “Substantial enforcement costs could be saved without sacrificing deterrence by reducing enforcement effort and simultaneously raising fines” (p. 72).

If individuals are risk averse, then fines can increase risk-bearing costs. Polinsky and Shavell (1979) stress that the argument for severe fines should be modified in this case. Another argument stresses the need for marginal incentives not to commit more serious offenses (Shavell 1991, Mookherjee and Png 1992). For example, if a trucker’s fine for exceeding the weight limit takes all the assets of the firm, then once the limit is exceeded there is no marginal incentive for hauling additional weight. For a crime like tax evasion by risk neutral firms, however, all existing work implies that the optimal fine seizes all assets of the convicted evader.¹

The current paper presents a model in which illegal activity takes the form of tax evasion. Our use of the model is twofold. First, we show that it is not optimal to impose the maximum fine on tax evaders, although firms are risk neutral. The basic reason is that tax evaders anticipate the possibility of a fine and would therefore undertake actions that raised the social cost of

¹ Boadway et al. (2000) show that an increase in the fine can, surprisingly, lead to increased tax evasion. In their model, tax evasion requires cooperation by both the buyer and seller, and the fine affects the ability of pairs of tax evaders to sustain cooperation. They do not perform a normative analysis, however.
production if the fine were set too high. These adverse incentive effects arise from the actual fine, rather than its expected value.

Second, we show that not only may the government wish to keep the fine below its maximum level, but it may also wish to reduce detection activities to levels that enable tax evasion to flourish, even if detection is costless. To illustrate these results, we consider the case where tax evasion takes the form of a black market, in which no transactions are reported to the tax authority. The government uses the black market to circumvent information asymmetries that are restricting its power to tax. In particular, it may wish to tax particular market transactions at relatively low rates but cannot do so because it encounters difficulties in distinguishing these transactions from others. Our model illustrates this problem by assuming consumers purchase goods with “qualities” that are unknown to the government at the time it collects taxes. By allowing agents to self-select into the black market, the government targets tax breaks to the desired transactions, thereby bringing the tax system closer to a tax system that optimally discriminates among different transactions.

Previous literature has also shown how illegal activities can enhance welfare. In the smuggling literature, for example, several papers show that evasion of tariffs can move prices below the tariff-inclusive world price and bring about improved resource allocation. The argument in this work is that tariffs represent a sub-optimal policy instrument and evasion mitigates the damage. Our result, on the other hand, is that even when the government is optimally financing its expenditure requirements, given its available information, some tax evasion may be efficient, even in cases where it is costless.

In the optimal commodity and income tax literature, tax evasion may be beneficial simply because the second-best nature of the problem implies that it is not always optimal to treat

---

2 See Bhagwati and Hansen’s (1973) model, where smuggling drives out legal trade entirely, or Pitt’s (1981) model, where smuggling coexists with legal trade at a price below the tariff-inclusive world price. See also Martin and Panagariya (1984).
identically the consumers or suppliers of particular commodities. The representative-consumer models of optimal commodity taxation, where only efficiency matters, focus on the moral hazard problems created by a commodity tax system. The optimal income tax literature typically assumes that innate differences between workers cause incomes to differ, in which case adverse selection problems are added to the analysis. Our model follows the optimal commodity tax literature by focusing on only efficiency issues (risk-neutral firms and consumers), but it departs from this literature by making adverse-selection problems a major part of the analysis. In particular, the black market is designed so that those firms that select it are the ones that the government desires to tax relatively lightly.3

In the next section, we present the model and use it to prove that maximum fines are not desirable. Section 3 briefly describes some alternative models that yield the same message. Section 4 illustrates the usefulness of a black market in the special case where there are two taxed goods, and then Section 5 extends the analysis to a continuum of goods. Section 6 concludes.

2. The Model and Optimal Fine

Consider an economy with a continuum of consumers, indexed by a taste parameter \( \alpha \). Each consumer is endowed with \( w \) units of a composite commodity, or “endowment good,” which is used to produce a public good and private goods of variable quality, via a linear technology. The consumer purchases either zero or one unit of a variable-quality good at price \( P(\theta) \), which depends on quality. If \( P(\theta) \) denotes this dependence, then a good with quality \( \theta \) yields utility

\[
U(w - P(\theta), \theta; \alpha) = w - P(\theta) + \alpha v(\theta) \tag{1}
\]

3 For an analysis of random income taxation, see, for example, Brito et al. (1995). Cremer and Gahvari (1995), among others, explicitly incorporate tax evasion and costly audits into an optimal nonlinear income tax model, thereby extending the work on linear income taxation originally undertaken by Sandmo (1981). In contrast to our work, none of these papers consider the taxation of risk-neutral firms.
for a type-$\alpha$ consumer, where the function $v$ is concave. The public good is suppressed from the utility function, because it is held fixed throughout the analysis. The parameter $\alpha$ possesses a continuous distribution, $h(\alpha)$, on the unit interval, $[0,1]$, and the population is normalized to equal one. From (1), higher values for $\alpha$ represent a greater marginal willingness to pay for quality. For the case in which quality is a continuous variable and the price schedule is differentiable (we consider both cases), utility maximization yields the following first-order condition:

$$\alpha v'(\theta) = P'(\theta).$$  

(2)

By the second-order condition, the chosen $\theta$ is an increasing function of $\alpha$.

Each unit of a good with quality $\theta$ is produced at a constant unit cost, represented by the function $C(\theta)$, where $C'(\theta) > 0$. For simplicity, we assume that $C(\theta)$ is also strictly convex over positive $\theta$. However, we do allow for the existence of some costs that are independent of quality, i.e., $C(\theta)$ converges to a positive number as quality goes to zero. This assumption implies a U-shaped curve for average cost $C(\theta)/\theta$. More general forms of $C(\theta)$ that generate this U-shape over the relevant range could be used, but at the cost of added expositional complexity.

There is a minimum quality, $\theta_m > 0$, that is verifiable. For example, only cars with qualities at least equal to $\theta_m$ can be driven off the lot. Any higher quality is private information of the seller at the time of sale. The buyer perceives the actual quality through the experience of consuming. We assume that the lowest quality chosen by consumers, $\theta_L$, is greater than $\theta_m$, thereby ensuring that there is an information asymmetry at all chosen qualities.

This information asymmetry leads sellers to offer “guarantees.” In particular, customers are informed that they will receive a payment if they do not receive the reported quality. This payment must be large enough to ensure that the firm does not have an incentive to reduce quality to $\theta_m$. In the absence of tax evasion, the payment is
\[ B^l(\theta) = C(\theta) - C(\theta_m), \]  

where the superscript \( l \) indicates that the firm is not evading taxes, i.e., the firm is in the “legal sector.”

Such guarantees are useless unless the firm has sufficient assets to honor them. Thus, we assume that the firm accumulates assets worth \( B^l(\theta) \), in addition to undertaking activities designed to convince consumers that these assets exist. The combined cost of holding these assets and conveying this information to consumers is \( rB^l(\theta) \), which we term the “carrying cost.” One example of such assets would be a bond, whereby the firm gives money to a third party, such as a bank, with instructions to turn over the money to customers if the bank ascertains that the firm has cheated on quality. In this case, the carrying cost should be low. For concreteness, the bond interpretation is used in the subsequent analysis. But another example would be real estate. In the current static model, the purchase and subsequent sale of some of the endowment good could be used to represent real estate operations, in which case \( r \) would represent the various transactions costs involved in these activities. In both examples, the basic idea is that it is relatively difficult for firms to quickly hide or dissipate these assets in an effort to avoid honoring the guarantee. The next section considers a dynamic version of the model, in which the assets represent the future value of the firm itself. We could add a cost of obtaining the assets (e.g., litigation expenses) without affecting the results, however, because the incentive effects from the guarantee lie not in how much consumers obtain, but in how much firms give up.

The government’s expenditure needs are financed by taxing the revenue firms earn from the sale of variable-quality goods at a constant rate \( t \). The tax authority cannot observe \( \theta \). As discussed below, it can perform an audit to detect a firm’s earned income, but it does not know the number of customers at the firm and therefore cannot deduce \( \theta \) from its audit.

Turning to market structure, we may assume either a large number of competitive firms, or we may allow for Bertrand competition, under which product prices serve as strategic
variables. In either case, firms sell their output at a price equal to unit cost, which consists of taxes, the input cost \( C(\theta) \), and the bond’s carrying cost, denoted \( rB'(\theta) \), \( 0 \leq r < 1 \). Thus, the price in the absence of tax evasion, or the “legal price,” is given by

\[
P_l(\theta)(1-t) = C(\theta) + rB'(\theta)
\]  

(4)

Firms may evade taxes by entering into an illegal “tax shelter,” which means that an exogenous percentage of their income is not reported. For concreteness and notational simplicity, we assume that this exogenous percentage is 100 percent, i.e., the firm enters a “black market.” By doing so, firms risk detection and punishment. The government audits a fraction \( \pi \) of firms and assesses monetary fines. The audit reveals the firm’s total assets. The firm’s scale of production is irrelevant, given the assumption of a linear transformation between the endowment good and output. Thus, we may examine assets per unit of output sold. These assets consist of the bond and the revenue obtained by selling the good at the black-market price, \( P_b(\theta) \). A fine is paid at the rate \( f \) on these assets. Note that the firm owners are not able to pass the burden of the fine on to input suppliers by reneging on the payment of \( C(\theta) \) and instead allowing the government to use this amount to pay the fine. One may assume that the firm owner is also the supplier of the endowment good (“self-employment”), or that the firm owner sells some of his own endowment good to other firms and uses the proceeds from this sale to purchase the endowment good, prior to the production and sale of output.

However, the bond can be used to pay the fine. A firm that cheats on quality surrenders the bond to consumers only after production has occurred, sales have been made, and consumers have “experienced” the good. Before then, the government conducts its audit and assesses the fine. Consumers therefore take into account the possibility that a black-market firm will have to surrender a portion of the bond to the government, leaving them with a less-than-full guarantee. Consistent with practice, we are assuming that the tax collector is first in line among the firm’s
creditors. Note here that consumers know which firms are black-market firms, because only firms with particular quality levels choose to operate in the black market (as we show below), and each firm guarantees its quality in equilibrium.

To summarize, an audit reveals a firm’s assets, \( P_b(\theta) + B_b(\theta) \), and then the coercive powers of the state are used to collect a fine at rate \( f \) on these assets. The firm’s owners are assumed to be able to avoid surrendering other assets, either due to limited-liability rules or their ability to hide such assets from the tax collector. In this way, the model places a ceiling on the severity of the fine. Such ceilings typically represent binding constraints in the literature on crime and punishment, since that literature advocates raising fines as much as possible, to minimize the enforcement costs needed to achieve the desired deterrence. In contrast, we shall argue against the usefulness of a 100 percent fine on those assets to which the government has access.

A central insight from this model is that the fine and audit probability have important implications for the size of the bond needed to guarantee quality. If the firm knows that some portion of its bond is going to be paid out to the government if it is audited, regardless of whether it cheats on quality, then the effective penalty from cheating on quality will be less severe. Let us therefore solve for the size of the bond needed to guarantee quality in the black market. If the firm cheats on quality, then its production costs are \( C(\theta) \). With probability \((1 - \pi)\), it is not audited and pays out the bond \( B^b \) to customers. With probability \( \pi \), it is audited and pays the government a total fine equal to \( F(\theta) = f(P_b(\theta) + B_b(\theta)) \). For zero expected profits, the unit price must satisfy \( P_b(\theta) = C(\theta) + rB^b(\theta) + \pi f(P_b(\theta) + B^b(\theta)) \); or,

\[
P_b(\theta) = \frac{C(\theta) + (r + \pi f)B^b(\theta)}{1 - \pi f}.
\]  (5)
If $F(\theta)$ is no greater than $P^b(\theta)$, then it can be paid out of revenue alone, allowing the bond to be handed over to the customers. In this case, the penalty from cheating is $B^b(\theta)$, regardless of whether the firm is audited, and so $B^b(\theta)$ must satisfy (3) to induce the firm not to cheat. Suppose, however, that $F(\theta)$ exceeds $P^b(\theta)$. Using the formula for $F(\theta)$, we can calculate the excess payment that must come from bond assets:

$$F(\theta) - P^b(\theta) = fB^b(\theta) - (1-f)P^b(\theta).$$

(6)

Substituting from (5) for $P^b(\theta)$ on the right side of (6), we obtain the following condition for $F(\theta)$ to be at least as great as $P^b(\theta)$:

$$f(1-\pi r f) \geq (1-f)\frac{C(\theta) + (r + \pi r f)B^b(\theta)}{B^b(\theta)}.$$  

(7)

If we hold fixed the expected fine per unit of assets, $\pi r$, this condition tells us that raising the fine rate and lowering the audit probability increases the likelihood that the firm will be thrown into the range in which it will have to use a portion of the bond to pay the fine. On the other hand, a large cost $C(\theta)$ relative to the bond obviously reduces this likelihood.

If auditing involves any costs, then the government will clearly desire to raise the fine until (7) is satisfied at least with equality. Up to that point, there is no cost to raising the fine. But it is not clear that the government would desire to raise the fine above this level (where $F(\theta) > P^b(\theta)$). To investigate this issue, we first solve for the minimum bond needed to ensure quality if the inequality holds. For a firm that has cheated, the expected payout from the bond is $(1 - \pi)B^b(\theta) + \pi[B^b(\theta) - (F(\theta) - P^b(\theta))]$. Using (6), this is equivalent to

$$(1 - \pi)B^b(\theta) + \pi(1-f)[B^b(\theta) + P^b(\theta)]$$

(8)
Substituting for \( P^b(\theta) \) from (5), we then obtain an expression for the expected payout, which must exceed the cost savings from cheating on quality:

\[
\left(1 - \pi + \frac{\pi - \pi f}{1 - \pi f}\right)B^b(\theta) + \frac{\pi - \pi f}{1 - \pi f}\left(C(\theta) + rB^b(\theta)\right) \geq C(\theta) - C(\theta_L). \tag{9}
\]

The right side reflects the absence of any tax benefits or costs from cheating on quality: the fine is based on revenue \( P^b(\theta) \) and the bond \( B^b(\theta) \), which are both determined by the quality that has been guaranteed by the firm, not the actual quality delivered.

It is immediately evident that if there is no chance of detection (\( \pi = 0 \)), then the bond equals the difference between \( C(\theta) \) and \( C(\theta_L) \). But if detection is possible and the fine is large enough to satisfy (7) with a strict inequality, then this condition implies that the \( B^b(\theta) \) in (9) exceeds the cost difference on the right side. Thus, the fine itself can generate a positive resource cost, in the form of the higher carrying cost associated with a larger bond. In contrast, previous literature emphasizes the benefit of the fines over detection activities by appealing to the cost-free nature of fines. The current paper helps level the playing field.

Indeed, the following proposition states that if we assumed that detection is costless, thereby leveling the playing field according to the previous literature, then we would in fact overshoot a level playing field:

**Proposition 1.** If detection is costless and the government’s goal is to achieve a given expected fine, \( \pi f < 1 \), at the lowest social cost, then the government should never set \( f \) equal to one. In other words, if \( f = 1 \), then raising the audit rate and lowering the fine rate could always reduce social cost.
Proof. As already described, the social cost referred to here is created by the positive impact of the fine on the equilibrium size of the bond. If \( f = 1 \), then (7) must hold, indicating that \( B^b \) exceeds \( C(\theta) - C(\theta_L) \). Holding \( \pi_f \) fixed, the derivative of the left side of (9) with respect to \( \pi \) is

\[
\left( -1 + \frac{1}{1-\pi_f} \right) B^b (\theta) + \frac{1}{1-\pi_f} \left( C(\theta) + rB^b(\theta) \right),
\]

(10)

which is positive. In other words, increasing \( \pi \) enables us to satisfy (9) with a smaller bond. As a result, carrying costs fall. Q.E.D.

3. Alternative Models

A. A Dynamic Model

The quality guarantees can be modeled in a dynamic context, with no change in results. In particular, suppose that the economy lasts forever, with the events previously described taking place in each period. Time is measured in discrete intervals, and we consider a steady-state equilibrium over an infinite horizon.

In any given period, customers must again be able to obtain assets from a firm in the event that it has cheated on quality. But if the firm can earn future profits, then the firm has a positive present value, which can be realized immediately by selling the firm. It follows that goods will be sold at a price greater than cost. The reason is that the firm itself is a valuable asset, which can be used to back guarantees on quality. In the absence of tax evasion, the firm’s discounted future value is

\[
B^q(\theta) = \left( \frac{1}{r} \right) \left[ P^q(\theta)(1-t) - C(\theta) \right],
\]

(11)

where \( r \) is the discount rate, and the term in the square brackets is after-tax profits. The no cheating condition is then
\[ [P(\theta)(1-t) - C(\theta)] + B'(\theta) \geq P(\theta)(1-t) - C(\theta_m), \] 

which implies, \( B'(\theta) \geq C(\theta) - C(\theta_m) \), as before. Moreover, tax evasion and enforcement activities affect this inequality in the same way as before.

It is clear then that the argument against high fines remains unchanged. Consumers recognize that a high fine will limit the ability of the firm to maintain its guarantee. As a result, this size of the guarantee must increase, which here takes the form of higher profits. Although the increase in profits is a transfer from consumers to firm owners, rather than a social cost, the resulting rise in product prices distorts consumption decisions. By keeping the fine sufficiently low, the government can minimize the size of this distortion, since the excess profits are then fully available for use as the guarantee.

In standard full-information models with Bertrand competition, excess profits cannot exist, because competitors will attempt to undercut each other’s price. In the current model, any undercutting would drive away customers, because they would then reason that the firm lacks an incentive not to cheat on quality. But with positive profits and constant unit costs, a new problem arises -- it is not at all clear how the market shares of different firms are determined. One possibility is that market share is effectively auctioned to the highest bidder, perhaps through expenditures on “advertising,” resulting in an equilibrium where the discounted value of profits equals the initial payments used to obtain market share. In this case, the profits identified above now represent a social cost, which is increased further through the use of excessive fines. With this interpretation, the cost of excessive fines is even higher than in the previous model.

B. Reputation

Instead of offering guarantees, firms may seek to establish a reputation for selling goods with the reported qualities. If a firm reports that it’s output possesses quality \( \theta > \theta_m \) but then sells a good with quality \( \theta_m \), it will make greater profits in the current period. If, however,
consumers do not believe any future reports, the firm will be unable to sell at a price above \( P(\theta_m) \) in the future, implying that its future profits drop to zero. Since we assume that consumers do not buy this minimum quality, this means that the firm exits the industry, with a new firm entering to provide the original quality. To prevent cheating, the firm’s one-period gain from cheating on quality must be no greater than the lost future sales. As a result, (11) and (12) continue to give the present value of future profits and the no-cheating condition, and (12) again implies \( B^f(\theta) = C(\theta_m) - C(\theta) \). Thus, the excess profits needed to ensure quality are identical to those in the previous model. In other words, reputation does no better or worse than guarantees.

However, the models differ in their implications for tax evasion and enforcement. Suppose, as before, that the fine is levied on total assets, \( P^b(\theta) + B^b(\theta) \). The no-cheating (on quality) condition is

\[
[P^b(\theta)(1 - \pi f) - C(\theta)] + B^b(\theta)(1 - \pi f) \geq P^b(\theta)(1 - \pi f) - C(\theta_m).
\]  

(13)

Solving for \( B^b(\theta) \) gives,

\[
B^b(\theta) = (1/r)[P^b(\theta)(1 - \pi f) - B^b(\theta)\pi f - C(\theta)].
\]

or.

\[
B^b(\theta) = \frac{1}{r} \left[ P^b(\theta)(1 - \pi f) - C(\theta) \right].
\]

In contrast to the previous models, only the expected fine matters, and this fine raises the required level of profits. In particular, (14) gives

\[
B^b(\theta)(1 - \pi f) = C(\theta) - C(\theta_m).
\]  

(15)

Unlike the model with guarantee, the profits in the reputation model are not distributed to consumers in the event of cheating. Rather, they go to a new firm, independent of the level of \( f \).

However, we are perhaps not giving the government enough power to levy fines. Consider an alternative strategy. Once an audit has occurred, the government obtains a value for
the firm, perhaps by putting it up for sale. If the firm has no value, then the government takes one
hundred percent of the revenue earned in the current period. If the firm does have value, then the
government sets a fine below 100 percent. The resulting no-cheating condition becomes
\[
[P^\text{th}(\theta)(1 - \pi_f) - C(\theta)] + B^\text{th}(\theta)(1 - \pi_f) \geq P^\text{th}(\theta)(1 - \pi) - C(\theta_m). \tag{16}
\]
where $B^\text{th}(\theta)$ is again given by (14).

We now see that the government’s enforcement policy is not only penalizing tax cheaters,
but also quality cheaters, and it does this through the use of differential fines. To maximize the
penalty associated with a given expected fine, $\pi_f$, the government should raise the audit rate to the
point where all profits are eliminated (i.e., both $B^\text{th}(\theta)$ and each side of (16) equals zero), unless
the audit rate first equals the tax rate, $\pi = t$. Any higher audit rate would induce a quality-cheater
to pay the tax $t$, thereby rendering the fine irrelevant as a deterrent to quality cheating.

In equilibrium, the 100 percent fine on quality cheaters is never observed, since no firm
cheats. Rather, the lower fine $f$ is observed. Once again, we see that relatively low fines can be
desirable. In particular, they represent a device for rewarding desirable private-sector behavior.
In the current model, firms that are not cheating customers are the ones that have a positive value,
and therefore fines should be kept low on these firms.

C. Differential Fines for Approximating a Nonlinear Tax System

The previous section discussed the use of differential fines on firms that cheat on quality
and those that do not. Another form of differentiation would be to use quality-specific fines, $f(\theta)$
for a firm that provides quality $\theta$. Consider the original model from the previous section (though
similar points can be made with the other models). In this case, the enforcement system could be
used to approximate an optimal commodity tax system, where each good is taxed at a different
rate. The effective tax rate on good $\theta$ would be the expected fine $\pi_f(\theta)$. It is easy to show that
under this system, the highest and lowest quality goods purchased by consumers should be taxed.
at a zero marginal rate, a result that corresponds to the finding in optimal income tax theory that the highest and lowest incomes should face a zero marginal tax rate. Thus, we now see that not only should fines be kept below 100 percent, but sometimes they should be reduced to zero!

Given that both audits and large fines are costly in this model, however, it will not be desirable to exactly replicate the optimal commodity tax system. But neither is there a role for setting the uniform tax rate low enough to eliminate tax evasion by some firms. Since all firms are randomly audited, regardless of their behavior, creating a legal sector in this model would not lower audit costs. The model could be complicated in a way that gave rise to a legal sector, but it seems instructive to investigate further the role of a black market under the simple (and therefore more feasible) form enforcement discussed in the previous section. This task is taken up in the next section, but first we discuss one more alternative model.

D. Debt

The basic problem with fines in Section 2 is that they can raise the costs needed to guarantee quality. In particular, fines increase the assets that firms must carry in case they are needed to compensate consumers. A similar problem exists for a firm that undertakes capital investments financed by debt. Consider an industry in which firms employ capital and labor to produce output. Suppose that capital expenditures are financed by bank loans, and that the capital assets themselves provide collateral. Suppose that some firms evade taxes. If detected, they must pay the fine. If the fine is less than the firm's ability to pay (in our model, the revenues on hand), then there is no impact on the cost of capital. On the other hand, if the fine exceeds the ability to pay, the tax authority may seize the capital assets of the firm, even though they needed to repay the loans. The tax authority stands in line ahead of private creditors. Thus severe fines represent a problem for creditors, as they make it more likely that the full amount of the loan will not be repaid. This possibility increases the cost of capital for the industry. In particular, if labor is paid in full prior to the audit, then the burden of the audit on capital will distort the firm’s mix
of labor and capital towards too little capital and too much labor. In a two-sector model, where the sector just described evades taxes but the other sector pays them, this factor-market distortion will move the economy inside its PPF. Once again, there is a cost to large fines.

4. Two Qualities

We now return to the model introduced in section 2 and use a simple example with two qualities, \( \theta_H > \theta_L \), to investigate conditions under which the government will choose to tolerate a black market, even in cases in which auditing activities involve little or no cost. We find that if a black market does arise, only low-quality firms populate it. Since consumers are aware that only two qualities are sold, there is no need for the low-quality firms to guarantee quality. Hence, the previous link between the fine and the size of bonds does not arise. We therefore simplify the analysis by assuming no carrying costs \((r = 0)\). To shorten notation, we let \( C_H = C(\theta_H) \), \( C_L = C(\theta_L) \), \( v_L = v(\theta_L) \), and \( v_H = v(\theta_H) \).

Before deriving equilibrium, let us first consider a hypothetical situation in which the tax authority prevents tax evasion but contrary to our information assumptions, is able to tax goods with different qualities at different rates. Zero profits in equilibrium implies the following prices:

\[
P_j(1 - t_j) = C_j \quad \text{for } j = L, H
\]

Equilibrium must also satisfy self-selection constraints. Let \( \alpha_H \) be the \( \alpha \) possessed by a consumer who is indifferent between the low- and high-quality goods, and let \( \alpha_L \) be the \( \alpha \) for a consumer who is indifferent between the low-quality good and consuming neither good. As in the continuous quality case, high-\( \alpha \) consumers tend to choose a high quality, except that the existence of only two qualities implies that consumers are now bunched at these qualities. In particular, a consumer with an \( \alpha > \alpha_H \) buys the high-quality good, a consumer with \( \alpha \in [\alpha_L, \alpha_H] \) buys the
low-quality good, and a consumer with $\alpha < \alpha_l$ buys neither good. Indifference between the low- and high-quality goods requires identical consumers’ surpluses:

$$\alpha_{hvH} - P_H = \alpha_{hvL} - P_L,$$  \hspace{1cm} (18)

or,

$$\alpha_H = \frac{P_H - P_L}{v_H - v_L}. \hspace{1cm} (19)$$

For the consumer who is just indifferent between consuming the low-quality good and neither good, consumer surplus must be zero:

$$\alpha_{lvL} - P_L = 0,$$  \hspace{1cm} (20)

or,

$$\alpha_L = \frac{P_L}{v_L}. \hspace{1cm} (21)$$

Figure 1 shows indifference curves for type-$\alpha_L$ and type-$\alpha_H$ consumers.

The optimal tax problem consists of choosing taxes $P_L$ and $P_H$ to maximize total consumer surplus, subject to a government budget constraint:

$$\text{Max} \quad \int_{\alpha_L}^{\alpha_H} (\alpha v_L - P_L)h(\alpha)d\alpha + \int_{\alpha_H}^{1} (\alpha v_H - P_H)h(\alpha)d\alpha$$

$$\text{s.t } G = (1 - \alpha_H)(P_H - C_H) + (\alpha_H - \alpha_L)(P_L - C_L), \hspace{1cm} (22)$$

where $G$ represents the government’s revenue requirement and $\alpha_H$ and $\alpha_L$ are given by (19) and (21). The unit tax rates are then $t_H P_H = P_H - C_H$ and $t_L P_L = P_L - C_L$. Attaching the Lagrange multiplier $\lambda$ to the constraint, we obtain the following first-order conditions for $P_H$ and $P_L$: 
\[
\frac{\lambda - 1}{\lambda} = \frac{h(\alpha_H)(t_H P_H - t_L P_L)}{[1 - H(\alpha_H)](v_H - v_L)} = \frac{h(\alpha_L)(t_L P_L - t_H P_H)}{[H(\alpha_H) - H(\alpha_L)](v_H - v_L)} + \frac{h(\alpha_H)t_L P_L}{[H(\alpha_H) - H(\alpha_L)]v_L}
\] 

(23)

Note that changes in the limits of integration in the objective function have no first-order impact on consumer surplus. This is an envelope-theorem result: since agents optimally chose whether to consume one or none of the two qualities, small tax-induced changes in these decisions cannot affect welfare, if taxes are held fixed. However, (23) does reflect the impact of tax-induced changes in behavior on the budget constraint.

With this first-order condition, we describe the optimal tax system as follows:

**Proposition 2.** The optimal discriminatory tax system satisfies the following tax rule:

\[
\frac{t_H}{t_L} = \frac{v_H}{P_H} \left\{ \frac{h(\alpha_L)[1 - H(\alpha_H)] v_H - v_L}{h(\alpha_H)[1 - H(\alpha_L)] v_L} + 1 \right\}
\]

(24)

Furthermore, \( t_L < t_H \) if

\[
\frac{\alpha_L h(\alpha_L)}{1 - H(\alpha_L)} > \frac{\alpha_H h(\alpha_H)}{1 - H(\alpha_H)}.
\]

**Proof.** Multiplying (23) by a common denominator gives

\[
v_H h(\alpha_H)[1 - H(\alpha_L)](t_H P_H - t_L P_L) - (v_H - v_L) h(\alpha_L)[1 - H(\alpha_H)] t_L P_L = 0.
\]

(25)

Solving (25) for \( t_H/t_L \) yields (24). To complete the proof, assume that the right-hand-side of (24) is less than one and make use of the equations given for \( \alpha_L \) and \( \alpha_H \) in (21) and (19).

Q.E.D.
This optimal tax rule tells us that the good that should be taxed more heavily depends upon the distribution of the taste parameter $\alpha$ and the equilibrium values for $\alpha_L$ and $\alpha_H$. It should be clear that the government will want to tax the low-quality good more lightly if $h(\alpha_H)$ is small compared to $h(\alpha_L)$. A distribution function with this property is depicted in Figure 2.

Because the tax authority cannot observe quality, this direct form of tax discrimination is not possible. With tax evasion, however, limited amounts of discrimination are available. In particular, we now argue that the introduction of a black market enables us to tax the low quality good relatively lightly, which improves welfare if the conditions given in Proposition 2 hold.

Thus, let both goods be taxed at a uniform rate, $t$, and choose the fine and audit probability so that the expected fine, $\pi_f$, is slightly less than $t$. For this system to replace a discriminatory tax system with the low-quality good subject to the lower tax rate, two questions must be answered affirmatively: First, will low-quality firms choose to evade taxes? Second, will high-quality firms choose not to evade taxes? The answer to the first question is “yes,” because the expected fine is less than the taxes owed under the uniform tax ($\pi_f < t$), and the low-quality firms base their decision on only this difference.

Turning to high-quality firms, observe that if they enter the black market, then they pay an expected fine given by $\pi_f(P_H + B^b)$, where $B^b$ is the bond required to guarantee quality in the black market. Recall that the black-market bond is at least as large as the cost savings from cheating on quality: $B^b \geq C_H - C_L$. In particular, we have seen that the possibility of a fine may raise the size of the bond that the firm is required to pay. However, the firm never actually has to surrender the bond to consumers, since it does not cheat in equilibrium, and we have assumed zero carrying costs in this section. Thus, moving to the black market does not raise the costs associated with issuing the bond. We therefore need only to compare expected tax burdens between the two markets to see which is preferred. Having bounded the size of the bond from below, a sufficient condition for the legal market to be at least as desirable is
\[ \pi f(P_H + C_H - C_L) \geq tP_H, \]

or,

\[ \frac{C_H - C_L}{P_H} \geq \frac{t - \pi f}{\pi f}. \]

(26)

For a sufficiently small fine \( f \), where (7) is not satisfied, this condition also becomes necessary, since then \( B^b = C_H - C_L \).

This condition is satisfied if the difference between \( \pi f \) and \( t \) is sufficiently small. Thus, we can replicate a discriminatory tax system if \( t_L \) is not too far below \( t_H \). The high-quality firms choose the legal market, paying the uniform tax \( t = t_H \), and the low-quality firms choose the black market, paying the expected tax \( \pi f = t_L \).

Proposition 2 offers a necessary condition for a black market to be welfare improving. One example satisfying this condition is illustrate in Figure 2. This condition implies that \( t_L < t_H \) under the optimal discriminatory tax system. In this case, we can start with an optimal uniform tax system, where all goods are taxed at rate \( t \), and then introduce a black market by setting \( \pi f \) slightly below the initial \( t \), while raising \( t \) to offset the revenue loss. Such a tax system raises welfare because the initial uniform tax violates the optimal tax rule, with the right side of (24) higher than the left side. This violation is easily seen to imply that a revenue-neutral tax change towards a lower \( t_L \) and higher \( t_H \) raises welfare. The introduction of the black market in the manner just described effectively implements such a tax change.

Note, however, that a black market is not always desirable. If the government desires to tax the low-quality good relatively heavily, then it cannot achieve this goal with a black market. As we have seen, only low-quality firms (or all firms) desire to produce in the black market, and these firms would not choose to do so if the expected fine \( \pi f \) were greater than the tax \( t \). Finally,
we have not shown that an optimal discriminatory tax with $t_L < t_H$ is achievable using a black market. But the economy can at least move towards such a tax.

5. Continuous Quality

We now assume that quality varies continuously, in which case consumers make their quality choices according to (1). The decision about whether to operate in the legal or black markets is based on a comparison of the tax payments and expected fines. There is also the issue of increased bond costs, dealt with in Section 2, but we shall assume for simplicity that the government is free to raise the audit probability enough to keep the fine below the point at which it starts interfering with the bond guarantee (i.e., below the minimum level that satisfies (7)). In this case, the bond satisfies $B(\theta) = C(\theta) - C(\theta_m)$ in both the black and legal markets, and any good $\theta$ produced in the legal market satisfies

$$\pi_f(P_l(\theta) + C(\theta) - C(\theta_m)) \geq tP_l(\theta). \quad (27)$$

Similarly, any good $\theta$ produced in the black market satisfies

$$\pi_f(P_b(\theta) + C(\theta) - C(\theta_m)) \leq tP_b(\theta). \quad (28)$$

The legal- and black-market price schedules are drawn in Figure 3. Where they cross, denoted $\theta^*$, both (27) and (28) hold, with firms indifferent between the two markets. A critical observation here is that the slope of $P_b(\theta)$ is greater than the slope of $P_l(\theta)$ where they cross, implying only a single crossing. This property can be checked by differentiating the price schedules given by (4) and (5), and evaluating the derivatives at $\theta^*$. The basic idea is that the black-market price schedule includes a fine based not only on the revenue obtained from sales of
the good, but also the bond, and the latter also rises with \( \theta \), thereby requiring a larger increase in the break-even price.

These properties suggest that the schedule of market prices will be the lower envelope of the black- and legal-market price schedules. But then firms operating above \( \theta^* \) must actually prefer the legal market, whereas those below \( \theta^* \) opt for the black market. These properties follow from (27) and (28), because whenever the two sides of either equation are equal, the left side rises more than the right side as \( \theta \) increases. Once again, the basic idea is that the tax base for the fines in the black market includes not only the firm’s revenue but also its bond assets, which become an increasingly large share of total assets as \( \theta \) rises. Thus, the black market becomes increasingly unattractive as \( \theta \) rises. In equilibrium, firms operating above \( \theta^* \) sell in the legal market, while those below \( \theta^* \) are in the black market.

However, no consumer buys \( \theta^* \) or any quality level nearby. The reason is that the price schedule faced by consumers has a kink at \( \theta^* \). For this reason, if both legal and black-market goods are consumed, there will be a consumer who is indifferent between consuming a quality, \( q^l \), which is the lowest quality in the legal market, and a lower quality, \( q^b \), which is the highest quality in the black market. Between these two qualities, no goods are sold. In other words, black-market goods are lower (in quality) than legal goods by a discrete amount.

For this reason, introducing a black market may or may not improve welfare. On the one hand, the black market can be used to induce new customers to consume variable-quality goods, in which case they now become taxpayers. On the other hand, it can only do so by inducing existing customers to jump into the black market, thereby discontinuously lowering their tax payments.

Whether a black market is desirable depends on the relative importance of this latter discontinuity, which depends critically how close the quality levels chosen by consumers are to the verifiable quality. If we set the expected fine \( \pi_f \) equal to the uniform tax \( t \), then the tax and
expected fine on each unit of good $\theta_m$ are the same, since no bond is posted at $\theta_m$. As a result, $P^b(\theta)$ and $P^l(\theta)$ converge at $\theta_m$, as illustrated in Figure 4. If the minimum quality purchased in the legal market, $\theta_l$, is slightly to the right of $\theta_m$, then this fine is too high to generate a black market. However, it does not take much of a reduction in $\pi_f$ to shift $P^b(\theta)$ down enough to cause some consumers to switch from the legal market to the black market. These consumers pay slightly lower taxes. On the other hand, the fine reduction also attracts consumers who have not been consuming variable-quality goods, and the tax payments for each of them go from zero to some positive amount. This latter effect clearly dominates the former, provided $\theta_l$ is sufficiently close to $\theta_m$ initially. Nobody is made worse off, but total tax revenue rises, and this new revenue can be distributed in the form of a reduction in all taxes and fines, making everyone better off.

As with the two-quality case, we once again see a role for a black market, because it allows the government to circumvent restrictions on its power to tax. Because there are only limited ways in which the black market can fulfill this role, it may not always be useful.

Paradoxically, expanding the government’s power to tax increases the likelihood that the black market will serve a useful role. We saw previously that a black market is necessarily desirable if the government can choose a different fine for each quality level. A more limited way to effectively base fines on quality is to distinguish between those assets financed by sale of the good and assets representing the bonds posted to guarantee quality. In this case, the government can base its fine of these two separate classes of assets, $P^b(\theta)$ and $B^b(\theta)$. Previously, we assumed that no such distinction was made, perhaps due to the added costs involved in making the distinction. If separate fines can be imposed on $P^b(\theta)$ and $B^b(\theta)$, then the government now has the power to adjust both the height and the slope of the black-market price schedule. It basically takes advantage of the fact that $B^b(\theta)/C(\theta)$ increases with $\theta$, implying that a fine on the bond becomes relatively more important at higher values of $\theta$. Thus, lowering the fine on the bond while raising the fine on revenue $P^b(\theta)$ enough to keep the total fine fixed at the given $\theta$ will
cause the slope of the price schedule to fall without changing the value of $P^\prime(\theta)$ at this $\theta$. Previous ambiguities arose because the slope of $P^\prime(\theta)$ was greater than the slope of $P'(\theta)$ by an amount that could not be independently manipulated. The government can now improve welfare by reducing the fine on the bond relative to sales revenue. In this way, it ensures that the black market is introduced at quality levels near the bottom of the range of legal qualities, in which case we have seen that welfare can be increased.

Finally, note that the power to distinguish between the bond and sales revenue suggests the power to similarly distinguish between production costs and the carrying costs of bonds in the legal market. This does not change our argument, however, because relative tax rates in the legal market need not be tied to relative fines in the black market.

6. Concluding Remarks

We have argued against the prevailing view that fines on illegal activities should be set as high as possible, thereby allowing the costs of detection to be reduced. We have also demonstrated the usefulness of a black market as a means of moving towards an optimal system of commodity taxes when informational asymmetries prevent the government from implementing such a system directly. It should be possible to extend both of these results to other settings, but we believe that our basic message about fines is particularly general. The traditional argument favoring audits over fines seems less clear-cut than currently thought.
References


Figure 1

Figure 2
Figure 3

Figure 4