Costly Auditing and the Revenue Function of Tariffs

William J. Polley[*]

Bradley University

October 2003

Abstract

I consider the problem of optimal taxation in an environment with costly monitoring. Because imports typically enter the country at locations which are more easily monitored, it may be less costly to enforce a tariff than an income tax. Given an exogenous amount of revenue to be raised, the government must balance the price distortion of the tariff against the potential savings in enforcement costs. The main finding is that when the cost saving is sufficiently large, the optimal tax system will include both income taxes and tariffs. The tariff rate decreases with increases in per capita income or reductions in income tax enforcement costs. This fits the stylized fact that countries rely on tariffs in the early stages of development and rely on less distortionary taxes as per capita income increases.

[*]I gratefully acknowledge my thesis committee, John Conybeare, Forrest Nelson, B. Ravikumar, and Steve Williamson. I especially thank my committee chair, Raymond Riezman, for many insightful comments. Versions of this paper have been presented in seminars at The University of Iowa and at the Midwest International Economics Meetings. All errors are my own. E-mail to: wpolley@bradley.edu.

1. INTRODUCTION

One implication of optimal taxation theory is that small economies should not erect trade barriers. Because small economies have no effect on world prices, the domestic price distortions caused by such actions cause utility to be lower even if the revenue generated by the tariff is returned to the agent as a lump sum rebate. In particular, the tariff can cause the home country to produce more of the imported good, decreasing total production valued at world prices and lowering utility. Even if production is unchanged, the representative agent is worse off because of the substitution effect due to the higher relative price of imports. However, a significant number of the very countries to which this warning applies (price takers) are the ones in which a high percentage of government revenue comes from taxes on international trade. Consider the following evidence both contemporary and historical.

According to the World Development Report (1992), there are large differences among countries in terms of the percentages of government revenue generated by tariffs. In general, OECD countries tend to have low percentages; all are under 10%. On the other hand, low income countries as a group have much more variability in tariff revenue. In some cases, over 50% of government revenue in these countries comes from taxes on international trade. Pakistan collects approximately 30% of government revenue from the use of tariffs. Uganda has the highest reported percentage at 75%. The percentage for the United States is 1.6%[1]. Thus, the evidence suggests that there are a number of price taking economies using tariffs when theory predicts that they would use a less distortionary tax for this purpose. Countries which can influence prices through the use of tariffs might be expected to do so for strategic reasons.[2] However, it is unlikely that these low income countries are using tariffs for strategic

[1] This issue of the report lists 1990 data.

[2] In some cases, becoming better off even when there is retaliation. See Kennan and Riezman (1988).

reasons.

As a historical example, figure 1 illustrates the main observation concerning trade protection in the United States. The figure shows the decline in the percentage of government revenue generated by customs from 1925 to 1970. Prior to this, the United States had generated a much larger share of revenue through tariffs. With the exception of the war of 1812 and a few years in the mid-19th century when land sales accounted for more revenue, approximately 90% of revenue was from trade taxes before the Civil War. After the Civil War, the percentages dropped to under 50% (Bureau of the Census 1975). Once the 16th amendment to the constitution was ratified in 1913, tariff revenue began to fall dramatically and became somewhat less volatile (in terms of percentages of total revenue). By the middle of the 20th century, tariffs were much less important for raising revenue relative to the new income tax.

What could cause these relationships to appear in the data? Riezman and Slemrod (1987) find that factors such as the literacy rate are significantly related to the percentage of tax revenue generated by trade taxes. They argue that collection costs are higher in countries with low literacy rates. The reason given is that implementing trade taxes requires dealing with a smaller number of more educated taxpayers at the points where imports enter the country. Lower collection costs might make tariffs a more attractive means of raising revenue. So while there may not be strategic reasons for these small countries to use tariffs, it may simply be optimal domestic tax policy to do so. Note, however, that there are many possible types of collection costs. These costs could be fixed or variable. A careful specification of the nature of these collection costs is necessary.

The idea of using collection costs to explain the use of tariffs is not new. Perhaps the first record of the idea is in Smith's (1776) *Wealth of Nations*, "First, the levying of it [a tax] may require a great number of officers whose salaries may eat up the greater part of the produce of the tax, and whose perquisites may impose another additional tax upon the people." (p. 453). More recently, Corden (1974) presents a graphical argument for using tariffs to raise revenue when collection costs of excise taxes are high enough. Gardner and Kimbrough (1992) present a model in which a certain fraction of the revenue generated by different taxes is lost. The analysis leads to the conclusion that the costs of income tax collection may at some point make income taxes less efficient, on the margin, than trade taxes. This result accords with the premise of Riezman and Slemrod (1987) and is a very plausible explanation of the use of tariffs to raise revenue.

The approach in Gardner and Kimbrough (1992), however, does not specifically take into account auditing and enforcement, two considerations which are intimately related to collection costs. The rest of this paper extends this research and gives explicit attention to auditing in constructing a theoretical model which explains the data. Rather than assuming that a certain fraction of the revenue is lost due to collection costs, this paper focuses on the costs of auditing different types of taxes. These auditing costs are presented in the framework of costly state verification similar to Townsend (1979). The reason to study audit costs rather than collection costs is that in the United States, income tax payment is largely voluntary in the sense that the government does not send tax collectors to confiscate the taxes directly. Rather, individuals make a report of their income to the government and send it along with their payment. Of course, the system has developed to the point where employers assist their employees by withholding the proper tax and sending it to the government. Even so, people can and do find ways to avoid paying some or all of the tax. If they are caught, the penalties for paying too little tax can include fines and/or jail time. Catching those who evade the tax requires that the government audit individual taxpayers. The government's decision of how frequently to audit is influenced by the cost of verifying the income of an individual taxpayer. Similarly, the taxpayer's willingness to comply is influenced by audit policy. The costly state verification model is a rich environment to study this interaction.

Townsend's (1979) model of costly state verification has become an important model for the study of debt contracts. However, there are also several examples of the costly state verification model's application to the optimal taxation literature. In this setting, we can think of the government as the "principal" and the taxpayer as the "agent" Reinganum and Wilde (1985), Mookherjee and Png (1989), Border and Sobel (1987) all present this type of model in the context of optimal taxation. In addition, Reinganum and Wilde (1988), Kaplow (1990), and Graetz, Reinganum, and Wilde (1986) are models with the possibility of evasion and costly enforcement (though not costly state verification in the sense of Townsend (1979)).

The central theme of this paper is to explore the possibility that auditing costs and costly state verification could cause tariffs to be part of an optimal tax system, and to determine if such a model would fit general features of the data described above. Because of the nature of the problem, explaining the co-existence of multiple tax instruments, this research differs significantly from much of the previous work on optimal taxation. In fact, in a survey article, Slemrod (1990) writes that the current literature on optimal taxation is lacking in that it fails to take into account the differences in the costs of operating different tax instruments. He states that a research agenda should be directed towards the study of the choice of tax instruments and the choice of enforcement policy. The model presented here is aimed precisely at this goal. Slemrod's (1990) term "optimal tax systems" is used to describe normative theories of taxation which consider the abovementioned aspects of the problem. I will use the term similarly.

The remainder of the paper is organized as follows. Section 2 explains the economic environment including the problem of costly state verification. Section 3 derives the main result concerning the optimal level of taxes and tariffs. Section 4 discusses some implications of the results. Section 5 contains some concluding remarks.

2. THE MODEL

2.1. Preliminaries

The remainder of this paper will make extensive use of a helpful analogy. Think of the economy as a circle with a given land area. Agents are uniformly distributed over this land area so that at any point in the interior there exists a single agent with zero mass. The mass of agents on the entire circle is equal to one. Any individual agent must choose between producing goods and providing a service, namely the service of bringing imports into the country. This choice is motivated by the observation that only a rather small number of agents are involved in bringing goods from the foreign country into the home country market. Presumably this has something to do with the infrastructure and geography of the home country. Imported goods would come in at the ports located along the shoreline (or border) or, in the age of air transport, at a limited number of airports.[3] It would be difficult, if not impossible, for imported goods to come directly to any arbitrary point in the interior of the country. In this spirit, the agents who are located closest to the shoreline are the ones who handle the imports as they come off of the boats. They then sell the goods to the general population.

This distinction is not found in the traditional models of trade because in those models importing is costless. When it is costless for any agent to become an importer, we would expect to see everyone handling only his or her own imports. They would not be reliant on those agents closest to the shore, and the advantage associated with being near the shore would disappear. However, since we observe that certain agents specialize in handling international transactions and since it has clear implications for the enforcement of tax/tariff policy, it will be important to distinguish between

[3]Hereafter, I will simply use the term "shoreline" as if this country were an island. The analysis is unchanged if we think of this as a landlocked country. Only the terminology is different.

producing agents and importers.[4] A primary motivation for this explanation of the role of tariffs is that the auditing of income taxes and the auditing of tariffs are fundamentally different activities. This model delivers that difference.

Specifically, I make the assumption that there are location dependent costs which preclude the possibility of every agent handling his or her own imports. Thus, the land held by agents close to the shore gives them a comparative advantage in the provision of the service of importing, and they specialize in that service. The service of providing imports is a non-traded good, and the production of the non-traded good is affected by the tax system.

The remainder of this section describes the tax system and the optimization problem of the agents involved in production of the consumption goods and the problem of the agents in the service sector (importers).

2.2. The Tax System

The import tariff is one of the instruments which the government has at its disposal to raise its operating revenue. The other instrument is an income tax. The government is required to raise a certain amount of revenue, $\overline{G}$, to provide a public good, and chooses taxes so as to maximize the expected utility of the agent. The problem of choosing taxes optimally can be cast in an environment in which there is the possibility of "cheating" or underreporting income. In that case, the government might decide to expend resources to audit taxpayers with the goal of catching those who cheat as well as discouraging cheating. This expenditure of resources to verify incomes of the agents and choosing an optimal scheme of taxation make this problem similar to Townsend's (1979) landmark paper. The tax system can be thought of as the "contract." Reinganum and Wilde (1985) build a model based on this premise

[4] These "importer" agents may be thought of as import/export firms. This model, however, largely ignores their role as exporters.

with the assumption of risk neutrality[5]. Here, I extend this basic idea to allow more than one tax instrument as well as risk averse agents. In keeping with the central theme of explaining the use of tariffs, which are less efficient in environments with costless verification and enforcement, I allow for direct taxes on individual income of producers and tariffs on imported goods.[6].

The government cannot observe the income of the producers or the incoming shipments of the importers unless they pay the cost of auditing them.[7] If this cost is paid, then the government learns the agent's true income. These costs can be thought of as constant marginal costs of auditing. The total audit cost to the government equals this constant marginal cost multiplied by the mass of agents who are audited. Slemrod and Yitzhaki (1987) model auditing costs as a function of enforcement and do not require constant marginal costs. The constant marginal cost assumption is made here to simplify the analysis. This assumption does not fundamentally alter the basic result derived below. In fact, allowing the cost of auditing income taxes to be increasing with enforcement would bias the model in favor of positive tariffs in equilibrium. The result shown in section 3 that positive tariffs are possible even with constant marginal costs of auditing is therefore a somewhat stronger conclusion.

The fact that the government cannot observe income unless they pay a cost raises the possibility of underreporting income. The next section deals with the problem of how much agents should report and how much they should produce and consume.

[5] Other models in the literature also make this assumption.

[6] Hereafter, direct taxes on individual income will simply be referred to as "income taxes."

[7] From this point on, the term "agent" will only be used when referring to both producers and importers. Otherwise, to avoid confusion, the terms "producer" and "importer" will be used.

2.3. Agent's Optimization

The sequence of actions is as follows. Any individual agent in this economy must first decide whether to produce or to provide the service (import). At that point, the producers allocate labor between the production of two consumable goods. Simultaneously, the importers bring goods from the rest of the world to sell on the world market. Since only the importers have access to overseas markets, they may charge a price higher than the world price. This price will reflect the cost of becoming an importer. Finally, the producers sell some of their goods on the domestic market and some of the goods are exported. At the end, everyone consumes all the goods that were produced (less exports) and imported.

This problem has multiple steps, and it makes sense to solve it by backward induction. The reason for using backward induction is that by the time the agents get to the stage at which they allocate labor or determine the domestic price of the imported good, they already know the prices and tax rates as well as how many importers there are. The costs of importing are sunk ahead of time during a "planning period." All agents take this as given and optimize. Working backward to the first step, the "planning period", the agents choose the "occupation" with the greater rate of return. This determines the equilibrium mass of importers. The result is an equilibrium allocation which cannot be improved upon by any agent. We now set up and solve the problem in this manner beginning with the production and consumption decision of producing agents.

2.3.1. Producing Agent's Problem.—

The economy is small and open in the sense that the world price vector is taken as given, and any import tariff imposed by the country has no effect on the world price. Two goods are produced in this economy. Call the goods x_1 and x_2. The numeraire good is x_1, and the domestic price of x_2 is P. The world price of x_2 is p. Without loss of generality, assume that in equilibrium x_2 is imported (though it is also be produced domestically.) Assume that there is an *ad valorem* tariff on the imported good, x_2, and this is reflected in the domestic price P which is higher than the world price p.

Producers earn income equal to the value of their production. Each producer owns a unit of labor, receives an individual productivity shock, and based on the realization of the shock, decides how much labor to devote to the production of good x_1 and x_2. For an individual agent indexed by j,

$$\begin{aligned} x_{1j} &= \mu_j L_{1j}^{\nu} \\ x_{2j} &= \mu_j L_{2j}^{\eta} \end{aligned} \tag{1}$$

where μ_j is the individual shock to total productivity which is independently and identically distributed across producers. This is a convenience to keep the number of states small.[8] Assume that $\mu_j = \mu_L$ with probability Θ, and $\mu_j = \mu_H$ with probability $1 - \Theta$. ($\mu_H > \mu_L$). The amount of labor devoted to the production of x_i by agent j is denoted L_{ij}. Labor is supplied inelastically from a unit endowment, so $L_{1j} + L_{2j} = 1$. The only restriction on the technology parameter is that $0 < \nu < 1$ and $0 < \eta < 1$. This specification of the technology in this economy yields a standard production possibilities frontier for each individual agent which is the boundary of a convex subset of $\Re_+^2 \times \Re_+^2$. The random shock, μ, shifts the frontier between two positions. For simplicity, the technology function impacts the production of both goods in the same proportion.

Income is allocated over the consumption of the two goods. Consumption of good 1 (x_1) is denoted c_1, and consumption of good 2 (x_2) is denoted c_2. The consumer has a utility function, $U(c_1, c_2)$. Assume $U(c_1, c_2) = u(c_1) + u(c_2)$ where $u(\cdot)$ is concave

[8] With more then two possible states, the notation as well as the analysis becomes much more complex.

and increasing. and $u(0) = 0$. Also assume that preferences are homothetic.

To keep the analysis simple, assume that there exists no mechanism by which agents can insure against the low income state. Clearly such insurance will not be supplied be the agents themselves as they are *ex ante* identical. Government provided insurance would cause moral hazard problems due to the fact that income is observable only at a cost. There would be incentive to file false insurance claims. Ruling out insurance keeps the focus on the taxation problem.

Income is denoted y and is a function of x_1 and x_2. The amount of income reported to the government to determine tax liability is y_r. The income tax rate is t. The producer/consumer's indirect utility, $V(\cdot,\cdot)$, is a function of disposable income and the tariff.

The individual producer's optimization problem in expected utility terms (with the individual j subscripts suppressed) is as follows. At this stage, the technology shock is known.

$$\max_{L_1,L_2,y_r} V((y-ty_r),\tau) = \max_{c_1,c_2,L_1,L_2,y_r} U(c_1,c_2) \tag{2}$$

$$\begin{aligned}
\textit{subject to} \quad c_1 + Pc_2 &= y - ty_r & (3)\\
x_1 &= \mu L_1^{\nu} \\
x_2 &= \mu(1-L_1)^{\eta} \\
V(y_H(1-t),\tau) &\geq \rho_{tL}V(y_H - T_t,\tau) + (1-\rho_{tL})V(y_H - ty_L,\tau) & (4)\\
V(y_L(1-t),\tau) &\geq \rho_{tH}V(y_L - T_t,\tau) + (1-\rho_{tH})V(y_L - ty_H,\tau) & (5)
\end{aligned}$$

The constraints (4) and (5) are the incentive compatibility constraints. Consider the decision of how much income to report. If the government audits someone who reported less than their full income, the government has the power to tax them T_t, where T_t is the penalty function for individual income tax underreporting. If the government does not audit, the tax is simply ty_r. Then, if ρ_{tL} is the probability of an audit of a low report and $y_H - T_t$ is the disposable income received if a high income person is audited and found to have incorrectly reported their income, constraint (4) says that telling the truth is at least as good as falsely reporting their income.

At this point, we would like to be able to call upon the Revelation Principle to assure us that we need only concentrate on mechanisms which induce truthful reporting. The Revelation Principle states that if a feasible audit/tax/penalty policy yields an optimal reporting function from the consumer, there is another feasible policy which makes the taxpayer and the government as well off while causing the taxpayer to report truthfully[9]. However in this problem, the report of the agent affects the payoff directly. So we cannot be assured that the optimal solution involves truth telling.

One way to make the Revelation Principle apply would be to set the tax rate for high and low incomes differently (progressive or regressive) so that the payoff of the agent is entirely under the control of the principal. As long as we constrain the government to move the two tax rates in the same direction when they set income tax and tariff rates the results of this paper are qualitatively unaffected. This is because we will be concerned with the average tax rate. Because the basic issue of this paper does not hinge upon the application of the Revelation Principle, I will not carry this out in this paper. However, I raise the issue that the optimal progressivity of the tax code depends on this application. I proceed under the assumption of a single tax rate and consider the optimal (truth telling) equilibrium under this constraint.

It is feasible for the government to take all of the income of a person who is caught falsely reporting. Let $T_t = y_j$ for all j. In this case, the agent gets zero utility.[10] Reports of high income are never audited because the low income consumer has no

[9] See Myerson (1979) and Harris and Townsend (1981).

[10] Recall that $u(0) = 0$. This is so that we rule out punishments of infinitely negative utility.

incentive to lie and report high income (and the government is receiving the high tax payment anyway). That is to say that constraint (5) is not binding. Thus, if we ignore ρ_{tH} and let $\rho_t = \rho_{tL}$, we have

$$V(y_H(1-t),\tau) \geq \rho_t V(0,\tau) + (1-\rho_t)V(y_H - ty_L,\tau) \tag{6}$$

With the provision that consumers are truthful when they are indifferent between telling the truth and giving a false report, consumers are always truthful, and $y_r = y$. The fact that the government is choosing the least cost policy which induces truthful reports from the consumers implies that (6) holds with equality, which gives the following equation for $\rho_{t.}$

$$\rho_t = 1 - \frac{V(y_H(1-t),\tau)}{V(y_H - ty_L,\tau)} \tag{7}$$

We can then drop the constraint (4) from the agent's problem.

Next, consider the decision of how to allocate labor effort. Leisure is not valued, so the agent's utility is maximized when income from production is maximized. Taking prices as given, the agent must allocate labor between the production of the two goods to equate the value of the marginal product.

$$\mu_j \nu L_{1j}^{\nu-1} = \mu_j P \eta L_{2j}^{\eta-1} \tag{8}$$

In principle, equation (8) can be solved for L_{1j} and L_{2j}. It is evident that the tariff distorts production. Wages are given by the value of the marginal product; however, decreasing returns to scale in the production of both goods implies that there may be nonzero profits in equilibrium. To avoid having to specify how the distribution of profits affects individual income, assume that profits are returned to the agent in a lump sum which is in proportion to that agent's individual shock. In other words, each agent is a distinct decreasing returns to scale firm. The income of any agent is equal to his or her production valued at world prices.

$$y_j = x_{1j}^* + px_{2j}^* \tag{9}$$

where the $*$ denotes equilibrium values obtained from equations (8) and (1). It is clear from the above discussion that income of the individual is random, with $y_j = y_L$ with probability Θ (when production of both goods is lower) and $y_j = y_H$ with probability $1-\Theta$. Though individual income is random, aggregate income of the production sector is constant for given prices and government policies at

$$y = Ey_j = \Theta y_L + (1-\Theta) y_H$$

If we let $\overline{\mu} = E\mu$, we can write

$$y = \overline{\mu}\left(L_1^{\nu} + p(1-L_1)^{\eta}\right) \tag{10}$$

After making the appropriate substitutions, the first order conditions imply

$$\frac{\frac{\partial u}{\partial c_2}}{\frac{\partial u}{\partial c_1}} = P \tag{11}$$

Equations (3) and (8) through (11) completely solve the producer/consumer's problem.

Let us now consider the importer's problem.

2.3.2. The Importer's Problem.—

The nature of the importers has been set out in the beginning of this section. Here I solve their optimization problem, which turns out to be fairly simple. We can think of the importers in this model as being akin to the standard neoclassical firm. Their objective is to maximize profits. They are competitive, and they are risk neutral. Because importing is a costly activity, it is only done by a subset of the total measure of agents. The measure of the importers is defined to be $0 < M < 1$. Importers unload the import bearing ships coming into the ports. They sell the imports to the consumers at the price $P = (p(1+\tau)+\pi)$, pay the tariff, $p\tau$, to the government and pay the price, p, to the foreign exporter. The difference, π, is their profit on each unit of imports. So the income of those agents is equal to π multiplied by the quantity of their imports.

In this model, the government cannot observe who is receiving the imports without auditing them. Analogous to the producer's problem, this raises the issue of underreporting or smuggling. The observation that imports typically pass through a limited number of ports puts the smuggling decision in the hands of the importers and separates the decision on whether or not to evade the tariff from the consumer's problem solved earlier. In this way, we isolate the problem of monitoring the level of imports to be taxed to specific locations, i.e. ports, border crossings, and the like. The importers could, to use the interpretation given above, falsely report the quantity of imports coming off the ship and thus pay less to the government. If they make a false report and go undetected, they would earn additional profits. To combat this, the government, just as they determine how to enforce the tax laws concerning individual consumers, determines how to enforce the tariff laws concerning the importers. Smuggling by *individual* consumers who do not specialize in importing is ruled out.

Now define the cost of being an importer, $F(M)$. Assume that this cost is increasing with the distance from the border or shoreline. That is, $\frac{\partial F(M)}{\partial M} > 0$. Under this assumption, the importers will be the agents closest to the border. See figure 2 for a graphical representation. Assume also that the range of $F(\cdot)$ is all of the positive real numbers. This is a technical assumption to ensure existence of equilibrium. At this stage of the backward induction solution process when the importers optimize, M is taken as given.

Assume that the importing agent has the same preferences for consumption as the producing agent described above. This reduces the problem to one of maximizing income. If the importer sets the "mark-up", π, too high, then the producing agents will do their own importing. Note that for the service of handling the imports, high income producers end up paying πc_{2H} to the importer. If $\pi c_{2H} > F(M)$, then that high income producers on the margin shown in figure 2 would be better off paying $F(M)$ and handling their own imports. Thus it must be the case that

$$\pi = \left(\frac{F(M)}{c_{2H}} \right) \tag{12}$$

Importers face uncertainty in the quantity of imports they will receive. For our analytical purposes, the precise quantity of imports received at each location is not important. Therefore simply specify that importers receive Ψ units of the imported good with probability $(1 - \theta)$ and $a\Psi$ units with probability θ where $0 > a > 1$. This is arbitrary up to the requirements that importers have non-negative consumption and that trade is balanced.

As in the producer's problem, we also have to consider the possibility of underreporting. An incentive compatibility constraint for the high income importer must be satisfied.

$$\pi\Psi \geq \rho_\tau \left[(p\tau + \pi)\Psi - T_\tau \right] + (1 - \rho_\tau) \left[(p\tau + \pi)\Psi \right]$$

Where T_τ is the penalty for underreporting and ρ_τ is the probability that a low report is audited. As in the producer's problem, the incentive compatibility constraint for the low income importer is not binding, so it has been omitted.

Again, the same caution about the application of the Revelation Principle applies, and we proceed in the same way as in the previous section. The government can impose a large penalty, T_τ, on importers who underreport their level of imports. The larger the penalty, the lower are the profits for importers who are caught cheating. So there will be some probability of auditing the importers, ρ_τ, which will cause the importer to report truthfully.

As in the consumer's section, we can specify that the government will confiscate their entire income when an audit reveals that an importer cheated.

$$\pi \geq (1 - \rho_\tau)(p\tau + \pi) \tag{13}$$

This equation solves for ρ_τ. Using the fact that incomes of producers and importers are equal and that the government uses the least cost enforcement policy, we have

$$\rho_\tau = \left(\frac{p\tau}{p\tau + \pi} \right) \tag{14}$$

With the provision that importers report truthfully when they are indifferent between being truthful and underreporting, all importers report their truthful shipment of imports.

This completes the solution of the importers' problem. The last step is to determine equilibrium M and to impose the equilibrium market clearing conditions.

2.3.3. Market Clearing.—

Determining equilibrium M is straightforward given the assumptions we have made. There must be no possibility of arbitrage across occupations. The expected utility of being a producer must be the same as the expected utility of being an importer.

$$\begin{aligned} & \theta V\left(\left(\pi a \Psi - F(M)\right)(1-t)\right) + (1-\theta) V\left(\left(\pi \Psi - F(M)\right)(1-t)\right) \\ = \; & \Theta V(y_L(1-t)) + (1-\Theta) V(y_H(1-t)) \end{aligned} \tag{15}$$

There is one more market clearing condition which must be written down. The balance of trade condition must hold. Let c_1 and c_2 continue to represent the amount of each good consumed by an agent in the production sector, and let c_{1M} and c_{2M} represent the amount consumed by an agent in the importing sector.

$$p\left((c_2 - x_2)(1-M) + Mc_{2M}\right) = (x_1 - c_1)(1-M) - Mc_{1M} \tag{16}$$

From this point on, it will be assumed that M is not very responsive to changes in τ. Ultimately, by equation (15), this is an assumption on $F(\cdot)$. If $\frac{\partial M}{\partial \tau}$ is sufficiently small, it can be ignored in the next section. There is a fundamental intuitive reason for this assumption. If the M is highly sensitive to the tariff rate, then a small change in the tariff might have a very large effect on the tariff base – the number of imports coming to any given location. This could bias the results either in favor or against tariffs relative to the income tax. While it is interesting that the model is able to deliver such implications, it is not the focus of this exercise. Thus, this effect will be assumed to be small throughout the rest of the paper.

This assumption together with downward sloping demand curves, also implies that $\frac{\partial \pi}{\partial \tau} > 0$ by equation (12). This result will be called upon in the next section.

This completes the specification of the model and the optimal decisions of the consumer and importer. We now turn to the government's problem of determining the optimal tax and tariff.

3. THE OPTIMAL TAX SYSTEM

3.1. Benchmark Case – Zero Cost Auditing

Here I consider the government's problem in the case of costless auditing. If there is no cost to auditing, the government will always learn the true income or the true level of imports for each individual. It is useful to start here because it establishes the expected result that the tariff will not be used through a solution method which can easily be extended to the case of costly state verification. The government takes the agents' (both producers and importers) decisions as given and maximizes social welfare (weighting producers and importers proportionally) subject to the government budget constraint which requires the provision of a public good.

As noted above, $\frac{\partial M}{\partial \tau}$ and $\frac{\partial M}{\partial t}$ will be assumed to be sufficiently close to zero so that they can be left out of the equations below. This will eliminate any bias in the results coming only from arbitrary specifications of $F(\cdot)$.

Define y_M to be the aggregate income of the importers. Again, let y_j, c_{1j} , and c_{2j} denote the individual random values of consumption and income (to distinguish these variables from the aggregates y, c_1, and c_2). Finally, to save space, let $C_2 = c_2(1-M) + Mc_{2M}$.

$$\max_{t,\tau} EV(y_j(1-t), \tau) \tag{17}$$

$$\begin{aligned} s.\ t. \quad \overline{G} &= t\left((1-M)y + My_M\right) + p\tau\left(C_2 - x_2(1-M)\right) && (18) \\ C_2 - x_2(1-M) &> 0 && (19) \\ t &\in [0,1] \\ \tau &\geq 0 \end{aligned}$$

Constraint (19) ensures that imports of good 2 remain positive.

Income tax received is equal to $t\left((1-M)y + My_M\right)$ because importers also pay income tax. I assume that importers always report truthfully and their income reports are not audited. This could be because the government could costlessly compare their income tax report to the (truthful) report of imports they have already made.[11]

At this point, note the following result which is used extensively in the rest of the paper.

Lemma 1: $\frac{dy_j}{d\tau} < 0$. Higher tariffs lead to lower income for producers in every state.

Proof: See appendix.

This result summarizes the dependence of labor allocation and output on tariff policy. Under the specifications in this model, income tax policy does not distort

[11] There is also a technical reason for this assumption. Because the income of the importers does not necessarily equal the income of the producers, there would be more than two possible states for income, making the analysis much more difficult.

production. Therefore this result factors heavily in the determination of optimal tax policy.

Similarly, since the incomes of the producers and importers are linked by equation (15), it follows immediately that $\frac{\partial((1-M)y+My_M)}{\partial\tau} < 0$.

To solve the government's problem, it is useful to know $\frac{dt}{d\tau}$. Totally differentiating the government budget constraint, (18):

$$\begin{aligned}\frac{dt}{d\tau} &= -\frac{p\left(C_2 - x_2(1-M)\right) + p\tau\left(\frac{\partial C_2}{\partial\tau} - \frac{\partial x_2}{\partial\tau}(1-M)\right)}{p\tau\left(\frac{\partial C_2}{\partial t} - \frac{\partial x_2}{\partial t}(1-M)\right) + ((1-M)y + My_M)} \\ &\quad - \frac{t\frac{\partial((1-M)y+My_M)}{\partial\tau}}{p\tau\left(\frac{\partial C_2}{\partial t} - \frac{\partial x_2}{\partial t}(1-M)\right) + ((1-M)y + My_M)}\end{aligned} \tag{20}$$

Furthermore,

$$\frac{dy_j(1-t)}{d\tau} = (1-t)\frac{dy_j}{d\tau} - y_j\frac{dt}{d\tau}$$

Observe that $\frac{dy_j(1-t)}{d\tau} < pC_2 + \frac{\partial\pi}{\partial\tau}C_2$.

$$\begin{aligned}\frac{dy_j(1-t)}{d\tau} &= (1-t)\frac{dy_j}{d\tau} - pC_2 + pC_2 - \frac{\partial\pi}{\partial\tau}C_2 + \frac{\partial\pi}{\partial\tau}C_2 \\ &\quad + y_j\frac{p\left(C_2 - x_2(1-M)\right) + p\tau\left(\frac{\partial C_2}{\partial\tau} - \frac{\partial x_2}{\partial\tau}(1-M)\right)}{p\tau\left(\frac{\partial C_2}{\partial t} - \frac{\partial x_2}{\partial t}(1-M)\right) + ((1-M)y + My_M)} \\ &\quad + y_j\frac{t\frac{\partial((1-M)y+My_M)}{\partial\tau}}{p\tau\left(\frac{\partial C_2}{\partial t} - \frac{\partial x_2}{\partial t}(1-M)\right) + ((1-M)y + My_M)}\end{aligned}$$

In expected value terms, this can be rewritten as

$$
\begin{aligned}
E\frac{dy_j(1-t)}{d\tau} &= E(1-t)\frac{dy_j}{d\tau} + pC_2 + \frac{\partial\pi}{\partial\tau}C_2 \\
&+ Ey_j\frac{p\tau C_2\left(p+\frac{\partial\pi}{\partial\tau}\right)\left(\frac{\partial C_2}{\partial y_j(1-t)}+\frac{\partial C_2}{\partial P}\right)+p\tau\frac{\partial x_2}{\partial\tau}(1-M)}{p\tau\left(\frac{\partial C_2}{\partial t}-\frac{\partial x_2}{\partial t}(1-M)\right)+((1-M)y+My_M)} \\
&+ Ey_j\frac{-\frac{\partial\pi}{\partial\tau}C_2((1-M)y+My_M)-px_2(1-M)+t\frac{\partial((1-M)y+My_M)}{\partial\tau}}{p\tau\left(\frac{\partial C_2}{\partial t}-\frac{\partial x_2}{\partial t}(1-M)\right)+((1-M)y+My_M)}
\end{aligned}
$$

Notice that the denominator is positive (because it is the revenue generated by a marginal increase in the income tax rate holding the tariff rate constant.) All the other terms (except pC_2) are negative since $\frac{dy_j}{d\tau} < 0$, $\frac{\partial x_2}{\partial\tau} > 0$, $\frac{\partial\pi}{\partial\tau} > 0$, and $C_2\frac{\partial C_2}{\partial y_j(1-t)} + \frac{\partial C_2}{\partial P} < 0$ by the Slutsky Equation.

Now substitute the government budget constraint into the objective function and maximize (17) where t is a function of τ. Under these assumptions we can differentiate inside the expected value expression. The first order condition implies

$$
\begin{aligned}
E\frac{dV}{d\tau} &= E\frac{\partial u}{\partial c_{1j}}\left(\frac{\partial c_{1j}}{\partial P}\frac{\partial P}{\partial\tau}+\frac{\partial c_{1j}}{\partial y_j(1-t)}\frac{dy_j(1-t)}{d\tau}\right) \\
&+ E\frac{\partial u}{\partial c_{2j}}\left(\frac{\partial c_{2j}}{\partial P}\frac{\partial P}{\partial\tau}+\frac{\partial c_{2j}}{\partial y_j(1-t)}\frac{dy_j(1-t)}{d\tau}\right) \qquad (21)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
E\frac{dV}{d\tau} &= E\left(p+\frac{\partial\pi}{\partial\tau}\right)\frac{\partial u}{\partial c_{1j}}\left[\left(\frac{\partial c_{1j}}{\partial P}+\frac{\partial c_1}{\partial y_j(1-t)}c_{2j}\right)+P\left(\frac{\partial c_{2j}}{\partial P}+\frac{\partial c_{2j}}{\partial y_j(1-t)}c_{2j}\right)\right] \\
&+ E\left[\frac{\partial u}{\partial c_{1j}}\frac{\partial c_{1j}}{\partial y_j(1-t)}+\frac{\partial u}{\partial c_{2j}}\frac{\partial c_{2j}}{\partial y_j(1-t)}\right]\left(\frac{dy_j(1-t)}{d\tau}-pC_2-\frac{\partial\pi}{\partial\tau}C_2\right) \qquad (22)
\end{aligned}
$$

The first term in the brackets is zero (Hicks (1939)). The second term is negative since the indirect utility function is nondecreasing in income and since $E\frac{dy_j(1-t)}{d\tau} - pC_2 - \frac{\partial\pi}{\partial\tau}C_2 < 0$. Since $E\frac{dV}{d\tau} < 0$ for all nonnegative tariffs, the solution is to set the tariff as small as possible, i.e. zero. This establishes the traditional result that when the government can costlessly observe income and imports the optimal tariff is zero. This solution method generalizes to the costly auditing case.

3.2. Costly State Verification

Recall that ρ_t is the probability of auditing income taxes and ρ_τ is the probability of auditing tariff reports when the reports are low. Since agents are truthful, audits result only in the cost of the audit being lost to society. The cost of auditing one producer is denoted γ_t, and the cost of auditing one importer is γ_τ. Thus, when auditing is costly, equation (18) becomes

$$\overline{G} = t\left((1-M)y + My_M\right) + p\tau\left(C_2 - x_2(1-M)\right) - (1-M)\Theta\rho_t\gamma_t - M\theta\rho_\tau\gamma_\tau \quad (23)$$

The additional terms are the losses resulting from the costly audits. For the respective tax instruments the costs consist of the marginal cost, γ, multiplied by the fraction of low reports audited, ρ, multiplied by the measure of low reports, Θ and θ. Assume $\gamma_\tau < \gamma_t$, so that income tax returns of importers are never audited. The government knows the income of the importers by auditing imports, not income tax returns.

First we find $\frac{dt}{d\tau}$.

$$\frac{dt}{d\tau} = -\frac{p\left(C_2 - x_2(1-M)\right) + p\tau\left(\frac{\partial C_2}{\partial \tau} - \frac{\partial x_2}{\partial \tau}(1-M)\right)}{D - (1-M)\Theta\gamma_t\frac{\partial \rho_t}{\partial t} - M\theta\gamma_\tau\frac{\partial \rho_\tau}{\partial t}} - \frac{t\frac{\partial((1-M)y+My_M)}{\partial \tau} - (1-M)\Theta\gamma_t\frac{\partial \rho_t}{\partial \tau} - M\theta\gamma_\tau\frac{\partial \rho_\tau}{\partial \tau}}{D - (1-M)\Theta\gamma_t\frac{\partial \rho_t}{\partial t} - M\theta\gamma_\tau\frac{\partial \rho_\tau}{\partial t}}$$

Where D is the denominator in equation (20).

Thus, following the same procedure as with zero cost auditing, we can write the following expression which contains the elements of the main result.

$$E\frac{dy_j(1-t)}{d\tau} = E(1-t)\frac{dy_j}{d\tau}+pC_2+\frac{\partial\pi}{\partial\tau}C_2 \tag{24}$$
$$+Ey_j\frac{p\tau C_2\left(p+\frac{\partial\pi}{\partial\tau}\right)\left(\frac{\partial C_2}{\partial y_j(1-t)}+\frac{\partial C_2}{\partial P}\right)+p\tau\frac{\partial x_2}{\partial\tau}(1-M)}{D-(1-M)\Theta\gamma_t\frac{\partial\rho_t}{\partial t}-M\theta\gamma_\tau\frac{\partial\rho_\tau}{\partial t}}$$
$$+Ey_j\frac{-\frac{\partial\pi}{\partial\tau}C_2\left((1-M)y+My_M\right)-px_2(1-M)+t\frac{\partial((1-M)y+My_M)}{\partial\tau}}{D-(1-M)\Theta\gamma_t\frac{\partial\rho_t}{\partial t}-M\theta\gamma_\tau\frac{\partial\rho_\tau}{\partial t}}$$
$$+Ey_j\frac{(1-M)\Theta\gamma_t\left(\frac{\partial\rho_t}{\partial t}\left(p+\frac{\partial\pi}{\partial\tau}\right)C_2-\left((1-M)y+My_M\right)\frac{\partial\rho_t}{\partial\tau}\right)}{D-(1-M)\Theta\gamma_t\frac{\partial\rho_t}{\partial t}-M\theta\gamma_\tau\frac{\partial\rho_\tau}{\partial t}}$$
$$+Ey_j\frac{M\theta\gamma_\tau\left(\frac{\partial\rho_\tau}{\partial t}\left(p+\frac{\partial\pi}{\partial\tau}\right)C_2-\left((1-M)y+My_M\right)\frac{\partial\rho_\tau}{\partial\tau}\right)}{D-(1-M)\Theta\gamma_t\frac{\partial\rho_t}{\partial t}-M\theta\gamma_\tau\frac{\partial\rho_\tau}{\partial t}}$$

If this is always negative for $\tau \geq 0$, then the optimal tariff is again zero. Notice again that the denominators of the RHS is positive since it is the revenue gained by a marginal increase in the income tax rate. Many of the terms in the numerator are the same as in the costless verification case. However, there now may be a positive term.

The following lemmas are useful.

Lemma 2: $\frac{\partial\rho_t}{\partial t} > 0$. The probability of auditing low income reports is increasing in the tax rate.

Proof: See appendix.

Lemma 2 essentially is a result of risk aversion. Higher income tax rates make cheating more attractive. Cheating is a gamble, and the possible gain increases as the income tax rate increases. Thus, the probability of being caught must rise in order to keep the expected value of the gamble equal to the value of telling the truth.

Lemma 3: If $\frac{\partial y}{\partial P}$ is sufficiently small in absolute value, then $\left(\frac{\partial\rho_t}{\partial t}\left(p+\frac{\partial\pi}{\partial\tau}\right)C_2-\left((1-M)y+My_M\right)\frac{\partial\rho_t}{\partial\tau}\right)$ 0 for any γ_t.

Proof: See appendix

Lemma 3 is a technical lemma needed in the following proposition.

Proposition 1: If $\left(\frac{\partial\rho_t}{\partial t}\left(p+\frac{\partial\pi}{\partial\tau}\right)C_2 - ((1-M)y + My_M)\frac{\partial\rho_t}{\partial\tau}\right) > 0$ (see Lemma 3), there exist γ_t sufficiently large and γ_τ sufficiently small so that $\tau^* \neq 0$. (Positive optimal tariff)

Proof: Consider equation (24) when $\tau = 0$. If $\left(\frac{\partial\rho_t}{\partial t}\left(p+\frac{\partial\pi}{\partial\tau}\right)C_2 - ((1-M)y + My_M)\frac{\partial\rho_t}{\partial\tau}\right) > 0$, then $\left(E\frac{dy_j(1-t)}{d\tau} - pC_2 - \frac{\partial\pi}{\partial\tau}C_2\right)$ can be made as large as desired by letting γ_t be large. This is because all of the other terms in equation (24) are bounded by the production technology and by letting γ_τ be arbitrarily small.

Now refer to equation (22) and notice that by making $\left(E\frac{dy_j(1-t)}{d\tau} - pC_2 - \frac{\partial\pi}{\partial\tau}C_2\right)$ large, it is possible that $\frac{\partial V}{\partial\tau} > 0$. Given the continuity of $\left(E\frac{dy_j(1-t)}{d\tau} - pC_2 - \frac{\partial\pi}{\partial\tau}C_2\right)$, welfare improves with a positive tariff. Q.E.D.

The intuition is that if the cost of auditing income tax reports is high, then using a tariff to raise some revenue will decrease the social losses associated with the audits enough to offset the price distortions associated with the tariff. The proposition establishes the result that the tariff could be a part of the optimal tax system.

The following implications follow directly from Proposition 1 if $\left(\frac{\partial\rho_t}{\partial t}pC_2 - ((1-M)y + My_M)\frac{\partial\rho_t}{\partial\tau}\right) > 0$ and rely essentially on the continuity of equation (24). The proof of each implication is brief.

Implication 1: As the volatility of income decreases, income taxes are more preferred.

Proof: The only place $(y_H - y_L)$ enters the problem is in ρ_t. It is clear that $\frac{\partial\rho_t}{\partial(y_H - y_L)} < 0$, so the expected cost of income tax audits decreases while the expected cost of audits of importers remains the same. Therefore, *ceteris paribus*, tariff revenue can be replaced by direct income tax revenue. Q.E.D.

Implication 2: As income increases beyond a certain level, y^*, the optimal tariff is zero.

Proof: Consider equation (24). The positive term is not increasing in y. The negative terms are decreasing in y. As y increases, $\left(\frac{\partial\rho_t}{\partial t}pC_2 - ((1-M)y + My_M)\frac{\partial\rho_t}{\partial\tau}\right)$

eventually becomes negative for all $\tau > 0$. Then the optimal tariff is zero. Q.E.D.

Implication 3: As $\gamma_t \to 0$, there is a value, γ_t^*, below which the optimal tariff is zero.

Proof: Similar to Implication 2. As $\gamma_t \to 0$, the only positive term in (24) decreases. Eventually, $\left(E\frac{dy_j(1-t)}{d\tau} - pC_2 - \frac{\partial \pi}{\partial \tau}C_2\right)$ becomes negative for all $\tau > 0$. Then the optimal tariff is zero. Q.E.D.

Implications 2 and 3 show that, given the continuity of the optimal tax functions, there is a downward trend in tariffs as income increases or as the cost of collecting the income tax decreases.

4. IMPLICATIONS

4.1. Relevance to the Data

The results presented in the previous section suggest that this model is capable of explaining the data mentioned in the introduction. It is not unreasonable to suppose that in the early stages of the development of the United States (late 18th and early 19th century), the cost of auditing an income tax return (γ_t in the model) would have been very large[12]. Over time, as the frontiers of the Midwest and West became more settled, the cost of auditing (especially in the frontier areas), would have become less prohibitive. Today, there is an extensive network of Internal Revenue Service offices around the United States. Therefore, the cost of auditing a single individual, regardless of location, should be relatively very low compared to the cost in colonial times. Also, the fact that income has increased supports the notion that higher incomes raise the marginal benefit of a proportional income tax and make the income tax relatively more attractive.

[12] It is unlikely that we will ever have data on what audit costs at that early date would have been since there was no system of collecting income taxes at that time.

These results are also relevant to World Development Report (1992) data on a cross section of countries. Countries such as Nepal, which collected 31% of its government revenue from taxes on international trade and transactions and only 10.8% from income, profit, and capital gains taxes, may not have the technology to collect and audit income taxes as efficiently as OECD countries. However, in 1972, the figures for Nepal were 36.7% and 4.1% for trade taxes and income taxes respectively. Other low income countries, such as Ghana and Indonesia, have similar experiences, while some, like Liberia, Sierra Leone, and Uganda, show the percentages moving in the opposite directions. It is unlikely that auditing costs would fluctuate that much over time, so perhaps other factors were influencing the tax authorities.[13] It would be interesting to gather data on these countries to test this model. Since low income countries show more variation in the percentage of revenue generated by tariffs, it would make sense to test to see if, for example, income could be low enough at times that it seriously affects the revenue that can be collected from a direct income tax. If so, this model predicts that optimal tariff rates could depend on income.

The result that auditing costs are important can be reconciled with the findings of Riezman and Slemrod (1987). Literacy rates, which they find to be a significant factor in level of tariffs in a cross section of countries, may have implications for auditing costs of income taxes. Clearly, filling out an income tax form requires a level of literacy. If people are unable to file tax returns correctly because if illiteracy, the cost of auditing a single return may be higher. Perhaps it will take more labor hours per return to correct all of the errors made by illiterate taxpayers. It is reasonable to suggest, as Riezman and Slemrod do, that this is less of a problem for trade taxes. However, this model only applies to systems in which payments are made voluntarily (in the sense that agents give a report of their income level.) If it is truly that much more costly to correct the errors made by illiterate taxpayers, perhaps other types of

[13] Such other factors could cause us to broaden the notion of auditing costs.

taxes (or more coercive methods of collection) would be used.

Finally, the model presented here has the potential to be tested with additional data from a cross section of countries. It would be ideal if data on auditing costs and related aspects of tax enforcement were available. However, as Reinganum and Wilde (1988) point out, the tax authorities may have reasons to keep data concerning tax enforcement out of the hands of the public in order to maintain compliance with the tax laws. Even if data on the cost of audits is not available directly, perhaps such a measure could be approximated from data on the total spending by the tax authorities and other government figures. Also, the model predicts that enforcement (probability of an audit) of income taxes should be higher when the tax rate is higher. This would be an interesting test of the costly state verification environment which could have implications beyond this particular problem.

4.2. Dynamic Implications

Although this is a static model, there are some interesting possibilities for dynamics in extensions of this model. Suppose there is a technology which would allow the government to pay a fixed cost in order to have lower audit costs for one of the tax instruments from that day forward. Retaining all of the assumptions of the previous sections, the government's decision rests upon the sizes the costs and benefits. The cost is the decrease in today's utility from raising taxes to pay for the fixed cost. The benefit is the discounted sum of utility from lower future audit costs (which translates to lower future tax rates). If the cost is small, we should expect that they would undertake the project and society would benefit from lower taxes. In any case, of course, income would have to be sufficiently high to bear the cost.

4.3. Theory of Optimal Tax Systems

This model has clear implications for the theory of optimal tax systems in the sense of Slemrod's (1990) explanation of the problem. Here we have a model in which two different tax instruments can co-exist or where only one of them may exist. The composition of taxes depends not only on the preferences of the agents and the production technology, but on the tax collection technology. Certainly it may be the case that certain preferences could cause the disutility of the tariff to be so great that it could outweigh even very large differences in auditing costs, and thus never be used. Depending on the parameters, it could also be the case that the taxes could co-exist, but at very different levels with the smaller tax audited less frequently.

A number of possibilities exist for extending the model on this front. There are many different taxes each with its own set of problems concerning compliance. Sales taxes and even inflation taxes can be studied in this framework. This model represents an important step in the research program suggested by Slemrod (1990) providing an intuitive, rich, and flexible environment to think about optimal tax systems.

5. CONCLUDING REMARKS

A model of costly state verification was set out in order to gauge the effects of auditing costs on the existence of multiple tax instruments. The model presented here is very general in that it allows for production inefficiencies resulting from tariffs. Even when considering the production inefficiencies in addition to the obvious attention to price distortions, the model predicts that a less efficient tax instrument may be a part of an optimal tax system if there exist high costs of auditing the more efficient tax.

The results proved here are consistent with the overall downward trend in tariffs as a percentage of government revenue in the past 200 years in the United States. They are also consistent with several branches of the literature which argue that collection costs are important in determining optimal tax and tariff rates. The main results of this paper can be a guide to future tests of hypotheses related to collection costs. In particular, it would be interesting to know if, for reasonable preference and technology parameters, low income countries behave as the model would predict – with the optimal tax and tariff functions depending on income. Despite whatever difficulties there may be in collecting data on collection costs, this is an interesting challenge. Perhaps using a proxy for collection costs, as Riezman and Slemrod (1987) did, would be the best option for future research in this area.

Another extension relates this model to the challenge presented by Slemrod (1990). It is not difficult to think of this model applying to more than simply the taxes assumed here. This is a step in the direction of a very broad view of optimal taxation and optimal tax systems. Given the many tax instruments which have been implemented, this could be an important area of research. It is appropriate to consider not only the preferences and production technology of the economy when setting tax rates, but the technology of tax collection should be considered. Just as this model shows how income taxes and tariffs can co-exist at a moment in time, a generalization of the model could predict what types of taxes would be chosen out of a menu of tax instruments over time.

Finally, these results do show that in the presence of auditing costs, which decrease the efficiency of otherwise more efficient taxes, a tariff may in fact be optimal in a small, open economy. The basic warning of elementary trade theory cited in the introduction still applies to the world of costless auditing, but may not apply when auditing is costly. In addition, the results also show that positive tariffs are more likely to be optimal in countries where income is small. Thus, the small, open economies, whose tariffs do not affect world prices and whose tariffs are ordinarily thought to be suboptimal, may in fact be the economies for which these results have greater impact.

APPENDIX

Proof of Lemma 1

Lemma 1: $\frac{dy}{d\tau} \leq 0$. Higher tariffs lead to lower aggregate income for producers.

Proof: Suppressing the individual subscript to indicate aggregate values, equation (8) implies

$$\nu L_1^{\frac{\nu-1}{\eta-1}} - ((p(1+\tau)+\pi)\eta)^{\frac{1}{\eta-1}}(1-L_1) = 0 \tag{25}$$

Totally differentiate equation (25) to obtain

$$\frac{dL_1}{d\tau} = \left(\frac{(1-L_1)\left(\frac{\eta}{\eta-1}((p(1+\tau)+\pi)\eta)^{\left[\frac{1}{\eta-1}-1\right]}\right)\left(p+\frac{\partial\pi}{\partial\tau}\right)}{\nu\frac{\nu-1}{\eta-1}L_1^{\left(\frac{\nu-1}{\eta-1}-1\right)} + ((p(1+\tau)+\pi)\eta)^{\left[\frac{1}{\eta-1}\right]}} \right) \tag{26}$$

The fact that $0 < \eta < 1$, $0 < \nu < 1$, and $\frac{\partial\pi}{\partial\tau} > 0$ implies that $\frac{dL_1}{d\tau} < 0$. Now recall equation (10) and differentiate.

$$\frac{dy}{d\tau} = \overline{\mu}\left(\nu L_1^{\nu-1} - p\eta(1-L_1)^{\eta-1}\right)\frac{dL_1}{d\tau}$$

Equation (8) implies that $\left(\nu L_1^{\nu-1} - p\eta(1-L_1)^{\eta-1}\right) > 0$. Therefore,

$$\frac{dy}{d\tau} < 0 \tag{27}$$

QED.

Proof of Lemma 2

Lemma 2: $\frac{\partial\rho_t}{\partial t} > 0$. The probability of auditing low income reports is increasing in the tax rate.

Proof: $\rho_t = 1 - \frac{V(y_H(1-t),\tau)}{V(y_H - ty_L,\tau)}$. Raising the income tax rate decreases disposable income, but does not affect production. Therefore,

$$\frac{\partial \rho_t}{\partial t}[V(y_H - ty_L), \tau)]^2 = [\frac{\partial V(y_H(1-t), \tau)}{\partial y_H(1-t)}(y_H)(V(y_H - ty_L), \tau) - \frac{\partial V((y_H - ty_L), \tau)}{\partial (y_H - ty_L)}(y_L)(V(y_H(1-t), \tau))] \quad (28)$$

Since V is concave given the concavity of u, $\frac{\partial V(y_H(1-t),\tau)}{\partial y_H(1-t)} > \frac{\partial V((y_H - ty_L),\tau)}{\partial (y_H - ty_L)}$. The hypothesis follows immediately. Q.E.D.

Proof of Lemma 3

Lemma 3: If $\frac{\partial y}{\partial P}$ is sufficiently small in absolute value, $\frac{\partial \rho_t}{\partial t} pc_2 - y\frac{\partial \rho_t}{\partial \tau} > 0$ for any γ_t.

Proof: Observe that

$$\frac{\partial \rho_t}{\partial \tau}[V(y_H - ty_L), \tau)]^2 = \left(p + \frac{\partial \pi}{\partial \tau}\right) \left(\begin{array}{c} \left(\frac{\partial V(y_H - ty_L,\tau)}{\partial P} + \frac{\partial V(y_H - ty_L,\tau)}{\partial (y_H - ty_L)} \frac{\partial (y_H - ty_L)}{\partial P}\right) V(y_H(1-t), \tau) \\ - \left(\frac{\partial V(y_H(1-t),\tau)}{\partial P} + \frac{\partial V(y_H(1-t),\tau)}{\partial y_H(1-t)} \frac{\partial y_H(1-t)}{\partial P}\right) V(y_H - ty_L, \tau) \end{array} \right) \quad (29)$$

Using Lemma 2 and equation (28), it follows that $\frac{\partial \rho_t}{\partial t} pc_2 - y\frac{\partial \rho_t}{\partial \tau} > 0$ whenever

$$\begin{array}{rl} & pc_2 \left[\frac{\partial V(y_H(1-t), \tau)}{\partial y_H(1-t)}(y_H)(V(y_H - ty_L), \tau) - \frac{\partial V((y_H - ty_L), \tau)}{\partial (y_H - ty_L)}(y_L)(V(y_H(1-t), \tau))\right] \\ > & y\left(p + \frac{\partial \pi}{\partial \tau}\right) \left[\begin{array}{c} \left(\frac{\partial V(y_H - ty_L,\tau)}{\partial P} + \frac{\partial V(y_H - ty_L,\tau)}{\partial (y_H - ty_L)} \frac{\partial (y_H - ty_L)}{\partial P}\right) V(y_H(1-t), \tau) \\ - \left(\frac{\partial V(y_H(1-t),\tau)}{\partial P} + \frac{\partial V(y_H(1-t),\tau)}{\partial y_H(1-t)} \frac{\partial y_H(1-t)}{\partial P}\right) V(y_H - ty_L, \tau) \end{array} \right] \end{array}$$

If $\frac{\partial \pi}{\partial \tau} = 0$, this simplifies to

$$\begin{array}{rl} & \left[\frac{\partial V(y_H(1-t), \tau)}{\partial P} + \frac{\partial V(y_H(1-t), \tau)}{\partial y_H(1-t)} \left(\frac{\partial y_H(1-t)}{\partial P} + y_H c_2\right)\right] V(y_H - ty_L, \tau) \\ > & \left[\frac{\partial V(y_H - ty_L, \tau)}{\partial P} + \frac{\partial V(y_H - ty_L, \tau)}{\partial (y_H - ty_L)} \left(\frac{\partial (y_H - ty_L)}{\partial P} + y_L c_2\right)\right] V(y_H(1-t), \tau) \end{array}$$

The terms $\frac{\partial V(y_H(1-t),\tau)}{\partial P}$ and $\frac{\partial V(y_H - ty_L,\tau)}{\partial P}$ are equal by the homotheticity assumption. This, together with concavity of V, implies that this inequality will hold if $\frac{\partial y}{\partial P}$ is sufficiently small in absolute value, (this is not dependent on γ_t.) If $\frac{\partial \pi}{\partial \tau} < 0$ as we have assumed, the inequality still holds. Thus, the inequality will hold if the tariff will alter production by a sufficiently small amount. Q.E.D.