

**PUTTING FIRMS INTO OPTIMAL TAX THEORY**

**Wojciech Kopczuk  
Columbia University  
1022 International Affairs Building, MC 3308  
420 West 118<sup>th</sup> Street  
New York, NY 10027  
(212) 854-2519  
fax: (212) 854-8059  
wkopczuk@nber.org**

**Joel Slemrod\*  
University of Michigan  
701 Tappan Street  
Ann Arbor, MI 48109-1234  
(734) 936-3914  
fax: (734) 763-4032  
jslemrod@umich.edu**

**Session Title: New Evidence about the Impact of Taxing Corporate-Source Income”**

**Conference draft. Not for quotation. Comments welcome.**

**January 10, 2006**

**\*Corresponding Author**

## PUTTING FIRMS INTO OPTIMAL TAX THEORY

Wojciech Kopczuk and Joel Slemrod\*

Firms are, for the most part, absent from the modern theory of optimal taxation. Their disappearance dates from the foundational models developed by Peter A. Diamond and James A. Mirrlees (1971) in which firms are simply mechanical vehicles for combining productive inputs into output in cost-minimizing proportions.<sup>i</sup>

In contrast, firms play a central role in all modern tax systems, mostly for a reason eloquently stated by Richard M. Bird (1996): “The key to effective taxation is information, and the key to information in the modern economy is the corporation.” In most countries, firms *remit* the majority of tax revenues to the government, either with regard to taxes legally owed by businesses or through withholding of taxes legally owed by employees or other businesses.<sup>ii</sup> Even when businesses are not required to remit taxes, they are often required to file information reports that can facilitate monitoring of tax liabilities.

The lack of a theoretical framework that features firms handcuffs rigorous welfare analysis of a number of important policy issues. One such example is the comparative evaluation of a uniform retail sales tax (RST) versus a value-added tax (VAT). In the standard model, these two taxes—both remitted entirely by businesses—are equivalent consumption taxes, but most experts consider the VAT to be clearly superior on administrative grounds, a view echoed in the recent report of the President’s Advisory Panel on Federal Tax Reform.

In this paper we propose a simple framework in which these and other issues can be analyzed. The new framework recognizes that a tax code must be backed up by an administrative and enforcement structure.

### I. Model Structure

Although there are tax-specific aspects of enforcement, for all taxes the difficulty of enforcement depends on two features, both of which are related to why firms play a central role in tax systems: (1) arms-length transactions, and (2) economies of scale. Transactions between unrelated, “arms-length” parties, and information reports of these transactions, greatly facilitate enforcement. In contrast, the profits of firms flow from a firm to its owners and therefore are not subject to arms-length information reports, and so audits must suffice.<sup>iii</sup> One consequence of this is that the rate of noncompliance for taxes on the profits of self-employed and other businesses is much higher than the rate for wages and salaries.<sup>iv</sup>

The model we propose simply captures the essential elements of our story. In the one-period model economy there are just two goods, where the technology of producing good 1 is given by  $X_1=L_1$  and the technology of producing good 2 is given by  $X_2=g(L_2, D)$ , where  $L_i$  represents the labor used in sector  $i$  and  $D$  is an intermediate input produced using the first technology. Thus, the output of technology 1 can be used as both the input in production of good 2 and as consumption good  $C_1$ , so that  $X_1=C_1+D$  and  $X_2=C_2$ . We assume that markets for all inputs and outputs are competitive, that both production functions exhibit constant returns to scale, and that a firm chooses to employ one technology or the other. (Vertical integration is addressed below.) We denote by  $p$  the relative price of  $C_2$ ,  $w$  the pre-tax return to labor time, and  $\pi$  the pre-tax profits. The economic profits are zero, but profits reported for tax purposes may be positive or negative, depending on tax incentives.<sup>v</sup>

The government needs to collect a fixed amount of revenue and can levy one or more of a set of linear taxes, but not lump-sum taxes. In particular, we model retail sales taxes (RST), value-added taxes (VAT), taxes on wages, and taxes on profits. To simplify the possible interactions among the taxes, we assume that a jurisdiction levies either a RST, a VAT, or some combination of wages and profits taxes; because wage taxes are withheld and remitted by firms,

all tax revenue is remitted to the government by businesses. We restrict attention to the class of utility functions for which the relative price of goods 1 and 2 at the optimum is undistorted.<sup>vi</sup> Ignoring administrative considerations, in this model this could be implemented equivalently either by a uniform retail sales tax, a uniform value-added tax, or a uniform tax on wages and profits (that are equal to zero).

Firms make decisions whether to comply with a particular kind of a tax. Noncompliance results in costs that we express in reduced form as  $A(x)$  or  $B(x)$ , where  $x$  is enforcement expenditure set by the government. In some cases, non-compliance requires cooperation of two arms-length parties; the cost of this kind of noncompliance is represented by  $A(x)$ . In other cases, non-compliance is a decision of a single agent; the cost of such noncompliance is denoted by  $B(x)$ . The costs of non-compliance may take many forms, and involve things such as the cost of restructuring transactions, hiring lawyers, investing in concealing activity from tax authorities, and foregoing opportunities that would increase the likelihood of detection or direct penalties. These costs are controllable by the government through the monetary expenditure  $x$ . We assume that it is much less costly to enforce taxes if the government can receive information reports from both sides of arms-length transactions, which applies to business-to-business transactions of a VAT and to wage payments, but not to profits or (by assumption) to sales from businesses to consumers. Formally, we assume that  $A(x) > B(x)$  and  $A'(x) > B'(x) > 0$ —the same administrative investment results in higher overall and marginal private costs of non-compliance in the presence of monitorable arms-length transactions.

A taxpayer either complies completely or does not comply at all. Given the potential tax liability of  $T$  applying in an arms-length context, the taxpayer involved will comply when  $T \leq A(x)$  and not otherwise. Similarly, in the absence of arms-length transactions, compliance will be guaranteed when  $T \leq B(x)$ . We consider policies that enforce taxes fully (and at minimal

cost).<sup>vii</sup> Therefore, the cost of compliance in the case of an arms-length relationship between parties is given by  $a(x)=A^{-1}(T)$ , and in the case of non-arms length transactions it is  $b(x)=B^{-1}(T)$ . Under full compliance, taxpayers effectively choose not to bear any private cost, which requires that the administrative expenses be present as means of deterrence. The total administrative costs are proportional to the number of effective entities that need to be monitored, either firms or, in the case of a wage tax, firms and employees. Holding the number of taxpayers constant and ignoring shifting across the tax bases, the administrative costs for each of the four taxes we consider can be written as follows:

$$\text{RST: } \varphi_R + b(t_R C_1)N_1 + b(t_R p C_2)N_2$$

$$\text{VAT: } \varphi_V + a_F(0)N_1 + b(t_V C_1)N_1 + b(t_V (p C_2 - D))N_2$$

$$\text{Wage: } \varphi_W + a_L(t_W w L)(N_1 + N_2 + m N_L)$$

$$\text{Profit: } \varphi_\pi + b(t_\pi \pi)(N_1 + N_2)$$

In these expressions  $N_i$  is the number of firms in sector  $i$ , and  $N_L$  is the total number of employees. The  $\varphi_i$  symbols refer to the fixed cost of having a given tax type at any rate.

According to this model, both wage payments and business-to-business sales generate monitorable arms-length transactions. Although the RST is based on arms-length sales, we assume that there is no practical way to involve the consumers in providing matchable information. The VAT has the same problem with matchable information reports for retail sales, but the tax base at risk is  $C_1 + p C_2 - D$  (compared to  $C_1 + p C_2$  for the RST) as long as the government does not provide refunds to retail firms. With respect to business-to-business transactions ( $D$ ), where sales taxable to the seller are deductible inputs for the buyer, there is no tax saving under an invoice-credit system to not paying tax on a good sold to another business; this explains why zero is the argument of the  $a_F$  function. We ignore the possibilities that firms

fabricate invoices to render invisible the nonpayment of VAT, and that under a RST final consumers have an incentive to falsely claim to be (tax-exempt) business purchasers.

The second factor that affects the cost of raising revenue—and that differentiates across taxes—is the number of effective tax units that need to be monitored. We use the fixed factor  $m$  ( $m < 1$ ) to translate the costs of dealing with an employee under a wage tax to that of dealing with a business, and use the  $a_F$  and  $a_L$  functions to differentiate the technologies required in monitoring business-to-business transactions and business-to-employee transactions, respectively.<sup>viii,ix</sup>

Systems of raising revenue that are equivalent under the standard setup are no longer such in this context. For example, while the proportional RST, VAT, and wage-and-profits taxes are equivalent in a static context, their administrative costs are different. In particular, this simple model would imply that a VAT is less costly than a RST because the stakes in the retail sector are lower. It also highlights what factors are critical to evaluating the desirability of VAT over wage-and-profits taxation. While the tax base is the same and costly audits are used in either case, the tax stakes supported by audits are lower under profit taxation if profits are low. On the other hand, wage taxes, while easy to enforce due to their reliance on arms-length transactions, involve many more parties. The key factors are therefore the relative cost of audits versus arms-length monitoring, and the relative number of parties involved.

To this point, we have not modeled behavioral responses that involve changing the number of (firm) taxpayers. But now consider that under a VAT firms will organize to maximize the difficulty of enforcement at the retail level. This incentive would be present for the last two firms in a production chain acting jointly, for example by vertically integrating. To illustrate the possible consequences in the context of our simple model, assume that  $N_1 = N_2 = N$

and that one integrated firm results from merging one firm from each sector, complete vertical integration under a VAT would result in administrative costs of

$$\text{VAT: } \phi_V + b(t_v C_1 + t_v p C_2)N,$$

where  $N$  is the number of resulting firms. Comparing this expression to the earlier expression for the VAT reveals two differences. First, the overall tax that has to be enforced using non-arms-length technology increases from  $t_v(C_1 + (pC_2 - D))$  to  $t_v(C_1 + pC_2)$ . Second, enforcement is now concentrated on a smaller number of businesses, with a higher tax stake per business. The first change represents an unambiguous loss to tax authorities (and therefore a social loss): part of tax revenue that was effectively self-enforced is no longer that. As for the second effect, given our assumptions and assuming additionally convexity of  $b(\cdot)$ ,  $b(t_v C_1) + b(t_v(pC_2 - D)) < b(t_v C_1) + b(t_v p C_2) < b(t_v C_1 + t_v p C_2)$ , and therefore it is more expensive to enforce taxes remitted by a vertically-integrated firm. This is, of course, an assumption, but one with interesting economic content. The vertically-integrated industry under either the VAT or RST is equally costly from the administrative point of view. However, because the tax incentives to vertically integrate are stronger under the VAT than under the RST due to higher potential tax savings, an extension of the model that would relax vertical integration as an automatic tax winner could produce situations where VAT results in vertical integration but RST does not, making the VAT a worse administrative option.

This reasoning implies that it is possible for production inefficiency to be part of an optimal solution, in contrast to the result in Diamond and Mirrlees (1971). This can occur under a VAT because of the lower cost of collecting revenue from business-to-business sales versus consumer sales. Indeed, any production benefits from vertical integration (not modeled here) may be offset by the negative administrative consequences of such integration due to the reduction of monitorable business-to-business transactions. This reasoning suggests that the

equilibrium boundary of the firm may not be the socially optimal boundary, because what is efficient from a private perspective may complicate tax collection.

This reasoning can be usefully restated using the terminology of the marginal efficiency cost of funds (MECF), as developed in Joel Slemrod and Shlomo Yitzhaki (1996). At an optimum each tax system instrument used must have an identical MECF. But the term “instrument” needs to be defined more precisely than, say, “a tax of rate  $t$  on commodity 1” and, for example, needs to differentiate a RST from a VAT. If a RST is more difficult to enforce than a VAT, it will have a higher MECF, because ignoring compliance costs the MECF equals  $X/(MR(1-H))$ , where  $X$  is the marginal revenue absent any behavioral response,  $MR$  is actual marginal revenue with behavioral response, and  $H$  is the marginal administrative cost, here (as in our model) assumed to be proportional to revenue. If, *ceteris paribus*,  $H$  is higher for a RST compared to a VAT, then its MECF is higher.

The revenue consequences of the incentive to vertically integrate will be taken account of in the calculation of  $MR$  for a VAT. In the Diamond-Mirrlees framework, if all commodity taxes are available, the MECF of a tax that disturbs production efficiency will always be higher than the common MECF of each optimally-set commodity tax. But in a world with administrative costs, this need not be the case. For example, a production-inefficient tax on large retail enterprises under a VAT might have a low MECF because it has a low value of  $H$ —it saves on administrative costs relative to other taxes. The existence of this margin of response increases the MECF of the VAT, either because it increases the marginal “leakage” of revenue (i.e.,  $MR=X-ML$ , where  $ML$  is marginal leakage) or, in our framework, it increases  $H$ . Taking the appropriate policy action to address this margin of response ensures that the VAT generates its optimal MECF.

Our argument suggests that optimal policy considering administrative cost externalities may involve instruments that violate the standard notion of production efficiency.<sup>x</sup> A related point has been made in the context of environmental externalities by Lans Bovenberg and Lawrence Goulder (1996) who showed that polluting intermediate inputs should be subject to taxation. In our context, administrative concerns alone could justify taxes or subsidies of bases or activities other than consumption, wages, or profits to the extent that they “pollute” or “clean” the tax system due to their administrative implications.

## **II. Conclusions**

Many tax policy choices revolve around administration and enforcement issues, for which firms and firm-to-firm transactions are critical, and which cannot be informed by the kind of theoretical underpinning that Diamond and Mirrlees (1971) introduced in the early 1970s. We offer the model of this paper as a springboard for thinking about an alternative underpinning, and argue that any such model must recognize the administrative efficiency advantages of business-based tax remittance, monitorable arms-length transactions, and economies of scale.

We highlighted three dimensions on which models with tax enforcement differ from the more standard optimal tax analysis. First, the budget-constraint-based equivalences among taxes on consumption and taxes on income and profits break down. Second, firms are critical in the tax system because they give rise to relatively easy-to-monitor transactions and can minimize the number of private agents the tax authorities must deal with. Third, certain behavioral responses that matter in this context have not yet been adequately addressed in the public finance literature, such as the size of and boundaries between firms as well as income shifting responses across different bases.

## References

- Bird, Richard M. "Why Tax Corporations?" prepared for the Technical Committee on Business Taxation, Canada, Working Papers: No. 96-2, 1996.
- Bovenberg, Lans and Lawrence Goulder. "Optimal Environmental Taxation in the Presence of Other Taxes: General-Equilibrium Analyses." *American Economic Review*, 1996, 86 (4), pp. 985-1000.
- Diamond, Peter A. and James A. Mirrlees. "Optimal Taxation and Public Production I: Production Efficiency." *American Economic Review*, 1971, 61 (1), pp. 8-27.
- Heller, Walter P. and Karl Shell. "On Optimal Taxation with Costly Administration." *American Economic Review*, 1974, 64 (2), pp. 338-345.
- Plumley, Alan. *Overview of the Federal Tax Gap*. Washington, D.C.: U.S. Department of the Treasury, Internal Revenue Service, NHQ Office of Research, 2004.
- President's Advisory Panel on Federal Tax Reform. *Simple, Fair, and Pro-Growth: Proposal to Fix America's Tax System*. Washington, D.C., 2005.
- Slemrod, Joel and Shlomo Yitzhaki. "The Costs of Taxation and the Marginal Efficiency Cost of Funds." *International Monetary Fund Staff Papers*, 1996, 43 (1), pp. 172-198.

---

## FOOTNOTES

\* Wojciech Kopczuk, Economics Department, Columbia University, 1022 International Affairs Building, MC 3308, 420 West 118<sup>th</sup> Street, New York, NY 10027, [wkopczuk@nber.org](mailto:wkopczuk@nber.org); Joel Slemrod, Stephen M. Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109-1234, [jslemrod@umich.edu](mailto:jslemrod@umich.edu). We thank Dhammika Dharmapala and Shomo Yitzhaki for helpful comments on an earlier draft. Kopczuk acknowledges financial support from the Alfred P. Sloan Foundation.

<sup>i</sup> Diamond and Mirrlees (1971) proved that, under certain conditions, production efficiency is a necessary condition of an optimal tax system even when lump-sum transfers are not feasible.

<sup>ii</sup> For example, over 80 percent of U.S. federal taxes are remitted by businesses, although only 10 percent are nominally business taxes.

<sup>iii</sup> For public corporations, the information in financial statements provides some information to the authority.

<sup>iv</sup> According to Alan Plumley (2004), in 1992 the net misreporting percentage for wage and salary income was 0.9 percent, compared to 31.8 percent for amounts subject to little or no information reporting—predominantly business income.

<sup>v</sup> If the true economic profit was always uniformly equal to zero, government would use this information to question any other reported value. However, consider a simple extension of the model to account for uncertainty that represents production as  $X_1=L_1+\varepsilon_1$  and  $X_2=g(L_2, D)+\varepsilon_2$ , respectively, where the  $\varepsilon_i$  terms are mean zero. This would produce a distribution of firm profits with (assuming linear tax structures with full loss offset) no changes to the behavior of risk-neutral firms, and therefore wouldn't affect our model except for breaking the equivalence between reporting profits and tax avoidance.

<sup>vi</sup> We also sidestep the possibility that, once administrative costs are introduced, it may be optimal to distort this relative price.

---

<sup>vii</sup> This assumption means that the social cost of evasion (or, later, income shifting) shows up not as the excess burden of revenue-lowering behavioral response but solely as administrative cost. A more general model would allow enforcement intensity to be chosen optimally.

<sup>viii</sup> Note that in this setup all firms are retail firms, although only some sell exclusively to consumers. In reality, the number of firms that have a retail business is significantly less than the total number of businesses, so that a retail sales tax requires that the tax administration deal with less tax units than under any of the other three taxes.

<sup>ix</sup> For the sake of notational simplicity we do not distinguish among the  $b(\cdot)$  functions that apply to qualitatively different bases, although we recognize that the enforcement issues differ.

<sup>x</sup>Walter P. Heller and Karl Shell (1974) introduced an alternative notion of production efficiency by redefining the production possibilities frontier to account for administrative costs.