Abstract - This paper argues that who remits tax may be an important aspect of implementing a tax system, in spite of standard economic analysis that maintains that which side of a taxed market remits is completely irrelevant. The irrelevance proposition does not apply in the presence of avoidance and evasion (i.e., in all real tax systems) because the total resource costs of administering a given effective tax structure may vary depending on the remittance system and because the opportunities for avoidance and evasion and the technology of enforcement affect the incentive to demand and supply the taxed activity.

ISSUES AND MOTIVATION

Nearly all of the modern economic theory of taxation is concerned with what actions or states of affairs trigger tax liability, and virtually none is concerned with who or what entity remits funds to the government to cover that liability. Indeed, both elementary public finance textbooks and advanced surveys suggest that the remittance responsibility—such as whether the buyer or seller of a commodity must remit any sales tax triggered by the sale—is irrelevant to the consequences of a tax.

Some would find the dismissal of remittance to be surprising. For example, many conservatives consider the introduction of income tax withholding in 1943 to be a crucial characteristic of the U.S. income tax. Dick Armey, the former Republican House Majority leader, has called it “the crucial, deceptive device that has made big government possible” and Milton Friedman, who helped develop it when he was at the Treasury Department during World War II, later considered it a mistake and called for its abolition.1 To tax administrators and tax practitioners, particularly those who work in developing countries, the remittance system is not irrelevant, and may instead be critical to the efficient and equitable operation of a tax system. Certain aspects of remittance systems are pervasive. We observe, for example, that remittance by business

---

1 See Armey (1996, p. 99) and, reporting on an interview with Milton Friedman, Doherty (1995).
rather than by individuals pervades modern tax systems. Among business-based remittance systems, the value added tax has, largely on implementation grounds, come to dominate the retail sales tax for broad-based, non-graduated, consumption taxes. Despite their prominence in tax policy considerations, these implementation issues have received very little careful analytical attention.

In this paper I contribute to such a consideration by addressing some questions in the economics of tax remittance. I discuss the conditions under which the standard presumption of irrelevance holds, and conditions under which it breaks down. I then relate this theory to two issues. The first is income tax withholding, where I note the overwhelming importance of firms in tax remittance and argue that to understand the impact of taxes one needs to explicitly introduce firms into the positive and normative theory of taxation. Second, I consider remittance in the context of the choice between a value added tax (henceforth VAT) and a retail sales tax (RST). This paper is not intended as a comprehensive review of remittance issues that arise in taxation.2

Before proceeding, some limits on this inquiry should be noted. Although remittance is an important aspect of a tax system, it is not the only important aspect of how a tax system is implemented. Also important are the administration and enforcement of the tax rules, as well as the design of the tax rules with the administrative and enforcement issues in mind. One key element is the extent of information reporting—reports to the tax authority from one party about transactions that have tax consequences for another party.

For example, employers provide information reports regarding payments of wages and salaries to employees, presumably so that the tax authority can check that the employee reported the wages and salaries on their personal tax return and also to facilitate a reconciliation with the individual taxpayer’s claims about the amount withheld on his behalf. With final (and perhaps exact) withholding, the information report is not as critical, because no reconciliation with the other party (employee, in this example) is required.3

The economics of remittance is distinct from the framing effects of tax policy—whether the form (holding substance constant) in which a policy is presented affects its consequences, because it affects how people react to it. For the most part, I will put framing issues aside, although I address briefly the issue of the role of remittance in the perception of tax policy.

TERMINOLOGY

Because one of the objectives of this paper is to clarify the distinction between what triggers tax liability and how these rules are implemented, I will try to be very careful about the usage of some key terms. This may be a bit disconcerting because the common, and sometimes official, terminology used to describe a tax system might not match up with the consistent terminology I employ, and recommend.

The key to the quest for semantic clarity is the careful use of the term “remit” and its various forms. I will here use (and urge the use of elsewhere) “remit” tax to refer to some person or entity in the private sector

---

2 For example, it does not discuss alternative forms of corporate integration as alternative withholding or remittance regimes.

3 There may be legal sanctions for not providing information reports, which may be supplemented by a link between the deductibility of payments and the provision of the information reports. For example, employer deductions for wages and salary payments could be made effectively contingent on the provision of the information reports. Note that, in general, withholding/remitting implies that an information report has been made, but the converse does not always hold.
writing a check or otherwise transmitting funds to the tax authority.\(^4\) For example, under a value added tax all businesses (perhaps above a prescribed turnover threshold) may be required to remit funds to the tax authority, while under a pure retail sales tax only retail businesses are required to remit money. It may be that, under certain conditions, a broad–based uniform VAT and RST end up having the same consequences, but even so the remittance pattern is quite different. Indeed, my principal point is that in many situations the remittance rules, and more generally the implementation, of a tax system matter for its consequences.

A compact lexicon based on the concept of remittance defines a tax system as a set of rules, regulations, and procedures that define 1) what events and/or states of the world trigger tax liability and how to calculate the liability (tax bases and rates), 2) who or what entity must remit that tax and when (remittance rules), and 3) the procedures for monitoring and ensuring compliance, including the information reporting requirements and the consequences (e.g., penalties) of not remitting the liability in a timely fashion (administrative and enforcement rules). Under the assumption that what triggers tax liability can be ascertained and collected costlessly, 2) is irrelevant and 3) is unnecessary. These elements of a tax system together determine the incidence of a tax system—who bears the burden—the allocation effects, and the total costs of collecting the tax revenue, including the distortion, the administrative, and compliance costs.

It is crucially important to distinguish between 1) the person who ultimately bears the burden of a tax and 2) who remits tax. This is because a tax system generally causes changes in pre–tax prices, and thereby the burden may be shifted away from the apparent statutory bearer. For example, a tax triggered by labor earnings will in general increase the pre–tax wage, so that the after–tax wage falls by less than the tax. Thus, the burden of the tax is shared between employers, who face a higher pre–tax wage (cost of labor) than otherwise, and employees who receive a lower after–tax wage rate than in the absence of the tax. Exactly how this burden is shared (i.e., how much the pre–tax wage rate increases and how much the after–tax wage declines) depends largely on the relative elasticity of the demand for labor and supply of labor: more inelastic labor supply and more elastic labor demand makes it more likely that employees will bear the burden of the tax; more inelastic labor demand and more elastic labor supply makes it more likely that employers will bear the burden of the tax.\(^6\) I will say that an individual “bears the burden” of a tax to the extent that he or she experiences a loss of utility (sometimes referred to as “real income”) because of the tax.\(^7\)

\(^4\) In some situations, such as a negative income tax, when withholding exceeds the tax liability, or for exporters under a VAT, and there is a reconciliation (“refund”), the government transmits funds to a private individual or entity. I discuss later what is often referred to as withholding by government entities.

\(^5\) This is true in a static, partial equilibrium analysis of a tax; in a general equilibrium analysis, other factors come into play.

\(^6\) The burden of a tax may be shifted to parties who are on neither side of the taxed transaction. The classic example concerns how consumers of margarine might be made worse off because of a tax on butter. To the extent there is substitutability between the two commodities, the tax on butter will cause the consumer price of butter to rise somewhat, which will cause some consumers to shift consumption to margarine. This will raise the price of margarine, causing a utility loss to those who purchase margarine even if they consume no butter at all.

\(^7\) Economists often assert that businesses do not bear the burden of taxes. This is not a statement that taxes will never reduce the rate of return to investing in, or running, a business. Rather it is a methodological statement that, in order to understand the distributional impact of any tax (i.e., its “incidence”), the burden must be traced beyond the impact on the profitability of a legal entity to its impact on the well–being (i.e., utility) of individuals. The owners of the business may bear part of the burden of a tax that is legally “on” business, but part may be shifted to others.
The compact lexicon I propose avoids the use of several terms that are commonly used to describe tax systems. One such term is the statutory incidence, with a loose shorthand of who the tax is levied "on." I will certainly talk about withholding, which refers to a situation in which the remitter differs from the statutory bearer. But, framing issues aside, what anyone calls a tax, or who the law says the tax is "on," is irrelevant to its impact except insofar as this affects who has the legal responsibility to remit and the consequences of failure to remit. For example, under the U.S. personal income tax, the amount of tax remitted by the employer/withholder does not generally equal the actual liability, and it is the responsibility of the individual to either remit the difference (if positive), or to file for a refund. The code also specifies how much the employer must remit on the employees' behalf, and specifies penalties for failure to remit appropriately. In common parlance, we say that the personal income tax is a tax on the individual, even though much of it must be remitted by the individual's employer and, as discussed below, the ultimate incidence may be shifted from the employee–individual as well as from the employer–remitter.

Because of the imprecision of its meaning, I will avoid saying that an individual (or a legal entity) "pays" taxes. Writers on taxation, including economists, use the phrase to "pay" taxes sometimes to refer to the remittance of money to the tax authority, and sometimes to refer to bearing the burden of a tax. In some situations the remittance is due to the tax authority at some time after the transaction (or state of the world) that triggers the tax liability; in this case the effective tax rate is lower than the stated tax rate because of the time value of money, sometimes referred to as the "cash–flow" value of the delayed remittance. The incidence and other consequences of this way of delivering a tax reduction are not

---

8 The equivalent term in VAT discussions is "charge."
9 For example, a business may tell its employees it is withholding income tax on their behalf, and not actually remit the tax on their behalf. My semantic stance is that this does not constitute "collection," nor can I think of a meaningful way to determine whether or not a business has collected a tax given its remittances. Some are comfortable using "collect" to mean "benefit from," but I prefer the latter term in that situation.
10 As Tanzi (1977) has pointed out, in periods of high inflation, this can be a significant factor.
fundamentally different from an explicit tax cut.

WHEN REMITTANCE DOESN’T MATTER

In the standard theory of optimal taxation, all the information needed to assess tax liability is observed costlessly by everyone. This assumption about the availability of information precludes tax evasion and, thus, the necessity of an enforcement regime. It also implies the standard textbook canon: which side of a market remits (or is apparently liable for) a tax is irrelevant, both for the ultimate incidence (i.e., who bears the burden) of the levy as well as its efficiency and allocation consequences.11 Dalton (1954) called the former the “Theorem of the Invariance of Tax Incidence.”12

This invariance, or irrelevance, proposition occupies a central position in the economics of taxation. Consider its treatment in Rosen (2002), a leading American undergraduate public finance textbook. Rosen first defines the “statutory incidence” of a tax as who is “legally responsible for the tax,” and the “economic incidence” as the “change in the distribution of private real income induced by a tax.” The irrelevance proposition is then stated as follows: “knowledge of statutory incidence tells us essentially nothing about who really pays the tax” (p. 254); here the emphasis is in the original, and the term “pays” obviously refers to the change in utility, or real private income and “statutory incidence” refers to remittance responsibility, although the word remittance is not mentioned. Rosen later presents a compelling remittance metaphor: “It is irrelevant whether the tax collector (figuratively) stands next to consumers and takes u dollars every time they pay for a gallon of champagne or stands next to sellers and collects u dollars from them whenever they sell a gallon” (p. 261). Undergraduate public finance students (and instructors) are familiar with the diagrammatic version of the irrelevance proposition: the imposition of a unit tax can be analyzed as a shift in the demand curve facing suppliers or a shift in the supply curve facing consumers, with no difference in the equilibrium prices each faces and the quantity bought and sold. This way of thinking is, to be sure, an important insight into the incidence of taxation, and is worth the expositional effort that Rosen, and others, give it; it is not, I argue, true in many situations.

The treatment of the irrelevance proposition in more advanced discussions is also revealing. In their chapter on tax incidence in the authoritative Handbook of Public Economics, Kotlikoff and Summers (1987, pp. 1045–6) assert, in the context of a partial equilibrium model of an excise tax on a product, that “it is clear that the determination of the equilibrium quantity, the price paid by consumers, and the net of tax receipts of producers do not depend on which side of the market the tax is levied . . . .” They go on to say that this irrelevance proposition carries over to much more general contexts, and underlies the idea that the ultimate incidence of a tax cannot be assessed by simply looking at where the tax is “approximately levied.” Just before their statement of the irrelevance proposition in terms of “where the tax is levied,” Kotlikoff and Summers refer to the two alternative tax systems as one where “the tax is collected from sellers” and one in which “the tax is collected from buyers,” apparently referring to acts of remittance rather than statutory inci-

---

11 This paper considers remittance in the context of a single tax jurisdiction. Other remittance issues arise in the context of multiple jurisdictions.

12 A related question is whether there is an invariance theorem about the identity of the provider of information to the tax authority.
dence: “The real equilibrium is invariant to whom the government requires mail in the tax payment.”

The irrelevance proposition is asserted only in the context of a partial equilibrium analysis of a single market and two remittance alternatives (buyer or seller), and its generality has not been addressed. What conditions must be in place for irrelevance to hold? For example, what if the tax triggered by an individual’s earning labor income must be remitted by a large company that does business with the individual’s small employer? Are the incidence (and efficiency) consequences the same as in the case where the employer must remit the tax? Presumably the proposition requires that the remitter must have an ongoing commercial connection with the parties to the tax–triggering behavior, but the precise conditions have not been carefully spelled out.

One can imagine remittance systems that involve not only the buyer and seller of commodities that trigger tax liabilities. For example, Scott (1998) notes that the tax systems of pre–modern and early modern states dealt more with communities than with individuals. The collective form of taxation generally meant that it was in the interest of local officials (or nobles) to misrepresent the local situation to reduce the local tax remittances by, for example, understating the local population or the acreage under cultivation. In this case, to understand the incidence and allocation effects of the tax one would need to know how the required remittance affected the terms of the relationship between the community or noble, on the one hand, and the residents and peasants on the other hand. This question has not been addressed in the modern theoretical literature, undoubtedly because in modern economies taxes levied on communities are rare—although the use of (sometimes unrestricted) grants to jurisdictions is a central feature of fiscal federalism. Such fiscal arrangements raise the issue of the precise conditions under which an invariance proposition holds.

WHEN REMITTANCE MATTERS

Standard Theory

The standard literature occasionally refers to certain exceptions to the irrelevance proposition. But one oft–used example—the minimum wage—is not related to remittance. The common assertion is that the impact of the imposition of a minimum wage in the presence of existing taxes will be different depending on whether the taxes are imposed on the employer or the employee. This occurs when the statutory minimum wage is defined in reference to the after–business–tax, before–personal–tax wage rate, sometimes referred to as the “market wage rate.” The market wage rate has no relevance in a competitive market: the only price that matters for the worker’s labor supply and utility is the after–all–taxes wage rate, and the only price that matters for the employer is the before–all–taxes wage rate.

13 Liability for the Russian “soul tax” introduced by Peter the Great in 1718 was triggered by the number of individuals in a community and remitted by representatives of the communities or by the nobles whose subjects they were; failure to remit the required sum usually led to collective punishment. This is similar, but not identical, to what is called “tax farming,” in which the government would sell the right to collect taxes to a private citizen or business. In this case the tax farmer is the remitter, and will presumably employ resources to maximize tax collections (subject to the tax law), which will lead to a socially inefficient over–allocation of resources to enforcement. In colonial Vietnam a tax was levied on communities on the basis of their presumed population. If the sum was not remitted, the central tax authority would hold an auction of whatever they could seize until they had collected the required sum. Scott (1998, fn. 69, p. 38) notes that this system gave the village notables, who owned most of the goods worth seizing, an incentive to make sure the taxes were remitted.
Theory of Tax Systems

More to the point of this paper, the invariance result can also break down when the standard model is extended to recognize the realities of tax implementation—the possibility of noncompliance and the need to monitor tax remittances, which lead to administrative and compliance costs. Here the imagery of the tax collector standing at the cash register and receiving the tax liability either from the buyer or from the cashier is misleading. In that image the social cost of the tax authority collecting the exact amount of tax is small or zero and, critically, does not depend on who hands over the cash. In some situations, that image is misleading because it suggests that enforcing compliance is so trivial that the remittance details are irrelevant.

The implication of this metaphor can be supplemented by some simple partial equilibrium analysis of an excise tax. In equilibrium, supply (\(S\)) must equal demand (\(D\)). Thus, in the absence of taxes, this means that

\[
S(p) = D(p),
\]

where \(p\) is the market price of the good. Now consider what happens when an (additive, not ad valorem) excise tax must be remitted by the seller/supplier. Then it must be true that

\[
S(p' - t) = D(p'),
\]

where \(p'\) is the new equilibrium price paid by the buyer to the seller. Alternatively, if the consumer must remit the tax, then in equilibrium it must be that

\[
S(p'' + t) = D(p''),
\]

where \(p''\) is the equilibrium price paid by the buyer to the seller. Clearly whatever equilibrium holds for [2] also holds for [3], as long as \(p' = p'' + t\). When the consumer must remit the tax to the government, he or she pays the supplier less, but the after–tax price received by both the supplier and the consumer will be exactly the same. What drives this result is that i) both supply and demand depend only on the after–tax price suppliers and demanders face, respectively, and not on an arbitrary designation of what the “market price” is, and ii) the price, \(p\), is completely flexible.

A theory of tax systems generalizes this, so that if the supplier remits the tax the supply function is \(S(p, t, x)\), and if the consumer remits the demand function is \(D(p, \tau, y)\), where \(t\) and \(\tau\) are the tax rates remitted by the supplier and demander, respectively. The variables \(x\) and \(y\) refer to the avoidance technology, including the set of tax instruments such as audit intensity that affect supply and demand, respectively. The standard equilibrium condition is a special case of this when neither \(x\) nor \(y\) matters, and when \(p\), as well as \(t\) and \(\tau\) (given \(t + \tau\)), matter only insofar as they affect the after–tax price faced by the market participants.

To make this idea more concrete, consider the simple model developed in Slemrod (2001) in the context of the labor market and a proportional income tax. In the standard model of labor supply in which the individual chooses labor (\(L\)) to maximize utility, that is:

\[
\text{Max}_L U(Y, L), \text{ subject to } Y = (1 - t)wL + M,
\]

where \(Y\) is private consumption, \(w\) is the pre–individual–tax wage rate, \(t\) is a linear income tax rate, and \(M\) is non–labor income. The individual chooses both how much labor to supply and how much labor income to hide from the tax authority (\(H\)), and so solves the following problem:

\[
\text{Max}_{L,H} U(Y, L), \text{ subject to } Y = wL - t(wL - H) - C(wL, H) + M,
\]

where \(C\) is the private cost of avoidance, which depends on the amount of true
labor income and the amount of hiding/avoidance; and it is natural to assume that $C_2$ and $C_{22}$ are positive, so that the marginal cost of avoidance increases with $H$. As long as $C_1 = \frac{\partial C}{\partial (wL)} < 0$, the effective after-tax wage includes an implicit subsidy to working equal to $wC_1$, because earning more income lowers the marginal cost of avoiding taxes by that amount. In this case, labor supply cannot be written simply as a function of the apparent after-tax wage facing the worker, $w(1 - t)$, and also depends on the function $C(\cdot)$, so that it generates a labor supply function of the form $S(w, t, x)$, where $x$ refers to the employee avoidance technology captured by $C(\cdot)$.

One can construct the same type of model for a business deciding both how much labor to employ and how much avoidance to do, so that the demand for labor becomes $D(w, \tau, y)$, where $\tau$ is the employer tax on labor payments and $y$ captures the element of the firm avoidance technology, including tax system parameters other than the tax rates themselves. One can show, by manipulating the expanded version of [1], [2], and [3], that the incidence irrelevance proposition holds if and only if $D_w - D_\tau = S_w + S_\tau$, where $D_i$ and $S_i$ denote partial derivatives of $D$; note that $D_w D_\tau$ and $S_w S_\tau < 0$, but $S_w > 0$. The standard model trivially satisfies this. In the extreme, a truly lump-sum (i.e., unrelated to any behavior) compliance cost will to a first approximation be borne entirely by whom it is directly imposed on. The issue is a bit more complicated for firms in competitive markets with free entry and exit, because costs that are inframarginal to the supply of any particular activity will in equilibrium have to be offset by some other factor. For example, if all toy manufacturers were subject to a compliance cost that was unrelated to how many toys they produced, then the price of toys would almost certainly rise as a consequence.

14 Slemrod (2001) calls this the “avoidance–facilitating” quality of providing labor supply.
15 Involuntary costs would be equivalent to assuming that $C(wL, 0) > 0$.
16 In the extreme, a truly lump-sum (i.e., unrelated to any behavior) compliance cost will to a first approximation be borne entirely by whom it is directly imposed on. The issue is a bit more complicated for firms in competitive markets with free entry and exit, because costs that are inframarginal to the supply of any particular activity will in equilibrium have to be offset by some other factor. For example, if all toy manufacturers were subject to a compliance cost that was unrelated to how many toys they produced, then the price of toys would almost certainly rise as a consequence.
Efficiency Implications

If who remits has no consequences, then, aside from the kind of political and perceptual issues discussed later, the policy question “who should remit?” is not interesting. But the fact that most countries’ tax systems follow similar remittance patterns suggests that this feature of tax systems not only has consequences, but that some remittance arrangements are better than others. In other words, there is no efficiency invariance property with respect to the remittance system.

We can say a bit about the optimal pattern of remittances because the modern theory of optimal taxation can be extended in a straightforward way beyond the tax instruments that are the usual focus of attention of the standard model—such as tax rates and tax bases—to address the instruments that are central to tax systems, such as the pattern of tax remittances, as well as the extent and nature of tax enforcement and the pattern of information reporting. One lesson of this inquiry is that, other things (such as excess burden from behavioral response) equal, revenue-neutral marginal changes in tax instruments should minimize \( C + mA \), where \( C \) stands for compliance costs borne directly by private agents, \( A \) stands for administrative expenses incurred by government, and \( m \) represents the marginal efficiency cost of raising funds. The marginal efficiency cost of raising funds is the burden on society per net dollar raised and is generally greater than one, reflecting the fact that raising revenue usually generates costs over and above the transfer of money from the private to the public sector. Putting a relatively higher weight on administrative costs accounts for the fact that these expenses come out of funds that were raised with tax instruments that have a marginal efficiency cost of raising funds in excess of one.

Voluntary Reassignment of Remittance Responsibility

If the statutory remittance system is not the least costly, why don’t the interested parties contract among themselves to arrange a new system, sharing the cost savings to make all parties better off? This may happen, as the case of escrow for property tax, discussed below, suggests. However, what is in the interest of the parties to the tax–triggering transaction may not be in the general interest. What minimizes compliance cost may not minimize the sum of compliance plus (properly weighted) administrative costs, and may also reduce revenue collected.

One relevant example involves escrow arrangements for property taxes. In many cases at closing the buyer of a property and the mortgage lender agree to set up an escrow account for the remittance of property taxes and, often, insurance costs that are added to the payments due on the mortgage. The collections are placed in a depository account that the lender manages, remitting taxes to the local tax authority as the tax liabilities come due. A mortgage lender can require the borrower to have an escrow account if the loan–to–value ratio exceeds 80 percent. When the homeowner’s equity exceeds 20 percent, the lender is often willing to rely on the homeowner’s self–interest to remit the taxes (and insurance premiums), although lenders may require a one–time escrow waiver fee of

---

17 This argument is developed in Slemrod and Yitzhaki (1996). This does not address the distributional consequences of alternative tax systems.

18 Reconfiguring the remittance arrangements might not be legal. Presuming that minimizing the sum of compliance costs is the bargaining objective ignores the possibility that the total tax liability might change.

19 In 1934 the federal government mandated that lenders manage escrow accounts on all FHA–insured mortgages.
The system is not perfect. Sometimes the mortgage company and the homeowner both end up remitting the tax due, and in some cases the mortgage holder company fails to remit the taxes or remits on the wrong account number. If the homeowner is current on payments to the escrow account and the mortgage company fails to pay on time, state and federal law require that the lender, and not the homeowner, pay any late penalties or fees. However, the homeowner is ultimately responsible for the remittance of the tax due.

Lenders benefit from an escrow arrangement because it reduces the risk to them that a property (which is their collateral) can be seized by the local government for unpaid taxes. Some argue that a mortgage escrow account protects the homeowner from the “anxiety” of remitting property taxes (and insurance premiums) “all at once,” in addition to protecting them against inadvertently failing to remit and the hassle of remitting. The Internet site of a prominent mortgage lender asserts that it benefits local governments by providing a more efficient, less expensive means of tax collection: “Rather than working with millions of homeowners, municipalities need only collect from a few hundred lenders.”

The issue of designating workers as employees versus independent contractors is also instructive. The distinction has many non–tax implications, and an important tax consequence. For employees, the employer must withhold income and Social Security taxes. If, however, the worker is determined to be an independent contractor, the employee rather than the employer is responsible for remitting both income and payroll taxes, but the business must (for annual payments that exceed $600) file an information return detailing the payment, except if the contractor is a corporation. The legal distinction between an employee and an independent contractor is based on many factors, and is the source of ongoing controversy.

Even though eliminating withholding would probably reduce compliance costs for the employer to handle the tax remittances for the workers, it would facilitate evasion by the workers. This supposition is consistent with the evidence presented in an IRS study for tax year 1974 that concluded that while employees reported over 99 percent of their wage and salary income, independent contractors voluntarily reported just 74 percent. However, because 1974 predates substantial extension of information reporting to independent contractor income, the difference in evasion rates reflects the impact on compliance of both withholding and information reporting.

Opponents of extending withholding to independent contractors often cite two problems. First, to the extent that independent contractors have substantial business expenses that reduce annual net income and taxes owed, withhold-

20 A few states have enacted laws that require lenders to pay interest on the funds held in escrow accounts, which typically amount to a two–month surplus.
21 See Section 6(g) of the Real Estate Settlement Procedures Act.
22 This is in contrast to employer withholding, where in the United States in most cases the employee is not responsible if the employer fails to remit the withholding tax.
23 The quotations in this paragraph are taken from manhattanmortgage.com (2006).
24 Presumably, by either explicit collusion or an unstated understanding, independent contractors may accept a lower payment than would have to be paid to employees, so that the expected tax saving is shared among the employer and the worker.
26 A number of countries do require remittance of tax related to payments to independent contractors.
ing based on gross payments could substantially overstate net income and, thus, could adversely affect cash flow (in the period before over–withholding can be reclaimed); in our terminology, this increases the effective tax rate on that income. More generally, it may be complicated to estimate an appropriate withholding rate because of the difficulty of knowing the independent contractor’s total annual income. Another issue is that independent contractors may avoid withholding by organizing their business as a corporation.

Non–replicability

In some situations the remittance system constrains what actions or states of the world trigger tax liability (or compliance cost), so that some tax–triggering regimes cannot be replicated, or cannot easily be replicated, by any given remittance system. Consider two examples. The first concerns a retail sales tax, which is remitted by (retail) businesses. It is well–documented that the compliance cost of an RST depends on business size in a regressive way, that the marginal compliance cost is less than the average cost and, therefore, that the direct, that is before shifting, cost as a ratio to size is smaller for larger companies.27 If the retail sales tax instead had to be remitted by the consumer and, plausibly, the tax rate could not be adjusted to reflect the size of the business from which it was purchased, the two remittance systems would have different consequences. For example, while the business–remitted tax would favor larger firms, and provide an incentive for business consolidation in the retail sector, the consumer–remitted system would not—and could not—have this consequence.

Next consider a personal labor income tax that is global and graduated, so that tax liability depends on an individual’s (or married couple’s) total income and certain characteristics. To take an extreme example, a tax–and–transfer system that included a transfer to people with no labor income (e.g., a negative income tax) could not be entirely implemented by an employer remittance system—which employer would receive the subsidy on behalf of any given worker? More generally, the details of implementing a graduated tax schedule when workers have multiple employers may have real consequences. This is certainly true in the case of final withholding, when the withheld tax is the sole determinant of the tax liability, and will be true when the administration of withholding in the case of multiple withholders is imperfect.

Note that both of the examples in this section of the paper concern personalized28 taxes, by which I mean taxes for which the tax burden triggered may depend not only on the transaction or activity but also some characteristic of a transacting party. In the first example, the marginal compliance cost depends on the size of the retailer. In the second example, the total individual tax liability depends on the size of the taxpayer’s total income.

Business–remitted taxes might serve as a way to constrain the amount of personalization of tax burden that is practically possible. The argument goes as follows: it is difficult if not impossible to subsidize, for example, big families, with a VAT or RST. With individuals not filing returns, the natural process for providing information about a family’s number of dependent children is absent. Business–remitted taxes can to some extent incorporate these features, however. Under the British Pay

27 Although, as Slemrod (2004) notes, the noncompliance rate is also higher for small, compared to large, businesses.

28 This definition of personalized taxes is the closest I can think of as a precise definition of what a “direct” tax is, but I will not raise that semantic flag here.
As You Earn exact cumulative withholding system (with no individual tax returns for those individuals who qualify), there is tax liability differentiation by family size; the requisite information is passed from the taxpayer to the government to the employer, which adjusts the withholding accordingly. Under recent proposals for a national sales tax in the United States, the tax would be supplemented by direct payments to households that depend on family size, requiring a separate bureaucracy to implement.

WITHHOLDING

History

History provides many instances illustrating the importance of one prominent aspect of tax remittance systems— withholding. The first modern British income tax, introduced in 1799 by Prime Minister William Pitt, was judged to be a failure, in that actual revenue collected fell well short of the expected amount. A revised version of the tax, introduced in 1802 by Pitt’s successor William Addington, was much more successful, and the structure has remained in place for two centuries. The key change between Pitt’s 1799 fiasco and Addington’s 1802 success was the introduction of remittance at source, which applied to interest, dividends, rent, income from government securities, and compensation of certain civil servants.

Withholding was also an important feature of the first U.S. income tax, instituted in 1862 during the Civil War. Withholding was applied to dividends paid by certain financial institutions and insurance companies, and salaries of federal employees. The administrative advantages were well-recognized: “[It] was much easier and simpler to collect [tax] from the corporations than from the individual stockholders and bondholders...” (Hill, 1894, p. 416, 427). In 1865 almost 40 percent of income tax receipts came from withholding (Soos, 1990). The Civil War income tax was eliminated soon after the war ended, and a new income tax enacted in 1894 was not implemented because the Supreme Court ruled that it violated the Constitutional stipulation that all direct taxes had to be apportioned among the states according to each state’s population. The Sixteenth Amendment to the Constitution, passed in 1913, allowed a federal income tax, and the Congress immediately enacted one. From 1913 to 1916 the income tax law had broad withholding provisions. These proved unpopular among taxpayers and businesses, and they were replaced in 1917 by information reporting (Lent, 1942). Advocates of mass-based income taxes in the United States had to face the

29 There are fascinating historical examples of changes in remittance responsibility. In early 19th century Ireland, Catholic and other Irish farmers were refusing to remit tithes meant for the Protestant church, with the rate of noncompliance reported to be over one-half. An Act passed in 1838 made landlords responsible for remittance. According to Douglas (1999, p. 62) this mattered little (for, presumably, incidence) because the peasantry was “rack-rented,” meaning that the rent was set to be the entire produce minus what was necessary for bare subsistence. But, under the arrangement the payment of tithes was less visible and, therefore, less of an irritant, and noncompliance fell. In terms of the model presented earlier, the incentives to evade apparently differed depending on who was responsible for remitting.

McCulloch (1975, p. 156) reports that at one point the Roman emperor Nero ordered that the four percent sales tax on slaves be remitted by the seller instead of the buyer. McCulloch approvingly quotes the Roman historian Tacitus on the incidence of this change: “The tax of one twenty-fifth part of the goods for sale was remitted, nominally rather than with any real effect, because although the seller was ordered to pay it, it accrued to the buyer as part of the price” (italics added). Tacitus observed that the invariance theorem held in this case.

30 It raised the same revenue as Pitt’s tax, at half the rate.

31 One might note the similarity to another revolutionary development in public services soon after—the British introduction in 1840 of pre-paid postage stamps—that replaced a system in which the receiver of a letter or package paid the carrier, and led to an enormous expansion in the use of the post office.
reality that about one–third of Americans were occupied as farmers and another third owned or worked in small, usually unincorporated non–farm businesses and, for these people, the government had no way of discovering their identities. The mechanism to do so was not introduced until the 1930s with the Social Security system, including the Social Security number and the withholding of payroll taxes at source.32 It was not until 1943, in the face of the vast revenue needs caused by World War II, that withholding for wages and salaries was again enacted into law, at the same time the income tax was extended to a broad swath of the population by drastically reducing personal exemptions.

Practice in Developed Countries

Withholding for income tax is now widespread among developed countries. Employer withholding for wages and salaries is required in 28 of 30 OECD (Organisation for Economic Co–operation and Development) countries, all except France and Switzerland.33 The use of withholding on other sources of income varies. Twenty–two of 30 OECD countries have withholding on dividends, and 21 of 30 have withholding on interest; the United States does not have either except if the taxpayer does not provide an identification number to the payer.34 Twelve of 30 OECD countries have withholding on at least some independent personal services;35 the U.S. does not.

Of the 30 OECD countries, exactly half have a system of cumulative withholding for labor income that is mainly return–free. Nine, including the United States, have a system of non–cumulative withholding where a return is required. Four others use a reconciliation approach with pre–populated, or pre–filled, returns sent to taxpayers based on information known to the tax authority, requiring taxpayers only to validate, or correct, the returns. As mentioned, only two countries—France and Switzerland—have no withholding, and instead require an annual return from taxpayers and regular installment payments.

Principles of Withholding

The principal justification for withholding is that it economizes on the administrative and compliance costs of collecting a given amount of tax revenue, $C + mA$ in the earlier notation. These cost savings are more likely to be realized when the withholding agents are fewer in number than the taxpayers on whose behalf they are remitting the tax. In addition, withholding is facilitated when only those entities that have the necessary accounting and bookkeeping capabilities and are otherwise able to carry out the withholding are designated as agents. In order to withhold the correct amount of tax from a single payment, a payer theoretically must know the recipient’s total taxable income for the year; in an employment situation, where

32 This history is adopted from Brownlee (1996). Note, though, that Social Security was not extended to regularly employed farm and domestic workers and the non–farm self–employed until 1950, and not until 1954 for self–employed farmers.
33 All of the data cited in this paragraph comes from OECD (2006). In France, there is withholding for social contributions but not for personal income tax. In Switzerland, aliens in possession of a work permit are subject to withholding.
34 The most recent U.S. effort to enact withholding on interest and dividends was the Tax Equity and Fiscal Responsibility Act of 1982. The withholding provisions were repealed before they took effect. “Backup withholding” at a rate of 20 percent, which applies to taxpayers who do not provide identification numbers to payers, was enacted in 1983.
35 Two more countries have it when the taxpayer does not provide a taxpayer identification number to the payer.
there is a continuing relationship between the employer and employee, the employer usually knows the annual remuneration of the employees (from that job). Accurate withholding is facilitated when there is little graduation of the tax rate schedule; minimal factors, other than labor income, that affect tax liability; workers with one job only; and an individual–based tax system.

As already mentioned, a common source of tension is the conditions under which an employer is required to withhold (and perform other duties) for someone who works for the company; in the United States, this is the distinction between an employee (for whom the answer is “yes”) and an independent contractor (for whom the answer is “no”). In 1979 the U.S. Treasury proposed a ten percent withholding on payments made in the course of business for services provided by certain independent contractors, including salespersons, but Congress failed to enact it.

Because the income of small businesses is particularly hard to measure and tax, in some countries there is withholding on business income, under which tax must be remitted in conjunction with certain purchases from these businesses, on the grounds that these payments presumably reflect or indicate taxable income of the recipient. In a few cases, there is “reverse” withholding, under which tax must be remitted in conjunction with certain purchases made by the taxpayer. Here the link to income is less direct, although arguably there is an indirect relation, if the transaction is expected to result in taxable profits, as when importers, wholesalers, or retailers, purchase goods for resale. Countries that require withholding on payments to certain businesses usually exclude as withholding agents individuals in their capacity as consumers because they are too numerous and otherwise not suitable as withholding agents. One important result of excluding individual consumers is that most retail establishments remain unaffected by withholding. There have been exceptions, though. At one time individual consumers in Japan and Australia were required to withhold and remit in certain cases, and in Japan individual consumers are required to withhold and remit for business income if they are withholding agents for employment income.

Withholding agents bear compliance costs, which are generally deductible in computing taxable income. In some situations the tax authority provides explicit compensation to withholding agents. Although not related to what is officially labeled as withholding, an instructive example is the practice of American states that offer “vendor discounts” to retailers who remit retail sales tax, often justified as recompense for the compliance costs incurred. If these discounts were simply a fixed fraction of sales tax remittances, then this would be equivalent to a reduction in the effective retail sales tax rate. In practice, the vendor discounts are often contingent on prompt remittance, so they serve as inducements for that behavior. The vendor discounts also are often capped per retailer, approximately mimicking the regressive nature of the compliance costs. There is, however, no obvious efficiency–based justification for providing a subsidy to offset the real resource costs of implementing a tax system. On the surface, such subsidies restore a competitive balance between small and large businesses, but this argument ignores the real differential in the social cost of compliance per dollar of revenue raised.

Some commentators have argued that an offset to compliance costs is the

37 For a formal treatment of this issue, see Slemrod (in process).
“cash–flow benefit” of withholding, which arises when the deadline for tax remittance comes after the tax–triggering event. In this case, the withholder gets the interest on the tax due for the period between the tax–triggering event and the remittance, which lowers the effective sales tax rate by an amount that depends on the remittance lag and the nominal interest rate. It is important to note that neither vendor discounts nor cash–flow benefits reduce the compliance costs of a tax system, which represent real resource costs of tax collection; rather they reduce the effective tax rate, often in a firm–specific manner.

Working in the other direction is the fact that many developed countries’ tax authorities refund as much as 20 percent of the gross revenue they receive, often without payment of interest. This is ubiquitous under a VAT, but also applies for example with respect to overwithholding that leads to refunds in the personal income tax, which is common in the United States. This, of course, amounts to an interest–free loan to the government, and has the effect of increasing the effective tax rate above the stated rate. This argument is complicated by the fact that, although the system defaults lead to over–withholding, these defaults can be adjusted by the taxpayer to eliminate over–withholding, and some argue that many taxpayers prefer the over–withholding in spite of the foregone interest because it provides a vehicle for forced saving.

When the employer is the government itself, the economics of withholding changes, in part because presumably the employer has no incentive to resist. Indeed, including labor income tax withheld by government agencies as part of tax collections can be misleading as an indicator of the transfer of resources from private to public hands, as long as the after–tax wage is set in a competitive market. In the absence of withholding, presumably the wage paid could be set lower; this would reduce recorded taxes and recorded wage payments by government, but would not affect net transfers.

As mentioned, half of OECD countries operate “exact withholding” systems, under which there is no need for any reconciliation at the end of the tax accounting period. To the extent that tax liability depends on information other than wages paid by any one employer—such as marital status, other income, or a second job—for this system to work this information needs to be conveyed to the withholder. This can be done directly from the employee to employer, as in the U.S. income tax, or indirectly, as in the U.K. income tax, where the employee conveys the information to the tax authority, which transmits a summary “code” to the withholding employer. Under “final withholding” systems, although in principle there should be reconciliation between withholding and actual tax liability, this is foregone to economize on compliance and administrative costs; in this case the withholding schedule is equivalent to the effective tax schedule.

38 These payments are largely attributable to zero rating of exports.
39 More than three–quarters of individual taxpayers receive refunds.
40 There may be other reasons. Taxpayers may be (perhaps irrationally) risk–averse to owing a penalty for insufficient remittances.
41 Taken to the extreme, one can see why Ickes and Slemrod (1992, p. 385) might assert that “from a tax administration standpoint, abandoning the planned economy is the worst possible direction.” They go on to say, “Withholding at source is greatly facilitated when the state is on one side of most economic transactions. In fact, the withholding system can be completely invisible. In a pure command economy taxes are implicit, and can be understood only as the difference between a family’s real purchasing power and some unobserved shadow value of their contribution to production.”
Withholding Noncompliance

Compared to the paramount importance of withholding, very little attention has been paid by economists to the enforcement of these payments. A number of interesting questions arise, however, such as which party is liable when the withholder does not remit, and how the penalty structure for noncompliance, including evasion, is set, including whether the penalties for withholding noncompliance should be symmetric to the penalties for noncompliance for the income recipient.

In the United States an employee receives credit for withheld tax even if it has not been remitted by the employer, and the government’s only recourse for unpaid amounts is against the employer. The statute covering this eventuality refers to the employers holding the withheld amounts in “a special fund in trust for the United States.” Any responsible person, typically a corporate officer, who willfully fails to remit withholding tax is subject to a 100 percent “trust fund” penalty; this is a collection device that is usually assessed only when the tax cannot be collected from the employer, and results in a personal liability not dischargeable by bankruptcy. There is also a four-tiered graduated penalty, with rates up to 15 percent of the amount of underpayment, which applies to the failure to make timely remittances. This compares to a 20 percent accuracy-related penalty that may be imposed on employee taxpayers for negligence, disregard of rules and regulations, or substantial understatement of tax, and a 75 percent penalty for any understatement that is due to fraud.

THE ROLE OF FIRMS

The discussion of withholding highlights the critical role of firms in the operation of modern tax systems. The impetus behind the central role of business in tax remittance has been elegantly stated by Richard Bird, who wrote: “The key to effective taxation is information, and the key to information in the modern economy is the corporation. The corporation is thus the modern fiscal state’s equivalent of the customs barrier at the border” (Bird, 2002). Collecting taxes from businesses makes use of the economy of scale of the tax authority dealing with a smaller number of larger units, many of which for other purposes have already developed sophisticated systems of record keeping and accounting. Firms are also central to information reporting, as they provide the information that facilitates checking whether the correct amount of tax liability has been collected from all the remitting parties.

The importance of firms in remittance, and therefore in tax systems, is noteworthy in light of the fact that, with some exceptions, firms are absent from the modern theory of optimal taxation.

43 Baird, Bris, and Zhu (2005) argue that withholding tax, as well as sales tax, obligations are key drivers in small business bankruptcies. They say (p. 20) that, “In the absence of these tax obligations, the owner–manager would be free to walk away from the business and start a new one afresh. The tax obligations prevent him from doing this. Indeed, small firms sometimes file for bankruptcy because of tax liabilities, and the center of their restructuring is focused on the bargaining between the owner–manager and the tax collector. The business may have little or no value as a going concern, but the owner–manager may seek Chapter 11 protection because it offers her the best way to deal with the tax collector and avoid personal liability for the unpaid taxes.”
44 Gordon and Li (2005) stress the importance of financial institutions to tax systems. In many Latin American countries, tax remittances are made to banks, who then forward the money to the government.
45 The Meade Committee report suggested that relying on firms for tax remittances might also lead to an insistence that the process be simple, and that “to leave to competing firms the task of tax collection may induce a healthy search for the most efficient methods of carrying out the operation.” (Meade, 1978, p. 20).
Their disappearance dates from the foundational models developed by Peter Diamond and James Mirrlees (1971) in which firms (generally endowed with constant–returns–to–scale production functions, which renders firm size indeterminate) are simply mechanical vehicles for combining productive inputs into output in cost–minimizing proportions. To helpfully analyze the consequences of taxation in a modern tax system in which firms dominate remittance, the theory of optimal taxation needs to be reconsidered within models that feature heterogeneous firms, both in terms of their production technology (that leads to heterogeneous size) and their propensity and ability to comply with the tax system.46

The central role of businesses in the tax remittance process can be blurred by the loose language often used to categorize which taxes are “business taxes.” The public positions of business associations reveal that they are often adamant about cutting—or, certainly, not increasing—“business taxes,” even though the ultimate incidence may bear little connection to this labeling.47 Clarifying the precise meaning of a “business tax,” and distinguishing between a legal responsibility to remit money to the government and bearing the burden of a tax, is especially important.

One measure of the central role of business in the U.S. tax system is provided by Christensen, Cline, and Neubig (2001), who calculate that in 1999 to all levels of government businesses “paid, collected, and remitted”—in the terminology suggested here “remitted”—83.8 percent of total taxes. Of the 83.8 percent, Christensen, Cline, and Neubig (2001) label 31.3 percent as “tax liability of business,” 8.1 percent as the “business as tax collector,” and 44.4 percent as “business as withholding agent.” The distinction is, however, not always clear.48 An analysis of British data in Slemrod, Whiting, and Shaw (2007) found that firms there remit 88 percent of all taxes, nearly the same percentage as Christensen, Cline, and Neubig (2001) calculated for the United States. At least in these two countries, firms dominate the tax remittance process, even though a relatively small fraction of total taxes are what is commonly referred to as business taxes.

BUSINESS–BASED CONSUMPTION TAXES

Implementation Issues: VAT versus RST

The two most prominent broad–base consumption taxes, the value added tax (VAT) and the retail sales tax (RST), are both business–based taxes, in that final consumers generally have no role in tax remittance.49 Economics textbooks treat the RST and VAT as two essentially equivalent ways to implement a consumption tax, differing “only” in how they are implemented. In “textbook” RST, tax is remitted only by retail businesses, triggered by sales to final consumers.50 Under a VAT, in principle tax is remitted

46 For a start on this see Dharmapala, Slemrod, and Wilson (2008).
47 On business associations’ positions on tax policy and reform, see Slemrod (2008).
48 For example, they assert that, for sales tax on final goods, businesses are responsible for “collecting and remitting” the tax, although the statutory liability lies with the consumer: “If the business does not collect the tax, the business still is responsible for remitting the tax liability.” They contrast businesses collecting tax with businesses withholding tax by saying: “Unlike the tax collection responsibility, if an insufficient amount is withheld, the employee is responsible for paying the difference.”
49 Compare a personal consumption tax, in which the remittance system operates much like an income tax but allows a deduction for net saving.
50 In contrast, actual retail sales taxes as operated by the U.S. states often collect considerable revenue triggered by business–to–business sales; see Ring (1989, 1999) for quantitative estimates of the importance of this.
by all businesses; under the common invoice–credit system, tax liability is triggered by sales and credits are triggered by purchases from taxpaying businesses.

At first blush, RST might look to be better on implementation grounds because the tax authority need deal with fewer (only retail) firms. But the RST raises two thorny implementation problems. The first problem is that on one side of all taxable transactions is a multitude of consumers, from which it is difficult to obtain useful corroborating information. Second, implementation requires distinguishing final sales to consumers from business–to–business sales. RSTs generally require businesses to register so that (in a pure RST) their purchases are not subject to tax, providing an incentive for consumers (or businesses making purchases intended for their principals’ consumption) to fraudulently register.

The VAT addresses both of these problems. Regarding the first problem, the sales at the retail stage are a relatively small fraction of total taxable sales, so that the difficulty of monitoring retail sales puts less revenue at risk.\textsuperscript{51} Second, the tax treatment of business sales does not depend on the identity (i.e., whether it is a firm or a final consumer) of the buyer. The VAT is triggered by all sales, without regard to the purchaser, relying on tax credits rather than on exemptions to offset the tax triggered by business purchases. In addition, most VATs are based on a “credit–invoice” system, under which business purchases from other businesses generate credits if the purchaser can produce an invoice certifying that the supplier paid tax on its sales; this feature, buttressed by cross–checking of information across firms, provides a certain self–enforcing, although certainly not perfect, aspect of a VAT. As discussed below, though, the VAT has its own enforcement problems.

**Exemptions**

Although under a pure VAT all businesses must in principle remit tax, most countries’ VATs allow businesses below a certain threshold amount of sales to not have to register, that is, to be “exempt” from VAT.\textsuperscript{52} In an invoice–credit VAT, the implications of this are somewhat subtle. First of all, exempting a firm that sells to either an exempt business or to final consumers reduces revenue, but exempting a firm that sells to non–exempt firms may actually increase tax revenue because the tax that has been remitted in conjunction with the non–labor purchases of the exempt firm are never credited against any firm’s tax liability. Indeed, many firms that would like to sell to non–exempt firms would be less profitable if exempt. In recognition of this fact, most countries that feature size–based firm exemptions allow voluntary registration for firms below the size threshold; this option would appeal to firms selling mostly to taxable firms that, ceteris paribus, prefer to buy from firms that are in the tax net so that they can claim a tax credit on their purchases (see Ebrill, Keen, Bodin, and Summers (2001)).\textsuperscript{53}

Understanding how exemption affects revenue explains the textbook equivalence of an RST and a VAT with no exemptions. Start with a VAT, and imagine randomly exempting pre–retail businesses. Some of these exemptions will increase revenue (if the firm sells predominantly to non–exempt firms) and some will reduce revenue (if the firm sells predominantly to

---

\textsuperscript{51} This implicitly assumes that retailers cannot easily claim refundable credits for purchases that exceed tax triggered by their (understated) sales to final consumers.

\textsuperscript{52} In many VAT systems some goods and services may also be exempt from taxation, which has a distinct meaning from firm exemption.

\textsuperscript{53} Note that this will also lead to VAT “chains,” in which firms out of the tax net tend to sell to other firms that are out of the tax net or to consumers, and firms in the tax net are likely to sell to firms in the tax net.
exempt firms). Ultimately, though, these revenue effects will cancel out, so that a VAT will raise the same revenue as an RST. In one sense, then, an RST is equivalent to a VAT with exemptions for all pre–retail firms. Although it is rarely described this way, one can consider that a VAT is equivalent to an RST with a system of withholding by pre–retail businesses; some of the tax due is remitted further up the production and distribution chain, but downstream firms can credit the tax remitted (i.e., withheld) by upstream businesses. Irrelevance propositions also apply to this issue; a VAT and an RST are equivalent if, but only if, one assumes away avoidance and evasion, as well as administrative and compliance costs.

Tax practitioners and scholars are familiar with exemptions in personal income taxes (based on taxable income, not sales) as a way to achieve progressivity of the tax burden and also limit the number of taxable entities the tax authority has to deal with. But exemptions for firms raise a number of distinct issues. One is the relative ease of sub–dividing businesses (as opposed to individuals or households) for tax advantage. This issue applies not only to exemptions, but more generally to graduated rate structures applied to businesses.

A second issue is how to characterize the economy’s equilibrium when some (say, small) firms receive preferential tax treatment. To put the question simply, why don’t firms that receive preferential tax treatment drive out the firms that do not? In a standard model of perfect competition, they would. When the preference is limited to small firms, the answer certainly involves the production technology that allows firms of different sizes to co–exist. The optimal exemption for small firms should trade off the administrative and compliance cost savings of the tax authority having to deal with a smaller number of firms against the production inefficiencies that arise from differentiating the treatment of firms.

Gaining insight into these questions is critical to understanding both the incidence (and, therefore, distributional) and the efficiency effects of tax systems that are based on firm remittance and which are imperfect (i.e., that feature both statutory exemptions and “do–it–yourself” exemptions—noncompliance). Gaining this insight will require bringing firms back into tax analysis, and thinking through the impact of taxation on industrial organization.

**VAT Withholding**

A number of countries, particularly in Latin America and Africa, have introduced as anti–avoidance measures remittance arrangements for VAT known as “withholding.” These arrangements obligate certain entities registered for VAT to remit to the tax authority amounts of money related to payments they make to their suppliers. In some countries this obligation is restricted to specific classes of taxpayer (e.g., government agencies and/or large taxpayers), specific transactions (e.g., government contracts), or certain hard–to–tax economic sectors. In other countries the obligation is imposed

---

54 Krauss and Bird (1971, p. 1167) characterize the difference between a consumption VAT and a retail sales tax as a difference in *method* (italics in original) rather than in essence, but do not characterize the pre–retail part of the VAT as a *withholding* tax.

55 This is not to imply that it would be easy for a huge public company like General Motors to divide into tens of thousands of small firms to take full advantage of a graduated corporation tax structure.

56 In contrast, neither RSTs nor income tax withholding requirements generally have registration thresholds.

57 For a model of this see Keen and Mintz (2004). Their model does not, though, consider that firm exemptions may distort the organizational structure and size of firms. For a model that addresses that issue, see Dharmapala, Slemrod, and Wilson (2008).
on a wide range of taxpayers and transactions.

According to the International Monetary Fund (2006, p. 1), proponents argue that VAT withholding increases collections, “since unregistered traders are forced to pay at least some VAT” (emphasis added to clarify that here “pay” means “bear the burden of” rather than “remit”). In addition, proponents say that it forces these traders into the VAT system and, as a result of reporting requirements imposed on withholding, their identity is exposed to the tax authorities.

Imports, Exports, and Trade–Related Fraud in VATs

Most VAT systems are destination–based systems, in that they intend to tax consumption that takes place within the jurisdiction, rather than production (for ultimate consumption) that takes place within the jurisdiction. This intent is generally implemented by rebating (“zero–rating”) the tax that would otherwise be due on sales that are exported, and by effectively not considering imports as creditable purchases. Indeed, imports are generally taxed at the border in the same way as dutiable imports, and often the tax collections from imports dominate VAT collections. Indeed, Ebrill et al. (2001) calculate that, for 22 mostly developing countries, the VAT collected on imports averages 55 percent of VAT revenues, and is as high as 70 percent.

The importance of the remittance system for the relationship between a VAT and an RST is nicely illustrated by some of the suggested policy responses to the rising problem of “carousel” fraud in the VAT. Carousel fraud takes advantage of the destination basis of the VAT and its implementation through the combination of no crediting of imported business purchases and zero–rating of exports, making exporters generally eligible for net refunds from the tax authority. To see how carousel fraud works, consider the highly simplified example of a pure entrepot transaction, where a good is imported and immediately exported with no value added. If the importing and exporting are done by a single firm, no VAT is due because the export sale does not trigger tax liability and the imported inputs do not generate tax credits. But now imagine that the importer and exporter are two different businesses (and there is no transfer pricing manipulation); in this case, the importer would have net tax liability equal to the tax rate times the value of the goods, and the exporter would be eligible to receive an (offsetting) net refund. Under the carousel fraud schemes, the exporter receives the refund and the importer “disappears” before the tax due is remitted, resulting in a net transfer from the government to the conspiring importer and exporter; in practice, the schemes are more complicated, often with sometimes unwitting businesses interspersed in the distribution chain between the importer and exporter.

One proposed policy response to carousel fraud is called “reverse charging,” in which “liability [in a business–to–business transaction] is placed on the purchaser rather than the seller” (Keen (2007, p. 378); see also Keen and Smith (2006)). In the stylized entrepot example above, neither the importer nor the exporter would have a remittance responsibility, as would be true under a pure RST. “Reverse withholding” of VAT, discussed briefly above, would address the issue in an essentially similar way, by requiring that a purchaser in a business–to–business transaction make a remittance to the tax authority triggered by its purchases. The apparent difference compared to reverse charging is that the seller would technically remain liable for output VAT, receiving a credit for the amount “withheld” by the purchaser. If an exporting business purchaser was required to withhold all of the VAT, it would eliminate its eligibility for a refund.
entirely. This system would require that the tax authority ensure that the seller receive credit for the withholding by the purchaser only when the remittance actually took place. But, in a perfectly working system, complete reverse withholding would eliminate the refund for an exporter, and move the system toward the remittance pattern of an RST. This discussion highlights that the terminology using the terms “reverse” and “withholding” should not mask what are essentially alternative ways to assign the remittance responsibility of a tax on consumption.

**Sales and Use Tax**

The importance of the remittance system is well illustrated by the policy controversy regarding U.S. state retail sales tax and their complementary use taxes levied on in–state use of out–of–state purchases. The in–state purchaser of the out–of–state goods is supposed to remit the use tax, but the compliance rate for individual purchasers is notoriously low, indeed approaching zero, and some states insert a line on their income tax form reminding the taxpayer of the use tax liability and facilitating remittance of the use tax as part of the income tax remittance process, bypassing the usually unknown use tax form. Remittance by the out–of–state vendor is a more promising collection strategy, but in the National Bellas Hess case the Supreme Court ruled that, because of the complexity of such vendors having to deal with variations in both rate and base of 45 states and over 7,000 localities that levy sales tax, requiring vendor remittance would be too onerous and, therefore, violate the Commerce Clause of the U.S. Constitution that forbids states from enacting measures that substantially affect interstate commerce.

Similar issues arise in excise taxes levied on, for example, gasoline, diesel fuel, alcohol, and tobacco products, where the remittance possibilities range widely, but in practice the taxes must generally be remitted by businesses that are upstream in the distribution process: importers, refiners of motor fuels, brewers, distillers, and tobacco companies, and wholesalers. Tax administrators tend to believe that relying on retailers (let alone final purchasers) to remit these taxes would be much more costly in terms of administrative cost per dollar collected, and much less effective in terms of high evasion per dollar collected. The case of commercial fuel raises particularly interesting examples of non–replicability (i.e., how the remittance system may constrain the assignment of liabilities), because although the point of collection is upstream, the ultimate liability cannot be determined until later in the distribution chain. In the case of the federal diesel fuel tax, the ultimate liability depends on the nature of the end use, with home heating being exempt and use as truck fuel taxed, a situation which led, in the Omnibus Budget Reconciliation Act of 1993, to a requirement that tax–exempt diesel fuel be dyed so as to more easily distinguish it from taxed fuel in the possession of users.58 State fuel taxes are remitted at the point of purchase by truckers, but the ultimate total liability, and which jurisdiction receives the remittances, depends on the number of miles driven in each state. The reconciliation of remittances and receipts is coordinated through a system known as the International Fuel Tax Agreement, discussed in Denison and Facer (2005).

---

58 OBRA93 also moved the point of collection from the distributor to the terminal level, on the grounds that it “would reduce the number of times the fuel changes ownership prior to tax and reduce the number of taxpayers, so that taxes would be easier to collect, and payments of tax would be easier to monitor” (Beerbower, 1995).
THE POLITICS OF TAX REMITTANCE

The politics of income tax withholding in the United States are fascinating because, although tax administrators swear by it, many American conservatives dislike it because it reduces the visibility of tax collection and, thus, reduces the perceived cost of government. They prefer that people write checks—remit taxes—every week or every month. Some conservative legislators have introduced into Congress a bill entitled the “Cost of Government Awareness Act,” which would eliminate withholding and instead require individuals to remit income taxes in monthly installments. The version of the Hall–Rabushka flat tax introduced as a bill by then–House Majority Leader Dick Armey eliminated employer withholding, which Armey referred to as a “deceptive device that has made big government possible.” “Only by taking people’s money before they ever see it,” Armey wrote, “has the government been able to raise taxes to their current height without sparking a revolt” (Armey, 1996, p. 99).

To many of these same conservatives the value added tax is anathema, demonized as the hidden money machine that fuels European welfare states. But it is more than visibility that is at issue here and more than privacy, as well. There is, indeed, a profound dilemma for conservatives. Do they want the cost of government to be visible or do they want it to be high? Having taxpayers regularly write checks to the government may increase the visibility of the tax burden, but it is also highly inefficient. One might argue for inefficiency, if the political process is dysfunctionally biased toward overspending. Some American conservatives point to high spending in European countries with a VAT, and see the VAT as one cause of bigger European governments. The empirical evidence on this question is, however, mixed.

CONCLUSIONS

Who remits tax is an important aspect of implementing a tax system, in spite of standard economic analysis that maintains that which side of a taxed market remits a given amount of tax liability is completely irrelevant for the consequences—incidence, allocation, and efficiency—of taxation. In the presence of avoidance and evasion (i.e., in all real tax systems), the cost of administration and enforcement varies depending on the identity of the remitter. This can occur for two distinct reasons. The first is that the total resource costs of administering a given effective tax structure may vary depending on the remittance system. For example, it may be less costly (considering both administrative and compliance costs) to monitor one employer’s tax remittances.

---

59 H.R. 1364, 107th Cong.
60 See Becker and Mulligan (2003) for a formal statement of this view.
61 See, for example, Stockfisch (1985). The report of the President’s Advisory Panel on Federal Tax Reform (2005, p. 192), which could not reach a consensus on whether to recommend a VAT option, noted that some panel members were concerned that introducing a VAT would be a “money machine,” i.e., lead to “higher tax collections over time and facilitate the development of a larger federal government.” Keen and Lockwood (2006), using panel data for the OECD, find evidence that countries with a VAT raise relatively more revenue than those without.
62 Dusek (2003) exploits the fact that withholding for state income taxes was adopted at different times by different states, and finds that governments did respond to this administrative efficiency, but the major response was a shift in the composition of revenues toward the income tax rather than an increase in total revenues. On average, adopting withholding was followed by an eight percent increase in total tax revenues, but most of this increase was due to the higher demand for government that stimulated the adoption, not the fact that more efficient taxes lead to bigger government.
as opposed to thousands of employees. Thus, although the statutory tax schedule is unchanged, the true total tax burden does vary across remittance regimes. In this case, although the remittance pattern of a given total burden is irrelevant, the remittance system affects the total burden of a given statutory tax.

Second, the opportunities for avoidance and evasion and the technology of the enforcement mechanism affect not only the total cost of a given statutory system but also the incentive to demand and supply the taxed activity. If this effect is not symmetric across the identity of the remitter, then changing the remittance system will affect the equilibrium price and quantity of the taxed activity and, thus, have incidence, efficiency, and allocation effects.

Firms are central to modern tax remittance systems. For this reason, rigorously applying economic analysis to these issues will require developing a theoretical framework that features heterogeneous firms and a collection technology with fixed cost elements so that, other things equal, it is more cost–effective to deal with firms rather than employees or consumers and with large firms rather than small firms. For such a framework to have policy implications, an empirical agenda needs to address the effect of tax policy on industrial organization and the efficiency of production.

Practitioners who work with real tax systems are well aware of these issues, and pay substantial attention to who remits tax monies in addition to what triggers tax liability. Vigorous theoretical attention could, hopefully, sharpen their analysis of alternative policies.

Acknowledgments

For comments on an earlier draft I am grateful to Amaresh Bagchi, Richard Bird, Dan Halperin, James Hines, Richard Highfield, Louis Kaplow, Henrik Kleven, Wojciech Kopczuk, Leandra Lederman, Kyle Logue, Alex Raskolnikov, Reed Shuldiner, Shlomo Yitzhaki, two exceptionally thoughtful referees, and participants in seminars at Harvard Law School, the University of Pennsylvania Law School, Northwestern Law School, the Harvard/MIT Public Economics workshop, and the University of Michigan public finance lunch seminar. Errors and misstatements are, though, solely my responsibility.

REFERENCES


Dalton, Hugh.  

Denison, Dwight, and Rex L. Facer II.  
"Interstate Tax Coordination: Lessons from the International Fuel Tax Agreement."  
*National Tax Journal* 58 No. 3 (September, 2005): 591–604.

Dharmapala, Dhammika, Joel Slemrod, and John D. Wilson.  
"Taxation and the Missing Middle: Optimal Tax Remittance with Firm–Level Administrative Costs."  

Diamond, Peter, and James Mirrlees.  
"Optimal Taxation and Public Production. I—Production Efficiency."  

Doherty, Brian.  
"Best of Both Worlds."  
*Reason* (June, 1995).

Douglas, Roy.  

Dusek, Libor.  
"Do Governments Grow When They Become More Efficient? Evidence from Tax Withholding."  

Ebrill, Liam, Michael Keen, Jean–Paul Bodin, and Victoria Summers.  

Gordon, Roger, and Wei Li.  
"Tax Structure in Developing Countries: Many Puzzles and a Possible Explanation."  

Hill, Joseph A.  
"The Civil War Income Tax."  

Ickes, Barry, and Joel Slemrod.  
"Tax Implementation Issues in the Transition from a Planned Economy."  

International Monetary Fund.  

Keen, Michael.  
"VAT Attacks!"  

Keen, Michael, and Ben Lockwood.  
"Is the VAT a Money Machine?"  

Keen, Michael, and Jack Mintz.  
"The Optimal Threshold for a Value–Added Tax."  

Keen, Michael, and Steven Smith.  
"VAT Fraud and Evasion: What Do We Know and What Can Be Done?"  

Kotlikoff, Laurence, and Lawrence Summers.  

Krauss, Mel, and Richard M. Bird.  
"The Value Added Tax: Critique of a Review."  
*Journal of Economic Literature* 9 No. 4 (December, 1971): 1167–73.

Lent, George A.  
*The Corporate Franchise as a Basis for Taxation.* Austin: University of Texas Press, 1942.

McCulloch, John R.  

Meade, James E.  

The President’s Advisory Panel on Federal Tax Reform.

Ring, Raymond J., Jr.

Ring, Raymond J., Jr.

Rosen, Harvey.

Scott, James C.

Slemrod, Joel.

Slemrod, Joel.

Slemrod, Joel.

Slemrod, Joel.

Slemrod, Joel, John Whiting, and Jonathan Shaw.

Slemrod, Joel, and ShlomoYitzhaki.

Soos, Piroska.

Stockfish, Jacob A.

Tanzi, Vito.