

Working Paper

Insights From a Tax-Systems Perspective

Joel Slemrod

Stephen M. Ross School of Business
at the University of Michigan

Christian Gillitzer

Stephen M. Ross School of Business
at the University of Michigan

Ross School of Business Working Paper Series

Working Paper No. 1206

October 2013

This paper can be downloaded without charge from the
Social Sciences Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=2340048>

Insights From a Tax-Systems Perspective*

Joel Slemrod and Christian Gillitzer

3 October, 2013

Keywords: tax systems, evasion, avoidance, collection costs.

JEL codes: H20, H21, H24-H26.

Abstract

The tax-systems perspective considers a variety of costs and behavioral margins often ignored in standard tax analysis: administrative and compliance costs, evasion and avoidance behavior, and multiple non-rate tax-system instruments (e.g., withholding and public disclosure). We show how the standard optimal tax framework can be augmented to include these new sources of cost and behavioral response by considering some enduring tax policy questions: What is the optimal commodity tax base breadth? How does enforcement effort targeted to avoidance behavior affect optimal progressivity? What fraction of returns should be audited? Should small firms be excluded from a tax system?

*This article is an edited extract from the forthcoming book *Tax Systems* by Joel Slemrod and Christian Gillitzer, published by The MIT Press in 2014. Joel Slemrod delivered the Richard Musgrave Lecture 2013, in Munich on April 11, based on this material.

1 Introduction and Motivation

1.1 What is a Tax System

A tax system¹ is a set of rules, regulations, and procedures that 1) defines what events or states of the world trigger tax liability (tax bases and rates), 2) specifies who or what entity must remit that tax and when (remittance rules), and 3) details procedures for ensuring compliance, including information-reporting requirements and the consequences (including penalties) of not remitting the legal liability in a timely fashion (enforcement rules).

Much of modern economic analysis of taxation, and certainly the seminal contributions, presumes that tax liability can be ascertained and collected costlessly, in which case 2) is irrelevant and 3) is unnecessary. But this way of modeling taxation misses much that is important about taxes. It cannot address many current tax policy issues—should Greece raise revenue to meet its bailout conditions by raising tax rates, or by cracking down on tax evasion?—providing one reason why there is often a disconnect between topical tax issues of the day and the economic theory of taxation. And, in our view, it misses much of what is intellectually fascinating about taxes.

This paper is motivated by the aspects of reality that the standard model ignores. Some of these blind spots are:

1. Although some individuals may remit their tax liability dutifully, others view compliance as a tactical decision and will evade their liabilities if the odds of success seem favorable.
2. Taxpayers will re-arrange their affairs to legally reduce their tax liability, including efforts to reduce their tax liability without altering their real activities, which we will refer to as avoidance.
3. Limiting avoidance and evasion is costly, and tax authorities have limited administrative resources.

¹This name is inspired by the title of Richard Musgrave's 1969 book *Fiscal Systems*, which addresses how fiscal institutions and functions change with their environment, and what similarities remain even though the setting differs.

4. Tax authorities have limited policy instruments.
5. Taxpayers (and tax policy makers) have cognitive limitations.
6. The world is complex, complicating the collection of non-capricious taxes. Some of the complication is manufactured by taxpayers to obfuscate their affairs, and some exists because the tax system is used to achieve specific social goals in addition to raising revenue.

These issues are especially critical in developing countries, so much so that a former IMF official² once opined that, in developing countries, “Tax administration *is* tax policy.” But they are also critical in developed countries, where the operative issue is generally not the feasibility of certain taxes, but rather the comparison of alternative imperfect tax structures. For example, while in many developing countries an income tax that relies on self-reporting cannot be administered with tolerable cost and equity, in a developed country the question is to what extent tax design should reflect the reality of evasion, the necessity of enforcement, and the costs of collection. In addition, an important set of generic aspects of income tax structure, such as the absence of taxation of imputed rents from consumer durables, taxation of capital gains (if at all) on a realization basis, and pre-set depreciation schedules, are undoubtedly largely driven by practical concerns of administrability. For these reasons, consideration of evasion, avoidance, and administration is essential to the positive and normative analysis of taxation in *all* jurisdictions.

The remainder of this essay is organized as follows. Section 2 discusses the properties of administrative and compliance cost, and their consequences for tax system design. In Section 3 we analyze taxpayer efforts to reduce tax liability, by both legal means (avoidance) and illegal means (evasion). Several new tax-system instruments that can help the tax authorities limit evasion and avoidance behavior are introduced in Section 4: withholding and information reporting, market transactions, and public disclosure. Section 5 explains the assumptions under which the elasticity of taxable income (ETI) serves as a sufficient statistic for the marginal welfare cost of *all* behavioral response to taxes, discusses issues in empirically estimating its magnitude, and summarizes findings in the empirical ETI literature. We switch gears in Section 6 from positive to normative

²Casanegra de Jantscher (1990, p. 179).

analysis, studying three central questions in tax-system design: i) the optimal commodity tax base breadth, ii) the optimal level of audit enforcement (hint: don't equate marginal audit revenue with the marginal administrative cost of audits), and iii) the optimal degree of tax-system progressivity when the behavioral response to taxes can be controlled, at some resource cost, by varying tax base breadth. Section 7 discusses the role of firms in real-world tax systems; firms serve a vital role as remitting, information reporting, and withholding agents, but monitoring their activity creates administrative cost, and it can be optimal to exclude small firms—either implicitly or explicitly—from the tax system. Section 8 addresses lines and notches, discontinuities in characteristic space and budget sets, their consequences for tax system design, such as the optimal degree of differentiation in commodity tax rates, and their use in empirical work as a source of identification for behavioral elasticities. Some concluding thoughts are offered in Section 9.

2 Multiple Sources of Cost

2.1 Administrative Cost

Even if all taxpayers were scrupulously honest, any tax system requires an administrative system with a bureaucracy to calculate tax liabilities and to record and check remittance. But, not all taxpayers are honest, nor are taxpayers obliged to arrange their affairs in a way that suits the tax authority. As a result, some taxpayers go to considerable lengths to reduce the size of their tax bill. In response to avoidance and evasion, all tax authorities judge it worthwhile to expend resources to limit revenue loss as well as any undesirable efficiency, horizontal or vertical equity consequences. Moreover, it will always be relatively cheap to collect taxes in a capricious way—without measurement or verification—so capriciousness must be balanced by the value of horizontal equity or legitimacy.

For any given objective, there are more and less effective ways for a tax administration to operate. For example, what is the optimal use of computers and information technology? Should a tax administration be organized by tax levy (e.g., corporate tax versus value-added tax), or by tax-

payer segment (e.g., corporations versus high-income individuals)? How should it be organized to minimize corruption? These questions have been addressed extensively but informally, especially in a developing-country context in, e.g., Bird (1983), and are certainly context-specific.³

The administrative cost of obtaining information is a function of the physical size, tangibility/visibility and the mobility of the tax base (e.g., it is harder to tax diamonds than windows), whether there is a registration of the tax base (e.g., owners of cars, holders of drivers' licenses), the number of taxpayer units, and the extent of information sharing with other agencies, both non-tax agencies within a jurisdiction and tax authorities in other jurisdictions. Administrative cost is also an increasing function of the complexity and lack of clarity of the tax law, *ceteris paribus*.

Administrative costs possess two additional properties that complicate the formal modeling of tax system issues: they tend to be discontinuous and to have decreasing average costs with respect to the tax rate. To see the first property, consider two commodity tax rates, denoted by t_1 and t_2 . If $t_1 = t_2$, then only the total sales of the two commodities need be reported and monitored. If, however, the two rates differ even slightly, then the sales of the two commodities must be reported separately, approximately doubling the required flow of information. This undoubtedly explains why, in contrast to optimal commodity tax theory that suggests different tax rates on each commodity, real tax systems feature a very small number of tax rates. Second, there are decreasing average costs because the cost of inspecting a tax base does not depend on the tax rate (except to the extent that people may be more inclined to cheat with a higher tax rate). Hence, a higher tax rate reduces the administrative cost per dollar of revenue collected (Sandford, 1973). Administrative cost may also be a function of the combination of the taxes employed and their rates, because the collection of information concerning one tax may facilitate the collection of another tax (e.g., inspection of VAT receipts may facilitate the collection of income tax). Note, finally, that in cases of negative marginal and average tax rates, such as arise under the earned income tax credit, the administrative problem is of a different nature, as the evading taxpayer's incentive is to overstate, rather than understate, income.

Administrative costs have implications for the optimal boundary between firms. Coase (1937)

³A good description of the properties of administrative cost can be found in Shoup et al. (1937, pp. 337-351).

argued that (in the absence of externalities) the equilibrium boundary is optimal because firms consider all relevant costs and benefits of expanding their size. But Dharmapala et al. (2011) show that administrative costs are akin to an externality, because firm size, and industrial organization more generally, will affect the cost of administering a tax system at a given effectiveness, in part because they affect the availability of more easily observable arms-length transactions. They show that when administrative costs are fixed per firm, if a firm is taxed then it should face a fixed fee—akin to a Pigouvian tax—equal to the administrative cost of taxing it.

2.2 Compliance Cost

Compliance costs, in contrast to administrative costs, are incurred in the tax remittance and collection process directly by taxpayers and by third parties (such as employers who are required to remit tax on behalf of their employees and provide information to the tax authority). We say “directly” for two reasons: 1) the burden of the compliance costs may be shifted away from the party that expends the resources to comply, just as an explicit tax may be shifted, due to changes in relative prices, and 2) administrative costs also ultimately burden taxpayers, but show up in the first instance as government expenditures, not as monetary or time costs to taxpayers, even though ultimately citizens/taxpayers bear these costs.

Measuring the extent and nature of compliance costs is not, as in the case of administrative costs, as straightforward as culling the relevant data from government budgets. Surveys are the most commonly used method to measure compliance costs, although low response rates can impair the reliability of their findings. While re-weighting survey results based on observable demographics is often helpful, unobserved heterogeneity between respondents and non-respondents is likely to be important. The direction of nonrandom underreporting bias is generally unclear, so survey results may not even accurately bound compliance costs. While respondents may see compliance as a “vexatious cost” (Tait, 1988, p. 352) and overstate their costs in an attempt to influence policy, it is also possible that those who find responding to surveys most burdensome also find tax-system compliance particularly costly (Sandford, 1995).

Regardless of these methodological difficulties, one fact stands out among all studies of all taxes in all countries: compliance costs dwarf administrative costs. For example, Slemrod (1996) estimates that, for the U.S. income tax, the private compliance cost is about 10 cents per dollar of revenue collected, compared to an administrative cost of about 0.6 cents per dollar collected for all the taxes the IRS administers. The value of time spent greatly exceeds taxpayers' out-of-pocket expenditures in complying with the U.S. income tax system. Based on a survey of Minnesota taxpayers for the 1982 tax year, Slemrod and Sorum (1984) estimate that a majority of time spent by taxpayers complying with the income tax system is due to record keeping, followed by return preparation, research, and interacting with a tax advisor.

Some of what is correctly measured as compliance cost is an unavoidable cost of complying with the law, and some of it is voluntarily undertaken in an effort to reduce tax liability. In either case it approximately represents resource costs to society because otherwise-productive resources are diverted to compliance activity. That taxpayers will voluntarily spend time on compliance in an effort to reduce their tax liability does not make it any less of a social cost; efficiency cost arises because taxpayers are encouraged by the tax system to engage in avoidance activity they would not engage in were it not for taxes.

To some extent administrative costs and compliance costs are substitutes, in the sense that either the government or the taxpayer may have the lead role in collecting key information. As an example, consider when is it optimal to delegate to employers the authority to remit taxes on behalf of, and convey information about, employees, thus allowing the administration to audit both the taxpayer agent and the taxpayer himself, and when it is optimal to deal only with the employee. Clearly, given that the employer already has the necessary information, it would save administrative costs to require him to pass it along to the tax administrator. This might also reduce total social costs if the cost of gathering information by the administration is higher than the increase in cost caused by imposing a two-stage gathering system. Note that a withholding system requires two information gathering systems and might generate incentives for the withholding agent to evade the taxes it collects, or to collaborate with withholders in withholding less than required (Yaniv 1988, 1992).

However, the potential efficiency of involving taxpayers in the administrative process must be tempered with a practical consideration. Administrative costs must pass through a budgeting process, while compliance costs are hidden. Hence, there may be a tendency to view a policy that reduces administrative cost at the expense of an equal (or greater) increase in compliance costs as a decrease in social cost, because it results in a decrease in government expenditures. In addition, administrative costs should be weighted higher than compliance costs because they are funded by tax revenue raised through distortionary taxes, for which the marginal cost of funds exceeds unity.

3 Evasion and Avoidance

3.1 Evasion

In the presence of tax evasion we need to enrich both the positive model of taxpayer decision-making and the normative model of optimal government behavior. The natural starting point for the former is to consider the costs and benefits of tax evasion. And, indeed, the standard framework for considering an individual's choice of whether and how much to evade taxes is a deterrence model. This was first formulated in the context of a flat income tax by Allingham and Sandmo (1972), who adapted Becker's (1968) model of criminal behavior to the economics of tax evasion. In this model, which we will refer to as the A-S model, a risk-averse taxpayer decides whether and how much to evade taxes in the same way she would approach any risky decision or gamble—by maximizing expected utility.

In the basic A-S model, labor income, wl , is held fixed; the risk-averse taxpayer chooses only what income to report to the tax authority. The taxpayer chooses the amount of understatement, e , to maximize expected utility:

$$\max_e (1 - p) u(wl(1 - t) + te) + pu(wl(1 - t) - \pi e), \quad (1)$$

where von Neumann-Morgenstern utility, u , is concave, p is the probability that evasion is detected and triggers a penalty of π times the taxable income understatement, $wl(1 - t)$ is the taxpayer's

true after-tax income, and t is the tax rate. Each dollar of taxable income understatement offers a payoff of t with probability $(1 - p)$, along with a penalty of π with probability p . If and only if the expected payoff to this gamble, $(1 - p)t - p\pi$, is positive, a risk-averse taxpayer will engage in some evasion, with the amount depending on the expected payoff and the taxpayer's risk preferences; more risk-averse individuals will, *ceteris paribus*, evade less. Increases in either p or π decrease evasion. As pointed out first by Yitzhaki (1974), in many tax systems the penalty depends on the tax understatement, rather than the income understatement as stated in Equation (1). Under the Yitzhaki formulation, increasing t has a negative income effect, but no substitution effect, making, *ceteris paribus*, a reduction in evasion optimal if risk-aversion is decreasing in income. To most, this is a profoundly unintuitive notion, one that has undoubtedly inspired further theoretical development and empirical analysis.

Some social scientists have argued that the A-S deterrence framework misses important elements of the tax evasion decision, and question some of its central assumptions, including that (i) nothing *per se* about its illegality matters and (ii) everyone acts as a free rider, so that there is no issue of intrinsic willingness to pay, or "tax morale." Some have gone further to suggest that, in thinking about tax evasion, it is necessary to abandon the standard expected utility maximization model and incorporate "behavioral" considerations. Much of the evidence related to these non-standard behaviors comes from how people react to other people, as in lab experiments. But the psychological attitudes of individuals toward government might be fundamentally different than their attitudes toward other people, or even other organizations. Individuals might feel more dutiful and even obedient toward government.

Neither type of theory provides clear quantitative predictions about the determinants of evasion, and so a large empirical literature has appeared, albeit one with many challenges. The threat of punishment and social shame make taxpayers unwilling to respond accurately to surveys, and results from randomized audits are typically unavailable to researchers, so a range of indirect methodologies have been pursued in the literature. What Slemrod and Weber (2012) dub the "traces-of-income" approach makes progress by looking for a variable correlated with true income, allowing the researcher to predict true income and back out the extent of evasion by comparing

the prediction to taxpayers' reports. Pissarides and Weber (1989), who pioneered this approach, used U.K. Family Expenditure Survey data on food consumption to estimate the extent of evasion. Assuming that only the self-employed evade, and that the relationship between food consumption and true income is independent of employment status, they are able to predict true income—and therefore underreporting—for the self-employed survey respondents. Assuming income reports in the survey match those given to the tax authority, they estimate that self-employed people in the United Kingdom on average underreported their income by about one-third. Feldman and Slemrod (2007) follow a similar approach, but avoid the need to use survey data by instead using charitable giving as reported on income tax returns relative to reported income as the trace of evasion. They find that, other things equal, reported positive self-employment income of \$1 is associated with the same level of contributions as \$1.54 of wage and salary income, which implies—assuming a negligible wage and salary noncompliance rate and that the self-employed are not inherently more charitable than others—a self-employment noncompliance rate of 35 percent; for positive farm net income, the implied noncompliance rate is 74 percent. Intriguingly, negative reported values for self-employment income are also associated with more contributions than reported by taxpayers with no self-employment income, suggesting that on average these reported losses are associated with *higher* true incomes.

In contrast to indirect means of inquiry, randomized field experiments can offer compelling evidence on the causal impact of particular policy interventions. In one of the first examples of applying this method to tax evasion research, Slemrod et al. (2001) conducted an experiment in Minnesota, sending a treatment group of taxpayers a letter warning them that their tax returns would be “closely examined.” Kleven et al. (2011) have since conducted a similar, but more comprehensive, study in Denmark. They conclude that threat-of-audit letters (and prior audits) have a significant effect on self-reported (but not third-party-reported) income.

In a field experiment in Austria, Fellner et al. (In Press) use a randomized design to test the effect on compliance with Austrian television and radio licensing fees of various mailings to potential non-compliers. All mailings asked recipients to clarify within 14 days why no payment had been received, with the letters differing in their emphasis of either the threat of detection and sanction,

a moral appeal equating compliance with fairness, and a third variant providing social information on the overall high level of compliance. Those receiving any type of mailing were significantly more likely to make a payment within 50 days of receiving the letter, but only the variant emphasizing the threat of punishment induced an additional increase in compliance. Fellner et al. interpret the generic effect of the mailing as an “alert effect” signaling that non-payment had been noticed, with the consequences of noncompliance amplified by the threat variant.

Field evidence on Chilean firms’ compliance with the VAT highlights the connection between information reports received by the tax authority and levels of evasion. Because firms can only claim tax credits for inputs bought from tax-compliant suppliers, the invoice-credit VAT system has a built-in (albeit imperfect) self-enforcement mechanism. Firms purchasing inputs would like to overstate purchase costs to inflate tax rebates, but sellers need to understate sales proceeds to minimize VAT liability. Because these incentives conflict and—except for final sales to consumers—information reports are made by both parties to each transaction to the tax authority, the VAT system is believed to dramatically increase the probability of evasion being detected. Pomeranz (2011) tests this hypothesis by mailing increased audit threat letters to over 100,000 randomly selected Chilean firms, using a sample of over 300,000 firms receiving no letter as the control group. Consistent with theoretical predictions on the self-enforcement mechanism, the increase in VAT receipts (and therefore the level of inferred evasion) induced by the letters is concentrated at the level of sales from firms to final consumers, for which there is no paper trail, suggesting that for an individual firm, information reporting acts as a substitute for audit risk. In line with the findings of Fellner et al. and Blumenthal et al. (2001) from the Minnesota randomized tax experiment, Pomeranz (2011) found that a mailing appealing to tax morale but promising no increased enforcement had little effect on VAT payments.

3.2 Avoidance

Economists generally differentiate between illegal means of reducing tax liability, referred to as evasion, and legal means of reducing tax liability, referred to as avoidance. In some contexts the law is unclear and neither the taxpayer nor the tax authority, or both, may not know for

sure which is which. But, more importantly, this definition of avoidance does not correspond to another common usage among economists, in which the term avoidance does not include “real” behavioral responses to changing tax rates such as reduced labor supply, altering the time path of consumption, or consuming less highly-taxed cigarettes, but rather to a different class of behaviors. If not all legal behavioral response to tax rates is avoidance, how can it be usefully defined and distinguished from real behavioral response? A good starting point is the definition of avoidance akin to the one offered in Slemrod and Yitzhaki (2002): taxpayer efforts to reduce their tax liability that do not alter their consumption basket other than due to income effects. This definition covers a broad range of behaviors. It includes paying a tax professional to alert one to the tax deductibility of activities already undertaken. It covers changing the legal form of a given behavior, such as reorganizing a business from a C corporation to an S corporation, recharacterizing ordinary income as capital gain, or renaming a consumer loan as a home equity loan. It covers tax arbitrage, where economically equivalent, but differentially taxed, positions are held simultaneously long and short, producing tax savings. Finally, it covers (slightly) retiming a transaction to alter the tax year it falls under, or slightly re-engineering a vehicle to change its tax classification.

The distinction between real behavioral response and avoidance does not arise in, for example, the standard model of optimal commodity taxation, where consumption, and only consumption, triggers tax liability, so the model does not allow any other behavior that could reduce tax liability. The same is true of the standard optimal income tax model, where individuals’ only decision is how much labor to supply, which affects tax liability by changing taxable income.

However, in many cases what triggers tax liability in actual tax systems is different than what triggers tax liability in stylized models. Retail purchases rather than consumption trigger retail sales tax liability, receipt of labor income rather than the exertion of physical or mental labor itself generally determines the timing of tax liability. Sales of appreciated capital assets trigger tax liability rather than accrual of gain or consumption itself. Operational definitions of taxable income differ on many dimensions from the Haig-Simons definition of income: consumption plus the change in net worth. As an example, in the U.S. taxable income does not include employer contributions to employee health insurance, whereas Haig-Simons income would. These give rise

to avoidance.

3.3 The Relationship Among Evasion, Avoidance, and Real Activity

The relationship between real activity on the one hand and evasion and/or avoidance is of note. To illustrate this relationship, consider the example studied in Grubert and Slemrod (1998), who provide empirical evidence of such an interaction for corporate income shifting to Puerto Rico. The fact that corporate income earned by U.S. firms in Puerto Rico was essentially exempt from U.S. corporate income taxation provided strong incentives for U.S. corporations to use transfer pricing and other means of income shifting to declare as large a share as possible of their taxable corporate income in Puerto Rico. Grubert and Slemrod’s key empirical finding is that the effective marginal cost of income shifting is declining in the amount of real activity conducted in Puerto Rico: more real activity makes it easier for the firm to claim legitimate income shifting. Importantly, avoidance technology differs across firms and, as expected, Grubert and Slemrod (1998) found more evidence of tax-motivated production in Puerto Rico among firms for which arms-length transfer prices are difficult to determine, such as manufacturers with intangible brand value or pharmaceutical companies with high R&D expenditure. This provides an implicit subsidy to real investment in Puerto Rico, what Slemrod (2001) calls “avoidance facilitation.”

The notion of avoidance facilitation allows us to reinterpret the fascinating study of Rosen (1976) which, in a regression analysis of female labor supply, splits apart as explanatory variables the pre-tax wage rate, w , and the net-of-tax $(1 - t)$ terms, interpreting any difference in estimated coefficients (which Rosen (1976) did not ultimately find) as evidence of “tax illusion”—lack of salience in more modern language. But we can now say that, in the presence of avoidance and evasion, no appeal to tax illusion is needed to explain a differential effect of the w and $(1 - t)$ terms, because the effective relative price of the real decision depends on the avoidance or evasion technology—the tax can be “finessed,” but the pre-tax wage rate cannot be. *In general, one should not expect a homologous response to a pre-tax price and a net-of-tax term.*

4 Multiple Tax Instruments

4.1 Withholding and Information Reporting

Withholding refers to a situation where some or all of a tax liability must be remitted by someone other than the statutory bearer of the liability. It facilitates administration by allowing the tax authority to take advantage of economies of scale that exist in dealing with a smaller number of larger remitters who have relatively sophisticated record-keeping and accounting systems. It also acts as a revenue safeguard, ensuring that some tax is remitted even when the statutory bearer fails to file a return or otherwise disregards the tax obligations.

Withholding remittance responsibility is usually restricted to businesses and government agencies. Individuals in their capacity as employees and consumers are usually excluded—they are too numerous and not sufficiently capable in general to be suitable withholding agents. In order to be able to withhold an appropriate amount of tax, the withholding agent must have an ongoing relationship with the statutory bearer of the tax or, alternatively, the withholding scheme must be sufficiently simple to avoid the need for such a relationship.

Withholding is about which individuals or entities remit a given tax liability, not about what triggers tax liability. The tax remittance (or, alternatively, tax collection) structure is, however, a more general concept than what is called withholding. For example, a retail sales tax is a consumption tax under which all tax liability is remitted by retail businesses, while a value-added tax is a consumption tax under which the tax liability is remitted by firms all along the importing, producing and distributing chain. Although a VAT is generally not referred to as a withholding tax arrangement for a retail sales tax, the relationship between the two taxes is very similar, and is especially clear in an invoice-credit method VAT, where each business purchaser can credit against its own tax liability the tax remitted on the transaction by the seller.

Information is central to the operation of non-capricious tax systems that are based on observable, verifiable quantities. In self-assessment systems, much information is provided to the tax authority by the taxpayer, subject to verification by the former, which can be obtained via au-

dit, or third-party information reports. Third-party information reporting refers to a mandatory requirement on certain third parties (e.g., businesses, financial institutions, and non-tax-related government agencies) to report payments (and other tax-related transactions) and payee details (generally with a taxpayer identifying number) to the tax authority. This information can then be compared against tax actually remitted, allowing suspect returns to be identified and for enforcement actions to follow. Successful evasion now requires coordination between the party providing the information report and the party responsible for remittance, but—and here is the key—their incentives and willingness to falsify the data are unlikely to be the same (incentives can even work in opposite directions). Thus, a working system of information reporting discourages noncompliance by increasing the risk of detection for a given amount of tax authority resources. In practice, for such arrangements to be sufficiently efficient to make them attractive to revenue bodies there must be electronic reporting by third parties of information reports and the use of a high-integrity taxpayer identifier to facilitate the matching of information reports with tax authority records.

The central role of information reporting is reinforced by IRS data that shows that components of taxable income for which information reports are nonexistent or of limited value, such as other non-wage income and tax credits, have relatively high estimated misreporting rates. IRS (2012) reports that the net misreporting rate is 56, 11, and 8 percent for income types subject to “little or no,” “some,” and “substantial” information reporting, respectively, and is just 1 percent for those amounts subject to both withholding and substantial information reporting. These relative magnitudes are strikingly consistent with the deterrence model.

4.2 Market Transactions

Basing tax liability on market transactions—those between a willing buyer and an unconnected willing seller—has several advantages. First, in a market transaction information can potentially be obtained from either party, which provides a natural check on its accuracy. A second property is that market transactions tend to be better documented, and the more documented a transaction, the lower is the cost of gathering information on it. Finally, market transactions between unrelated parties establish arm’s-length prices, the availability of which greatly facilitates valuing the

transaction. VAT, for example, relies almost entirely on market transactions, while taxing capital gains on a realization basis rather than the theoretically-preferable accrual basis takes advantage of the measurement advantage of market transactions. In contrast, estate and wealth taxation cannot in general take advantage of market transactions to reliably value wealth.

Not all market transactions facilitate monitoring of the tax base, with market transactions carried out in cash being particularly difficult for a tax authority to monitor. Morse et al. (2009) document that many evading cash businesses construct a parallel cash economy in which they collect cash revenue, pay some expenses in cash, and then use the unreported cash they receive for cash purposes. A cash business might also pay for inventory and other expenses in cash, and not report the expenditure, thus reducing any suspicious discrepancy between revenues and costs (that should be nearly proportional to sales). One problem that arises for cash businesses is that cash transactions are susceptible to employee whistle-blowing, which Morse et al. (2009) suggest can be addressed by hiring family members. Kopczuk and Slemrod (2011) offer a stylized model of family firms in which the benefit of a trusted employee is balanced against the cost of hiring an otherwise ill-suited employee.

4.3 Public Disclosure

Another information-based tax-system instrument is public disclosure of tax details. As discussed in Lenter et al. (2003), supporters of public disclosure cite the gains to the transparency of tax policy, but also note that it may contribute to tax compliance because egregiously low income declarations might elicit private information that contradicts a non-compliant taxpayer's claim; thus, it is a way for the tax authority to collect relevant information.

The small empirical literature on the effects of public disclosure has so far provided mixed evidence about its effect on tax compliance. Hasegawa et al. (In Press) examine the Japanese disclosure system, which ended in tax year 2004, and required disclosure of corporate and individual taxable income only over a threshold amount. They find evidence that many corporations and individuals manipulated their reported income to be below the disclosure threshold, but do not find evidence

supporting an overall positive effect on compliance for either tax.

Slemrod et al. (2012) study the Norwegian disclosure system, making use of the fact that from 2001 the Norwegian tax data were made available on the Internet, whereas prior to that date they were easily available only in a select number of communities. They find a small average increase in reported business incomes after 2001 in communities that previously had limited disclosure, consistent with public disclosure deterring tax evasion.

5 Tax Base Elasticity

5.1 Elasticity of Taxable Income

Because it holds the promise of summarizing the welfare cost of all behavioral responses—real, avoidance, and evasion—undertaken to reduce tax liability, the elasticity of taxable income (ETI) has assumed a central role in measuring the marginal excess burden of income taxation (Feldstein 1999, Usher 1986). More generally, the argument for considering all behavioral responses applies to any tax base, not just personal income. The larger is the elasticity of response, the higher is the marginal excess burden per dollar raised. In most standard models, hours of work is the only dimension of behavioral response to changes in marginal tax rates, but the ETI concept generalizes the set of behavioral responses to all the margins of adjustment affecting taxable income. In addition to hours of work, higher marginal tax rates may affect taxpayer’s choices over intensity of work, quantity of income-tax-deductible consumption (e.g., charitable giving), career choice, form and timing of compensation, tax avoidance, and tax evasion (Saez et al., 2012).

These behavioral responses matter because, provided the private and social costs of sheltering one dollar of income from taxation are equal, the ETI is a sufficient statistic for welfare analysis. At the margin, a taxpayer will incur cost equal to one dollar, or sacrifice utility valued at one dollar, to save one dollar in taxes. Taxpayers adjust all means available to reduce taxes by one dollar up to the point that the marginal cost of doing so rises to one dollar. The anatomy of behavioral response does not, in principle, matter because at an optimum for the taxpayer the marginal cost

of saving a dollar in taxes is equal across all margins of behavioral response.

Empirically identifying the ETI boils down to estimating what taxpayers' taxable income would have been absent any change in marginal tax rates. Most recent empirical work has been based on sophisticated difference-in-difference (DD) methods, which rely on identifying some "control" group to measure the change in taxable income that would have occurred to the "treatment" group absent the tax reform. Unlike experimental studies, treatment and control groups are not randomly chosen, but are selected into by taxpayers according to income. This makes estimates biased if there is any systematic variation across the treatment and control groups—due, for example, to changes in inequality, real economic growth, and business-cycle dynamics—that is correlated with, but not caused by, the tax reform under investigation. Panel data, that allow some sources of non-tax related variation in incomes to be controlled for, can be potentially informative. These issues are discussed at much greater length in Saez et al. (2012). Based on a comprehensive reading of the empirical literature, Saez et al. (2012) suggest that the longer-term ETI lies between 0.12 and 0.40.

5.2 Issues in Applying the Elasticity of Taxable Income

Despite the attraction of the ETI as a sufficient statistic—in principle eliminating the need to identify the anatomy of behavioral response—the overwhelming evidence of fiscal externalities evident in the empirical ETI literature highlights the need to carefully adjust raw ETI estimates to arrive at an accurate measure of marginal welfare cost. Two main issues come up in applying an estimated ETI to welfare analysis: i) fiscal externalities and income shifting, and ii) changes in the definition of the tax base. Fiscal externalities arise if a change in marginal income tax rates induces taxpayers to shift income to another tax base, be it, for example, from the individual to the corporate income tax base, or to any base at a different date. When fiscal externalities arise, the ETI overstates efficiency cost because the observed reduction in taxable income is, to some extent, offset by socially valued revenue that is not accounted for. Hence, when constructing an ETI suitable for welfare analysis, it is necessary to determine for each tax change under consideration the extent to which revenue reductions in the applicable base lead to increased revenue elsewhere,

and to know the relevant tax rate applying to that income.

There are several prominent cases of income shifting across tax bases. For example, the Tax Reform Act of 1986 reduced the top-bracket marginal personal income tax rate to below the corporate income tax rate, increasing the attractiveness of partnerships and S-corporations, for which distributions are taxed as personal income, compared to C-corporations where income is first subject to the corporate tax rate. Thus, revenue shifted into the personal tax base from the corporate tax base. Similarly, anticipated increases in top-bracket marginal income tax rates for 1993 led some taxpayers to re-time income realization to the 1992 tax year (Feldstein and Feenberg, 1996). The ability of many high-income earners to re-time income suggests fiscal externalities across time periods are likely to be larger for temporary than permanent tax changes (Goolsbee, 2000). But, on the other hand, adjustment frictions and longer-run considerations, such as bequest planning, point to potentially larger fiscal externalities for permanent tax changes. A complete accounting of fiscal externalities would also take general equilibrium effects into account. For example, lower marginal income tax rates may increase labor supply and raise the pre-tax return on complementary factors, such as capital, that are taxed via other bases.

Some changes in the definition of the tax base also affect the magnitude of the ETI, and its policy implications. A narrow tax base is likely to have a higher ETI than a broad base with few exemptions because avoidance schemes are facilitated by loopholes in the tax code. Because the ETI is a function of the breadth of the tax base, it is not an immutable structural parameter, instead being partially under policymakers' control. Slemrod and Kopczuk (2002), whose model we discuss in the next section, argue that there is an *optimal* ETI.

6 Standard Instruments With New Costs

Having laid out the building blocks of tax systems, we turn now to what a tax-systems perspective implies about the appropriate design of tax policy. We take a fresh look at some standard optimal tax questions, and also examine what insights can be gained from applying the standard optimal tax analysis to non-standard aspects of tax systems.

6.1 The Optimal Commodity Tax Base

In the classic optimal commodity tax problem, first laid out by Frank Ramsey in (1927), at an optimum the marginal excess burden per dollar raised is equated across all taxed commodities. Assuming perfectly elastic supply of commodities, we need know “only” the matrix of compensated own and cross-price demand elasticities to determine the optimal set of relative commodity tax rates. In the special case where all cross-price elasticities of demand are equal to zero, but not generally, optimal commodity taxes follow an “inverse elasticity” rule, in which at an optimum the least price-elastic commodities face the highest tax rates.

Crucially, in this model the tax base—which commodities can be taxed—is held fixed. However, because collection costs—administrative and compliance costs—can vary substantially by commodity, the breadth of the commodity tax base is also an important real-world choice variable. In general, a broad commodity tax base distorts purchasing decisions less than a narrow tax base, but if breadth comes at the cost of additional collection cost, optimality requires trading off reduced excess burden from distorted consumption choice against higher collection cost.

Yitzhaki (1979) studies this problem in a model with a representative consumer and a set of commodities arranged in increasing order of fixed administrative cost per dollar of consumer expenditure, with untaxed items imposing no direct cost on the tax authority. The representative consumer is assumed to have Cobb-Douglas preferences, which greatly simplifies the choice of tax rates for each commodity, because the homotheticity and separability properties of these preferences together imply that excess burden is minimized with an equal tax rate on all taxed commodities. This highlights the trade-off between the tax rate and tax base breadth and sidesteps the question of the optimal pattern of taxes on the taxed commodities—it is always uniform at a rate high enough to raise the fixed required revenue.

Formally, the tax authority’s problem is to jointly choose the uniform commodity tax rate and the breadth of the tax base, defined as the sum of expenditure shares for the taxed commodities, subject to a revenue constraint. This formalizes a ubiquitous tax policy issue—to what extent to broaden the tax base and “use” the revenue raised thereby to lower tax rates or, in this model,

to lower *the* tax rate. For the retail sales tax, the recurring issue is whether to include services in the tax base; services are a growing fraction of consumer expenditures but are widely thought to be more costly to collect per dollar raised. To be sure, there are other justifications—real and imagined—for special treatment of some consumption goods, such as externalities or redistribution, but this model captures the essence of the policy tradeoff.

A narrow tax base creates a relatively large excess burden but has low administrative cost, while a broad tax base has small excess burden, but requires more resources to administer. At an optimum the marginal efficiency cost of funds must be equal whether it is raised from a broader tax base or a higher commodity tax rate. To be sure, the assumption of smoothly increasing administrative costs overlooks some important real-world practicalities. For example, it may in some settings be less costly to tax *all* commodities than all but one, in which case administrative costs are discontinuous, and the optimally taxed set of commodities no longer varies smoothly with revenue required.⁴

Broadening the tax base is also a perennial issue in income taxation where, for example, eliminating the deduction for charitable contributions would allow a lower tax rate(s) without sacrificing revenue. In a single-period framework, because the tax rate on all taxed goods is the same, the Yitzhaki model can be equivalently interpreted in terms of the share of goods whose expenditure is deductible from taxable income in a linear income tax. Permitting a large share of goods to be deducted from taxable income is equivalent to a narrow commodity tax base. But putting it this way highlights that it is not always true, as in the Yitzhaki model, that broadening the base raises collection costs: a comprehensive income tax base may be *less* costly to collect than one with exceptions.

6.2 The Optimal Size of a Tax Collection Agency

How many resources should government devote to tax enforcement? Intuitively, not as much as would be needed to eradicate evasion completely, just as it does not make sense to station a police

⁴The intuition gleaned from this model survives relaxing the assumption of Cobb-Douglas utility to the more flexible constant-elasticity-of-substitution form, in which case Wilson (1989) finds that the higher is the elasticity of substitution between taxed and untaxed commodities, the broader is the optimal tax base.

officer at every street corner to eradicate street crime. One needs to distinguish economically recoverable tax evasion just as one distinguishes economically recoverable oil reserves from total oil on the planet.

Because one measure of output of a tax agency, extra tax revenue received through direct collection efforts as well as deterrence, is denominated in dollars and is in principle measurable, some people have been tempted to assert a simple rule for the optimal extent of tax enforcement: maximize net revenue. Indeed, IRS commissioners seeking a higher budget appropriation and public finance textbooks of a certain vintage assert that increasing the audit budget for the IRS is appropriate because it would raise net revenue, implicitly presuming that at an optimum marginal audit revenue and the marginal cost of obtaining that revenue, say via additional audits, should be equated. But, as Slemrod and Yitzhaki (1987) show, this apparently intuitive argument leads to a socially excessive level of enforcement because it ignores the fact that real resources are being consumed to implement what is simply a transfer of funds between households and the tax authority; thus, the implied optimality condition compares (marginal) apples and oranges.

There is an excess burden arising from evasion because it involves paying taxes in the form of a gamble rather than with certainty. The optimal audit probability is set such that the marginal social cost of conducting audits equals the marginal reduction in the representative taxpayer's excess burden. Equating marginal audit revenue and marginal administrative cost—as IRS commissioners and certain textbooks suggest—always results in a socially excessive level of enforcement. Intuitively, a revenue maximizer raises no net revenue on the margin but takes a full dollar from taxpayers. Spending a dollar in administrative costs to take a dollar from taxpayers can never be optimal.

6.3 Endogenous Elasticity

Because the elasticity of behavioral response with respect to a tax rate is a central concept in modern tax analysis, it is important to note that behavioral elasticities with respect to tax rates depend on other tax system parameters. To see this, first consider the standard model of optimal

progressivity, where the social planner chooses a tax function to maximize social welfare assuming the elasticity of taxable income, which in many models is equivalent to the elasticity of labor supply, is a structural preference parameter. The change in focus from labor supply to taxable income is important, because numerous non-rate tax-system aspects under the control of policymakers—such as the breadth of the tax base, opportunities for sheltering, and enforcement effort—may affect the tax rate elasticity.

Recall that one of the motivating examples for focusing on taxable income and its elasticity, rather than on labor supply alone, is the idea that higher tax rates may reduce labor supply only slightly, but may induce significant shifting into inefficient untaxed fringe benefits, tax shelters, and so on. But tax policy could also directly address these issues by paring tax-exempt activities, restricting shelters, and so on, which could substantially reduce the ETI, which in turn might increase the optimal level of progressivity.

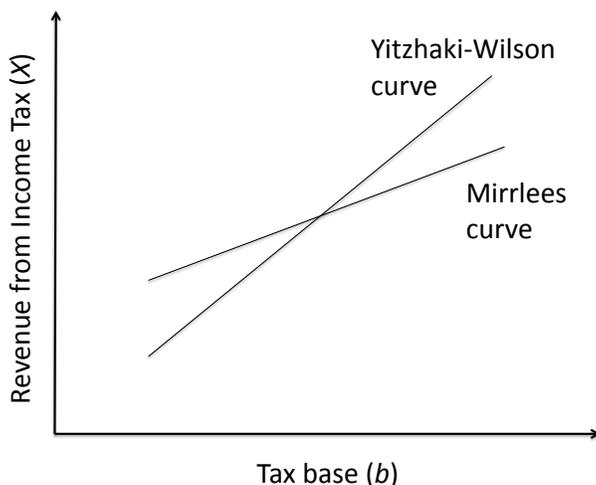
To address this issue, Slemrod and Kopczuk (2002) modify the standard problem by allowing the social planner to control a tax system instrument that affects the elasticity of taxable income with respect to a tax rate, albeit with a resource cost. They focus on an example in which the social planner controls the elasticity of taxable income by choosing the set of commodities that are tax-deductible, which in their one-period model is equivalent to choosing which commodities are taxed. This set-up links up to the literature on the optimal breadth of a commodity tax base discussed above.

In the model of Slemrod and Kopczuk, consumers have Cobb-Douglas utility over a continuum of goods. The vector of tax system instruments consists of a linear tax rate on income, a demogrant, and the set of goods subject to tax. Taxpayers are assumed to supply one unit of labor inelastically. Strikingly, the compensated elasticity of taxable income in this model with respect to the net-of-tax rate simplifies to equal exactly one minus the expenditure share of goods subject to tax: the broader is the tax base, the lower is the elasticity. This makes intuitive sense because there are fewer untaxed commodities toward which to substitute when the tax rate rises. As in Yitzhaki (1979) and Wilson (1989), a broader tax base lowers the excess burden of taxation, but is assumed to raise administrative costs. For any given size of the tax base, the commodities minimizing

administrative cost are assumed to be included in the tax base, so that administrative cost is a smoothly increasing function in the neighborhood of the optimum tax policy.

The first-order conditions for the demogrant and the tax rate define the optimal tax policy, holding the breadth of the tax base fixed, as in the Mirrlees (1971) problem, where the base was implicitly fixed to be all income. Solving for optimal policy at each value of tax base breadth produces an upward-sloping “Mirrlees” curve in (X, b) space, where X is the amount of revenue raised from distortionary income taxes, and b is tax base breadth. Because the elasticity of taxable income in this model is decreasing in the breadth of the tax base, ceteris paribus having a broader tax base implies it is optimal to have a more progressive tax system, with the amount of revenue raised from distortionary taxes serving as a sufficient statistic for progressivity.

Figure 1: The Mirrlees and Yitzhaki-Wilson Curves and the Optimal Tax System



In contrast, taking the demogrant and revenue requirement as given and solving for the optimal tax rate and tax base breadth, as in Yitzhaki (1979) and Wilson (1989), defines the optimal breadth of the tax base holding progressivity constant. Solving for all values of the demogrant produces an upward sloping “Yitzhaki-Wilson” curve in (X, b) space. This implies that more egalitarian societies (raising more revenue from distortionary taxes) will have a broader tax base, and so a lower compensated elasticity of taxable income.

The globally optimal tax policy is found at the intersection of the “Mirrlees” and “Yitzhaki-Wilson” curves, defining the optimal degree of progressivity and the optimal elasticity of the tax base. Slemrod and Kopczuk (2002) prove that, in the neighborhood of an optimum, the “Yitzhaki-Wilson” curve cuts the “Mirrlees” curve from below, so that an increase in egalitarianism (which shifts up the Mirrlees curve) will increase the amount of revenue raised from distortionary taxes, broaden the tax base and lower the compensated elasticity of taxable income at the optimum. Slemrod and Kopczuk (2002) also prove that an increase in the cost of maintaining a broad tax base—that is, an increase in marginal administrative costs holding the level of administrative costs constant—decreases the optimal progressivity of the tax system and narrows the optimal tax base.

7 Bringing Firms into Optimal Tax Theory

7.1 Remittance

In the standard model, who or what entity must *remit* the tax triggered is unspecified, and presumed irrelevant. Indeed, there is an invariance, or irrelevance, proposition emphasized in most, if not all, public finance textbooks: it doesn’t matter which side of a taxed transaction must remit tax (or which side has the statutory liability), the incidence is the same, and indeed *all* of the consequences (other than what the pre- and post-tax prices are called) are the same. The invariance theorem is a folk theorem, in that the assumptions are not made explicit. The theorem suggests, for example, that it doesn’t matter if a retail business or a consumer remits a retail sales tax: the outcome is exactly the same. It doesn’t matter whether only the retail businesses remit based on their sales, as under a RST, or whether all businesses remit based on their value added, as in a VAT. Current and historical practice suggest that the remittance structure does in fact matter. Over 30 countries levy a value-added tax at a rate of 25 percent or above, and more than 70 countries levy a value-added tax at a rate of 15 percent or above (KPMG, 2012), and few levy a retail sales tax at a rate over 10 percent; we can infer that these textbook-equivalent systems are not practice-equivalent. The widespread existence of withholding and remittance at the source of payment demonstrates the importance given to it by tax policy makers.

Whenever the tax evasion or avoidance technology differs between potential remitters the textbook irrelevance proposition no longer holds, and the assignment of statutory tax liability has real consequences. Kopczuk et al. (2013) provide empirical evidence in support of this proposition based on changes in the point of collection for state diesel taxes. They exploit variation across states and through time in the proximity of the point of collection to final consumption for diesel fuel. Collection at the retail level requires the tax authorities to monitor a large number of sellers, whereas collection at the distributor level requires monitoring many fewer sellers, and collection from distributors at the top of the supply chain requires monitoring an even smaller number of sellers. Kopczuk et al. (2013) find that a re-assignment of remittance responsibility up the supply chain—from retailers to distributors, distributors to wholesalers, or retailers to wholesalers—is associated with a higher level of pass-through of the statutory tax burden to consumers, which contradicts the textbook irrelevance proposition. They attribute their findings to evasion being more costly the fewer are the number of remitters of the tax, which will be the case, for example, if audit coverage is decreasing in the number of remitters, perhaps because there are fixed per-firm costs of monitoring compliance. Under their assumption of perfect competition, a reduction in tax evasion from a re-assignment of tax liability up the supply chain is passed through into higher consumer prices.

7.2 Taxes and the Missing Middle

Firms below a certain size are often exempt from some taxes for administrative reasons, leading to discontinuous tax treatment in firm size and few firms observed on the high-tax side of the cut-off. For example, in Japan small firms receive favorable VAT treatment, while in the U.S. corporations averaging less than \$5 million in gross receipts over the prior three years may use a cash rather than accrual accounting method and are exempt from the corporate alternative minimum tax.

Although explicit or implicit exemption, or more generally special tax treatment, of small firms might economize on collection costs (both compliance costs borne in the first instance by taxpayers and administrative costs borne in the first instance by the tax authority), it also generally causes production inefficiency, in part because it provides a tax-related incentive for firms to be small.

The tradeoff between the costs of collection and production inefficiency has not been closely addressed by the optimal tax literature. In part, this is because nearly all modern tax theory has been developed assuming that the party remitting taxes is economically irrelevant. In addition, the seminal Diamond and Mirrlees (1971) theorem on aggregate production efficiency posits that production inefficiencies, including from discriminating among firms in the same sector, should not be tolerated if the government faces no constraints on its ability to levy optimal commodity taxes. But their model of optimal taxation ignores administrative costs and assumes that there are no untaxed profits, due either to constant returns to scale—in which case firm size is indeterminate—or a 100 percent profits tax.

An exception is Dharmapala et al. (2011), who develop a model in which fixed per-firm administrative costs are responsible for the production inefficiencies. With administrative costs, Dharmapala et al. (2011) identify cases in which it is optimal to exempt small firms from taxation, creating production inefficiencies that are never part of an optimal tax system in the Diamond and Mirrlees framework. These inefficiencies occur because different firms in the same industry sell output at different prices, and also because some firms obtain the tax exemption by reducing their outputs to inefficiently low levels, creating a “missing middle” of firm size. The central claim—that a missing middle can potentially be generated by optimal tax policies—is theoretical in nature, rather than empirical (that observed “missing middles” correspond to existing tax thresholds). However, there is an emerging body of empirical evidence documenting cases in which the distribution of firm size has been affected by various tax and regulatory thresholds. For instance, Onji (2009) analyzes the introduction of a value-added tax (VAT) in Japan in 1989 that incorporated preferential treatment for small firms, with a cutoff for eligibility of 500 million yen in sales, and finds a clustering of firms just below this threshold following the reform.

8 Lines and Notches

8.1 Tax Driven Product Innovation

As any tax lawyer will say, real-world, in-the-trenches, scuffling about taxation is largely about drawing and interpreting lines that separate the cases where discretely different tax treatments apply.⁵ Is a given charitable donation deductible or not? Are food expenditures a deductible business expense for someone who sells their blood, or not? Is a worker an employee, for which an employer must withhold and remit income and payroll taxes, or an independent contractor, for which no withholding is required? Is a given hybrid corporate security debt or equity? In spite of its ubiquity and policy relevance, line-drawing is almost completely absent from economic analysis. Why?

Let's begin with commodity taxation. The major reason for ignoring lines is that the modern theory of optimal commodity taxation allows for a different tax on each good. But it is excessively costly to administer a separate tax on each good (or service), especially given that new goods are being created every day. Actual commodity tax systems feature a very small set of rates and each tax rate applies to a wide range of goods. Whenever selective commodity taxation is called for, a non-capricious tax system must have procedures for distinguishing among goods subject to different tax rates. Real-world consumption tax systems do that by appealing to the characteristics of the commodities. For example, the retail sales taxes of U.S. states often exempt food but not restaurant meals, requiring the tax law to draw a line between the two categories. This is done by appealing to a set of characteristics of a restaurant meal, and the line can be fine when, for example, grocery stores sell pre-prepared meals that may or may not be eaten on the premises, or set up in-store salad bars.

Some recent work (Kleven and Slemrod, 2011) addresses this set of issues by re-formulating optimal commodity tax theory in the language of characteristics, so that it matches up more easily with real tax systems. To do so, they make use of Lancaster's (1966, 1975) idea that it is the characteristics of goods, not the goods themselves, which are the direct objects of utility, and

⁵See Weisbach (1999) for an insightful legal-analytical view of line drawing.

there exists a mapping of each good into characteristics space. They formalize the relationship between characteristics, substitutability and optimal tax rates, which allows one to explore the notion that shared characteristics can be used to gauge substitutability and hence optimal tax rate differentials. As expected, the closer two goods are in characteristics space, the smaller the optimal tax rate differential.

With this reformulation, one can naturally address another important aspect of reality that has been heretofore ignored by the literature on optimal taxation, *tax-driven product innovation*, the creation of new products that requires no technical innovation, but which represent a re-packaging of characteristics so as to reduce tax liability. On Wall Street or the City of London, tax-driven product innovation is a major pre-occupation, where one objective is to design corporate finance vehicles that qualify for the interest deduction accorded to debt finance, but have most or all of the characteristics of an equity security. Sallee and Slemrod (2012) show that in response to the U.S. Gas Guzzler Tax, car manufacturers tended to produce cars with fuel efficiency just high enough to reduce the “notched” (defined below) tax liability. Our favorite examples, discussed by Harberger (1995), concern the response to high tax rates and tariffs on automobiles. In Indonesia the preferential tax treatment of motorcycles led to the creation of a new type of motorcycle with three wheels and long benches at the back seating up to eight passengers. In Chile the market responded to high taxes on cars, but not on panel trucks, by introducing a redesigned panel truck that featured glass windows instead of wood panels and upholstered seats in the back.

8.2 Notches

Lines create notches. Perhaps the simplest type of notch is a discontinuity in tax liability as a function of the size of the base. We need look no further than the U.S. federal income tax code for an example of an—admittedly fairly trivial—quantity notch. People with taxable incomes between \$3,000 and \$100,000 are permitted to calculate their tax liability via a tax table made up of brackets in increments of \$50, so tax liability is discontinuous—albeit only up to a maximum of \$17—for each \$50 of taxable income earned. Tying eligibility for tax and transfer programs to incremental changes in taxable incomes also creates notches in the income tax code. For example, the U.S.

Saver's Credit provides a 50 percent credit on deposits into retirement accounts for married filers earning less than \$30,000, but only a 20 percent credit for filers earning \$30,001 or more. With contributions eligible for a credit capped at \$2,000, there is a notch of up to \$600 at a taxable income of \$30,000.⁶

Abrupt changes in tax policy, often caused by the use of (usually annual) accounting periods, create notches whenever tax liability is non-linear in taxable income, or tax policy changes are anticipated before their introduction in the new year. For example, the yearly system of accounting used in most income tax codes provides incentives for taxpayers to re-time income realization whenever reporting in one period rather than another avoids incurring taxes at a higher rate. An example of this sort of re-timing was the legislation in 1992 of a higher top marginal income tax rate from 1993, which led many Wall Street firms to accelerate bonus payments so that in aggregate about two-thirds were paid as end-of-year bonuses, compared to the norm of two-thirds beginning-of-year bonuses (see Parcell, 1995), and Goolsbee (2000) documents a large increase in the taxable exercise of stock options in 1992.

When there are differences in tax systems across geographic borders, notches exist because an incremental change in the location of economic activity leads to a jump in tax liability. Consumers living close to borders purchase relatively little on the high-tax side of the border, while corporations choose their location taking geographic differences in tax rates affecting them into account. In principle, jurisdictions could remove these notches by setting graduated tax rates based on distance to the border. This would favor those living closest to the border, but so does a uniform rate policy in a high-tax jurisdiction because those living nearest the low-tax neighboring region face the smallest transportation costs. As one example, in the U.S. differences in state sales tax rates create discontinuities at state borders.⁷

Regardless of the desirability of notches as part of an optimal tax system, they can provide an excellent opportunity to identify behavioral elasticities. At a notch, marginal tax (or subsidy) rates exceed 100 percent, inducing behavioral response even in the presence of frictions. But,

⁶Ramnath (2013) provides evidence of bunching in taxable income at notches created by the Saver's Credit.

⁷Agrawal (2011) finds that, in order to reduce the extent of cross-border shopping, municipalities on the low-state-tax side of a border choose local sales tax rates that eliminate more than half of the difference in state tax rates.

despite the strong bunching incentives created by notches, optimization frictions for some taxpayers can still be large enough to meaningfully attenuate structural elasticity estimates. Exploiting administrative data on income tax filings in Pakistan, which has a tax system featuring notches between each income tax bracket (*average* tax rates are constant within each bracket), Kleven and Waseem (2013) find that about 90 percent of wage earners and 50-80 percent of self-employed filers reported incomes in strictly dominated regions. They go on to show that the mass of taxpayers in the dominated region can be used to separately identify the structural (i.e., ignoring optimization frictions) and the observed elasticity attenuated by optimization frictions. Roughly speaking, their technique for estimating structural elasticities requires inflating the observed elasticity inferred from a notch by the share of people in the dominated region, so in the case of wage earners in Pakistan the structural elasticity is about 10 times larger than the observed elasticity. Despite the observed elasticities being attenuated by optimization frictions, Kleven and Waseem (2013) estimate small structural elasticities; the unusually large marginal tax rates at a notch induce pronounced bunching behavior even with small structural elasticities for taxpayers not subject to prohibitively large optimization frictions.

9 Conclusion

For the most part standard theoretical tools of public economics can be applied insightfully to tax-systems issues, although future research should address some methodological weaknesses. The failure to formally integrate concerns about horizontal equity in a satisfactory way is particularly problematic for tax-systems issues such as evasion. Due to the practical importance of firms in tax collection, tax-systems theory needs to move beyond constant returns to scale models of production to models where heterogeneity of size naturally arises. Because the collection of information is central to taxation, the Holy Grail would be to insightfully integrate the theory of information into the theory of taxation.

As we seek the Holy Grail, our analysis must move beyond the either-or world of the first wave of optimal tax theory, where some information is available to government perfectly and costlessly,

while other information is impossible to obtain. The collection and verification of information is very costly, often at least 10 percent of revenue collected and, even with these resources devoted to collection, tax gaps are large in developed countries and even larger in developing countries. In the real world of taxation much of the action is in between two extremes: information is observable with error, to varying degrees, and its quality depends greatly on the type of administration and enforcement in place.

Policy makers are well advised to recognize the interrelationship among tax rates, bases, enforcement, and administration. There are many alternatives to raise revenue, and many types of costs, some that show up in government budgets and some that do not, and the costs of using one tax instrument often depends on the setting of others. Recognizing that tax policy is really tax-system policy can ward off substantial policy errors, such as foregoing tax rate increases because the existing (suboptimal) base is narrow, and misinterpreting large short-run responses to policy changes as indicative of long-run behavioral response.

Frank Hahn (1973, p. 106) once commented that “Optimum tax formulas are either guides to action or nothing at all.” The stylized treatment of information in the literature to this point has afforded tremendous insight into the consequences of taxation and its proper design. We believe that an important next step in completing this task is to pursue the issues this paper addresses.

References

- Agrawal, David R. (2011). The Tax Gradient: Do Local Sales Taxes Reduce Tax Differentials at State Borders? Working Paper, University of Michigan.
- Allingham, Michael G. and Agnar Sandmo (1972). Income Tax Evasion: A Theoretical Analysis. *Journal of Public Economics* 1(3-4): 323–338.
- Becker, Gary S (1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76(2): 169–217.

- Bird, Richard M. (1983). Income Tax Reform in Developing Countries: The Administrative Dimension. *Bulletin for International Fiscal Documentation* 37: 3–14.
- Blumenthal, Marsha, Charles Christian, and Joel Slemrod (2001). Do Normative Appeals Affect Tax Compliance? Evidence from a Controlled Experiment in Minnesota. *National Tax Journal* 54(1): 125–138.
- Casanegra de Jantscher, Milka (1990). Administering the VAT. In Malcolm Gillis, Carl S. Shoup, and Gerardo Sicat (Eds.), *Value-Added Taxation in Developing Countries*, pp. 171–179. Washington, DC: The World Bank.
- Coase, Ronald H. (1937). The Nature of the Firm. *Economica* 4(16): 386–405.
- Dharmapala, Dhammika, Joel Slemrod, and John D. Wilson (2011). Tax Policy and the Missing Middle: Optimal Tax Remittance with Firm-level Administrative Costs. *Journal of Public Economics* 95(9-10): 1036–1047.
- Diamond, Peter A. and James A. Mirrlees (1971). Optimal Taxation and Public Production I: Production Efficiency. *American Economic Review* 61(1): 8–27.
- Feldman, Naomi E. and Joel Slemrod (2007). Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns. *Economic Journal* 117(518): 327–352.
- Feldstein, Martin S. (1999). Tax Avoidance and the Deadweight Loss of the Income Tax. *Review of Economics and Statistics* 81(4): 674–680.
- Feldstein, Martin S. and Daniel Feenberg (1996). The Effect of Increased Tax Rates on Taxable Income and Economic Efficiency: A Preliminary Analysis of the 1993 Tax Rate Increases. In James Poterba (Ed.), *Tax Policy and the Economy*, Number 10, pp. 89–117. Cambridge, MA: MIT Press.
- Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler (In Press). Testing Enforcement Strategies in the Field: Legal Threat, Moral Appeal and Social Information. *Journal of the European Economic Association*.

- Goolsbee, Austan (2000). What Happens When You Tax the Rich? Evidence from Executive Compensation. *Journal of Political Economy* 108(2): 352–378.
- Grubert, Harry and Joel Slemrod (1998). The Effect of Taxes on Investment and Income Shifting to Puerto Rico. *Review of Economics and Statistics* 80(3): 365–373.
- Hahn, Frank (1973). On Optimum Taxation. *Journal of Economic Theory* 6(1): 96–106.
- Harberger, Arnold (1995). Tax Lore for Budding Reformers. In Rudiger Dornbusch and Sebastian Edwards (Eds.), *Reform, Recovery, and Growth: Latin America and the Middle East*, Chapter 9, pp. 291–310. Chicago: University of Chicago Press and NBER.
- Hasegawa, Makoto, Jeffrey Hoopes, Ryo Ishida, and Joel Slemrod (In Press). The Effect of Public Disclosure on Reported Taxable Income: Evidence from Individuals and Corporations in Japan. *National Tax Journal*.
- Kleven, Henrik J., Martin B. Knudsen, Claus T. Kreiner, Soren Pedersen, and Emmanuel Saez (2011). Unwilling or Unable to Cheat? Evidence from a Tax Audit Experiment in Denmark. *Econometrica* 79(3): 651–692.
- Kleven, Henrik J. and Joel Slemrod (2011). A Characteristics Approach to Optimal Taxation and Tax-Driven Product Innovation. Working Paper, University of Michigan.
- Kleven, Henrik J. and Mazhar Waseem (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *Quarterly Journal of Economics* 128(2): 669–723.
- Kopczuk, Wojciech, Justin Marion, Erich Muehlegger, and Joel Slemrod (2013). Do the Laws of Tax Incidence Hold? Point of Collection and the Pass-through of State Diesel Taxes, Working Paper, University of Michigan.
- Kopczuk, Wojciech and Joel Slemrod (2011). Taxation of Family Firms, Working Paper, University of Michigan.

- KPMG (2012). Corporate and Indirect Tax Survey 2012, KPMG International, Available at: kpmg.com/tax.
- Lancaster, Kelvin J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy* 74(2): 132–157.
- Lancaster, Kelvin J. (1975). Socially Optimal Product Differentiation. *American Economic Review* 65(4): 567–585.
- Lenter, David, Douglas Shackelford, and Joel Slemrod (2003). Public Disclosure of Corporate Tax Return Information: Accounting, Economics and Legal Issues. *National Tax Journal* 56(4): 803–830.
- Mirrlees, James A (1971). An Exploration in the Theory of Optimum Income Taxation. *Review of Economic Studies* 38(114): 175–208.
- Morse, Susan, Stewart Karlinsky, and Joseph Bankman (2009). Cash Businesses and Tax Evasion. *Stanford Law & Policy Review* 20(1): 37–68.
- Musgrave, Richard A. (1969). *Fiscal Systems*. Number 10 in Studies in Comparative Economics. New Haven, CT: Yale University Press.
- Onji, Kazuki (2009). The Response of Firms to Eligibility Thresholds: Evidence from the Japanese Value-Added Tax. *Journal of Public Economics* 93(5-6): 766–775.
- Parcell, Ann D. (1995). Income Shifting in Response to Higher Tax Rates: The Effects of OBRA93, Office of Tax Analysis, U.S. Department of the Treasury, Washington, DC.
- Pissarides, Christopher A. and Guglielmo Weber (1989). An Expenditure-Based Estimate of Britain’s Black Economy. *Journal of Public Economics* 39(1): 17–32.
- Pomeranz, Dina D. (2011). No Taxation Without Information: Deterrence and Self-Enforcement in the Value Added Tax. Working Paper, Harvard University.
- Ramnath, Shanthi (2013). Taxpayers’ Responses to Tax-Based Incentives for Retirement Savings: Evidence From the Saver’s Credit Notch. *Journal of Public Economics* 101: 77–93.

- Ramsey, Frank P. (1927). A Contribution to the Theory of Taxation. *Economic Journal* 37(145): 47–61.
- Rosen, Harvey S. (1976). Tax Illusion and the Labor Supply of Married Women. *Review of Economics and Statistics* 58(2): 167–172.
- Saez, Emmanuel, Joel Slemrod, and Seth Giertz (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review. *Journal of Economic Literature* 50(1): 3–50.
- Sallee, James M. and Joel Slemrod (2012). Car Notches: Strategic Automaker Responses to Fuel Economy Policy. *Journal of Public Economics* 96(11-12): 981–999.
- Sandford, Cedric T. (1973). *Hidden Costs of Taxation*. London: Institute for Fiscal Studies.
- Sandford, Cedric T. (1995). *Tax Compliance Costs: Measurement and Policy*. Bath, United Kingdom: Fiscal Publications.
- Shoup, Carl, Roy Blough, and Mabel Newcomer (1937). *Facing the Tax Problem: A Survey of Taxation in the United States and a Program for the Future*. By Committee on Taxation of the Twentieth Century Fund. New York: Twentieth Century Fund, Inc.
- Slemrod, Joel (1996). Which is the Simplest Tax System of Them All? In Henry J. Aaron and William Gale (Eds.), *The Economics of Fundamental Tax Reform*, pp. 355–391. Washington, DC: The Brookings Institution.
- Slemrod, Joel (2001). A General Model of the Behavioral Response to Taxation. *International Tax and Public Finance* 8(2): 119–128.
- Slemrod, Joel, Marsha Blumenthal, and Charles Christian (2001). Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota. *Journal of Public Economics* 79(3): 455–483.
- Slemrod, Joel and Wojciech Kopczuk (2002). The Optimal Elasticity of Taxable Income. *Journal of Public Economics* 84(1): 91–112.

- Slemrod, Joel and Nikki Sorum (1984). The Compliance Cost of the U.S. Individual Income Tax System. *National Tax Journal* 37(4): 461–474.
- Slemrod, Joel, Thor O. Thoresen, and Erlend E. Bø (2012). Taxes on the Internet: Deterrence Effects of Public Disclosure. Working Paper, University of Michigan.
- Slemrod, Joel and Caroline E. Weber (2012). Evidence of the Invisible: Toward a Credibility Revolution in the Empirical Analysis of Tax Evasion and the Informal Economy. *International Tax and Public Finance* 19(1): 25–53.
- Slemrod, Joel and Shlomo Yitzhaki (1987). The Optimal Size of a Tax Collection Agency. *Scandinavian Journal of Economics* 89(2): 25–34.
- Slemrod, Joel and Shlomo Yitzhaki (2002). Tax Avoidance, Evasion, and Administration. In Alan J. Auerbach and Martin Feldstein (Eds.), *Handbook of Public Economics*, Volume 3, pp. 1423–1470. Amsterdam: Elsevier.
- Tait, Alan A. (1988). *Value-Added Tax: International Practice and Problems*. Washington, DC: International Monetary Fund.
- U.S. Department of the Treasury. Internal Revenue Service. (2012). *The Tax Gap*. Available at: www.irs.gov/newsroom/article/0,,id=158619,00.html.
- Usher, Dan (1986). Tax Evasion and the Marginal Cost of Public Funds. *Economic Inquiry* 24(4): 563–586.
- Weisbach, David A. (1999). Formalism in the Tax Law. *University of Chicago Law Review* 66: 860–886.
- Wilson, John D. (1989). On the Optimal Tax Base for Commodity Taxation. *American Economic Review* 79(5): 1196–1206.
- Yaniv, Gideon (1988). Withholding and Non-Withheld Tax Evasion. *Journal of Public Economics* 35(2): 183–204.

Yaniv, Gideon (1992). Collaborated Employee-Employer Tax Evasion. *Public Finance* 47(2): 312–321.

Yitzhaki, Shlomo (1974). A Note on "Income Tax Evasion: A Theoretical Analysis". *Journal of Public Economics* 3(2): 201–202.

Yitzhaki, Shlomo (1979). A Note on Optimal Taxation and Administrative Costs. *American Economic Review* 69(3): 475–480.