International Tax and Public Finance manuscript No.
(will be inserted by the editor)

Evidence of the Invisible: Toward a Credibility Revolution in the Empirical Analysis of Tax Evasion and the Informal Economy

Joel Slemrod · **Caroline Weber**

Received: date / Accepted: date

Abstract Empirical research about tax evasion and the informal economy has exploded in the past few decades, seeking to shed light on the magnitude and (especially policy) determinants of these phenomena. Quantitative information informs the analysis of policy choices, enables the testing of hypotheses about determinants of this phenomenon, and can help with the accurate construction of national income accounts. Even as empirical analysis has burgeoned, some have expressed doubts about the quality and usefulness of some prominent measures. The fact that high-quality data is elusive is neither surprising nor a coincidence. The defining characteristic of tax evasion and informal economic activity—that they are generally illegal—often renders unreliable standard data collection methods such as surveys. Unlike invisible phenomena in the natural sciences, these invisible social science phenomena are hard to measure because of choices made by individuals. Analysis of tax evasion and the informal economy must proceed even in the absence of the direct observability of key variables, and theory should guide the construction and interpretation of evidence of the "invisible." In this paper, we address what can be learned using micro or macro data regarding tax evasion and the informal economy under given conditions and assumptions, and critically review some of the most common empirical methods in light of our conclusions. We conclude with an entreaty for researchers in this field to enlist in the "credibility revolution" Angrist and Pischke (2010) in applied econometrics.
JEL Codes H26 · O17

Keywords Tax evasion · Informal Economy

Joel Slemrod
University of Michigan
701 Tappan St.
Ann Arbor, MI 48109
Tel.: +734-936-3914. Fax: +734-936-8716
E-mail: jslemrod@bus.umich.edu

Caroline Weber
University of Michigan
611 Tappan St.
Ann Arbor, MI 48109-1220

"On principle it is quite wrong to try founding a theory on observable magnitudes alone. In reality the very opposite happens. It is the theory which decides what we can observe." Albert Einstein, quoted by Werner Heisenberg, in *Encounters with Einstein*, pp. 62ff.

1 Introduction

1.1 Background

Empirical research about tax evasion and the informal economy has exploded in the past few decades, seeking to understand the magnitude and (especially policy) determinants of these phenomena. Quantitative information informs the analysis of policy choices, enables the testing of hypotheses about determinants of this phenomenon, and can help with the accurate construction of national income accounts. Even as empirical analysis has burgeoned, some have expressed doubts about the quality and usefulness of some prominent measures. The fact that high-quality data is elusive is neither surprising nor a coincidence. The defining characteristic of tax noncompliance[1]—that it is generally illegal—often renders unreliable standard data collection methods such as surveys. Moreover, tax noncompliance is hard to measure because individuals purposely conceal it. This differs from the common situation in the natural sciences, where observation is often limited only by the measuring technology.[2] A classic example is the invention of microscopes and telescopes, which ended once and for all serious debate about whether there indeed existed worlds beyond the "naked eye," and enabled analysis of the microscopic and celestial worlds. In contrast, tax noncompliance shares some characteristics with other human activities that social scientists find difficult to measure: "victimless" crimes such as prostitution and drug use, for which victim surveys are of no use. Both the threat of punishment and shame (fear of public exposure) make measurement difficult.[3]

We share the sentiment expressed in the epigraph by Albert Einstein: analysis of tax evasion and the informal economy must proceed even in the absence of the direct observability of key variables, and theory should guide the construction and interpretation of evidence of the "invisible." In this paper, we address what can be learned about tax evasion and the informal economy under given conditions and assumptions, and critically review some of the most common empirical methods in light of our conclusions.

We conclude with an entreaty for researchers in this field to enlist in the "credibility revolution" Angrist and Pischke (2010) in applied econometrics that seeks to establish credible and explicit identification strategies both in archival research and randomized field experiments. This movement also aspires to overthrow poorly identified causal interpretations and casual use of instrumental variables, and to instill a

[1] We will use tax term "tax noncompliance" to refer to tax evasion and the informal economy, noting below the differences.

[2] The electron is not difficult to observe because it decides to be.

[3] One can, though, think of activities that are shameful for some people but not illegal (obesity, smoking), and some that are illegal but not shameful (speeding). One suspects that some people put tax evasion in the latter category, and even brag about successful evasion at cocktail parties.

skepticism with inference based on cross-sectional rather than longitudinal data for which one can plausibly argue that unobserved influences do not change over time. We do not attempt a comprehensive survey of the vast literature, and concentrate more on the tax, rather than the labor market, aspects of this set of issues.[4] We will say little about what has been learned, but will instead focus on how we can learn more about the methods of obtaining evidence and drawing inferences in a credible way, so as to facilitate the progressive accumulation of useful knowledge.

1.2 Definitions

To economists, tax evasion generally refers to efforts to not pay taxes by illegal means.[5] It is often contrasted with tax avoidance, which refers to the legal use of tax provisions to reduce tax liability. In practice, there is often considerable lack of clarity about whether a behavior is illegal, such that reasonable people (including the taxpayer and tax authority) may disagree. Definitions of the informal economy, which also goes by the shadow, black, or underground economy, vary a bit more. A reasonable one, adapted from Schneider (2005), is all market-based legal production of goods and services that are deliberately concealed from public authorities to avoid 1) payment of taxes or social security contributions, 2) having to meet labor market standards, or 3) complying with certain administrative procedures.

Clearly tax evasion and the informal economy are not identical but have some overlap. For example, overstating income-tax-deductible charitable contributions is tax evasion but is not part of the informal economy. A small business failing to register to avoid regulatory standards is part of the informal economy, but may not constitute tax evasion if the business's activities trigger no tax liability. Both are motivated by the desire to reduce the burden of some aspect of government, and may therefore respond to government policy. However, the informal economy refers explicitly to production, while tax evasion need not.

2 Theoretical Underpinning

Although in the epigraph Einstein asserts that theory determines what we can observe, for our purposes we need only claim that theory helps to make sense of the data patterns we observe. In particular, we need to refer to theory about the choices made by individuals, businesses, and governments. We therefore begin with an informal overview of some basic aspects of the relevant theory. This overview is not meant to be comprehensive, but rather to introduce the important structural relationships that should be considered in a credible empirical research design.

[4] Gërxhani (2004) and Slemrod (2007) offer concise recent surveys of the literature on the informal economy and tax evasion, respectively. Andreoni et al. (1998) provide a comprensive review of the economics of tax compliance.

[5] The legal definition varies across countries.

2.1 A Model of Agent Behavior

2.1.1 Evasion and Evasion-Facilitating Behavior

In the seminal treatment of income tax evasion[6] by Allingham and Sandmo (1972), a risk-averse individual with true income[7] Y subject to a flat income tax at rate t may report R, less than or equal to Y, where evasion, denoted E, is $Y - R$. If true income is underreported, there is some probability p that the evasion will be discovered by the tax authority, whence the true tax liability must be paid plus a penalty rate f that is related to either the understated income or the evaded tax liability. Although in the simplest model p is fixed, one can imagine that the probability of detection and punishment depends on three sets of factors: 1) a set of policy actions taken by the tax authority,[8] which we will denote, inclusive of the tax rate itself, as N; 2) a set of taxpayer-initiated evasion-facilitating activities, denoted B, designed to reduce the chance of being penalized; and 3) a set of exogenous characteristics of taxpayers and the non-tax-system environment they face, denoted H. In general, E declines when either p (given other choice variables) or f increases, but the effect of t is theoretically ambiguous.[9]

A large literature, summarized in Torgler (2007), has argued that the deterrence model is insufficient to explain variations in compliance behavior, and has introduced other considerations including social norms, intrinsic motivation such as duty, and perceptions about the fairness of the tax system and the purposes to which the tax revenue is directed. We accept that some people's behavior is affected by factors outside a simple private cost-benefit calculus.[10] This perspective broadens the vector of policy actions (to include, for example, taxpayer rights) and relevant exogenous characteristics, and can also blur the expected effect of traditional tax system actions (so that, for example, a heavy-handed enforcement may backfire by causing some people to switch how they frame compliance decisions from intrinsic motivation such as duty to extrinsic motivation based on playing the tax audit lottery).

2.1.2 A Model of the Informal Economy

The basic insights of the Allingham-Sandmo model, amended to allow for "behavioral" considerations, would apply to the informal economy, although some other issues arise. For example, the vector of tax system aspects N would have to be expanded to include regulations and other administrative burdens. Second, the model

[6] Some of the details of the model would be different if the tax in question was, for example, value-added tax evasion, but the basic points would be the same.

[7] True income is assumed to be equal to taxable income.

[8] N is exogenous to the taxpayer's choices but may vary systematically across jurisdictions or across time. As discussed later, the impact of N may vary across people or firms in a given jurisdiction at a point in time.

[9] As shown by Yitzhaki (1974), the effect of t depends in part on whether the penalty is assessed on the income understatement or the tax understatement.

[10] We do not, though, accept the argument, based on observed low audit coverage, that the deterrence model grossly underpredicts tax compliance; in fact, for income subject to information reports, the chance of getting away with evasion in developed coutnries is close to zero, in spite of low rates of formal auditing.

would have to consider the opportunity costs of informality, i.e. of joining the formal economy, as stressed by Loayza (1996) and others. This renders decisions about the informal economy unlike the tax evasion decision, where it is almost always the case that individuals can be free riders because their decisions do not affect the benefits they get from government at all.

Importantly, the Allingham-Sandmo model refers to a risk-averse individual.[11] Adapting the model to business decisions raises several interesting issues, which are discussed in Slemrod (2004). Small businesses, which are mostly the type at the margin for the informal economy, are reasonably thought of as individuals. However, as Slemrod (2004) points out, it is less clear whether a widely-held business ought to display risk aversion. It is also important to note that tax compliance decisions of businesses extend far beyond what are commonly considered "business taxes." For example, businesses commonly are required to withhold and remit taxes on behalf of their employees to meet their labor income tax obligations, and for many other payments they make. Indeed, recent studies have calculated that, in the U.S. and U.K., over 80 percent of all taxes are remitted by businesses,[12] and anecdotal evidence suggests that this figure is even higher in developing countries. Thus, noncompliance for withholding is an important issue. The incidence and efficacy implications of business noncompliance are certainly different than those of noncompliance by individuals in their role, say, as consumers.

2.1.3 Other Decisions

At the same time private agents are contemplating noncompliance and evasion-facilitating actions, they are making scores of other decisions. Individual consumption of any particular good X (as well as total consumption), depends on true after-tax income, or more precisely true after-tax expected permanent income.[13] Similarly, firms' demand for inputs is derived from their desired output. Thus, we might expect that, at the micro and macro level, there is a relationship between true income or output Y, and consumption or input demand, X.

2.2 A Model of Government Behavior

It is useful to think of governments and their tax authorities as choosing the tax and regulation systems to maximize some objective function.[14] Thus, the choice of N depends on H and on the tax authority's beliefs about how E and R will react to its choices of N. The tax system vector N includes the tax authority's attempts to

[11] Of course tax evasion may also be done by businesses, from small firms to multinational corporations.

[12] See Christensen et al. (2001) and Shaw et al. (2010).

[13] We could go further and recognize that labor supply (and therefore true income) is itself a decision that depends in part on the expected after-tax return to working, and therefore depends on aspects of the tax system including the tax rate and the enforcement regime. See Slemrod (2001) for a model of the joint choice of noncompliance and labor income in response to income taxation.

[14] The maximand may be a social welfare function of citizens' utilities, but could also include the private objectives of policy makers and thus reflect Leviathan considerations.

acquire information about Y, R, and E, through auditing of self-reports and matching of third-party information.

Noncompliance research must account for the fact that any analysis is undertaken in an environment where there are operational attempts by the tax and regulatory authorities to monitor noncompliance. The tax authority, as well as regulatory authorities, wants to enforce the tax laws and regulations. They also want to understand the marginal impact of policies, so as to guide rational resource allocation. Indeed, as we will discuss later, some of the most detailed information about the nature and magnitude of tax evasion comes from government studies whose main purpose is to inform the auditing strategies of the tax authority.

One clear implication of this discussion is that government policy cannot be considered exogenous. For this reason, finding "natural" experiments to exploit is difficult. Across countries (or other jurisdictions), policies may differ because country characteristics dictate that different policies are optimal. Unless those characteristics are completely controlled for, a cross-sectional regression analysis may ascribe the direct effect of these characteristics on noncompliance to the policy variables. Similarly, within a country over time, policies often change as country characteristics change. Within a jurisdiction at a given time, the policy usually does not vary, which is not promising for learning about the effect of policy variations. However, the application of the policy may vary. For example, the probability of detection may vary across individuals or firms, as might the applicability of the information reporting regime. Yet, this variation comes with its own problems. In particular, it is likely not exogenous, because the impact is likely correlated with preferences or ability to evade.

2.3 System

Before moving on to empirical research design issues, it will be helpful to collect all the structural equations our discussion so far suggests:

$$\text{Noncompliance: } E = E(Y,B,N,H) \text{ or } R = R(Y,B,N,H), \text{ where } E = Y - R. \tag{1}$$

$$\text{Traces of True Income or Output: } X = X(Y-T,N,H) \text{ or } X = X(Y,N,H). \tag{2}$$

$$\text{Traces of Noncompliance: } B = B(E,N,H). \tag{3}$$

$$\text{Policy Determination: } N = N(B,H). \tag{4}$$

3 Micro-based Inference about Noncompliance

3.1 Using Audited Returns or Surveys

There is a good reason why the credibility revolution has been late arriving to the study of tax evasion and the informal economy. Empirical analysis in this field faces extraordinary challenges. Many years ago, at a conference on tax evasion, a colleague and friend of the senior co-author of this paper began his remarks by saying that "the empirical analysis of the determinants of tax evasion is straightforward except for two things: you can't measure the left-hand side variable, and you can't measure the right-hand side variables."[15] This paper focuses mostly on the former problem. We begin, in this section, by discussing the most direct attempts to address this problem: intensive audits conducted by the tax authority and surveys conducted by researchers. Both can examine a random sample of individuals and produce an estimate of non-compliance for each individual. However, the former applies only to tax evasion, and can be spearheaded only by the tax authority at substantial resource cost. Despite its costs, such a study has been done regularly since the 1960's by the U.S. tax authority, the Internal Revenue Service (IRS). These studies were originally called the Taxpayer Compliance Measurement Program (TCMP), but more recently the name has changed to the National Research Program (NRP). The studies generate information, by line on the individual income tax form, about what the taxpayer reported and what the examiner concluded was correct. The primary purpose of this exercise is to improve the process for selecting returns for operational audits. It is also used to estimate the magnitude and nature of the aggregate "tax gap" by combining information from the intensive audits with information obtained from ongoing enforcement activities and special studies about sources of income, like tips and cash earnings of informal suppliers such as nannies and housepainters, that can be difficult to uncover even in an intensive audit.

Several studies have used the micro data from these IRS studies to examine the determinants of evasion. For example, Clotfelter (1983) uses the micro data from the TCMP study for 1969 and examines a regression specification of the following sort:[16]

$$ln(E) = \lambda_0 + \lambda_1 ln(Y - T) + \lambda_2 N + \lambda_3 B + \lambda_4 H + u, \tag{5}$$

where E is estimated income understatement, $Y - T$ is true after-tax adjusted gross income, N includes the marginal tax rate, B includes a dummy variable for whether or not more than four forms were filed and the ratios of wages, interest and dividends to adjusted gross income. H includes U.S. region, marital status, and age.

There are several substantive issues with this approach. First, as with any cross-sectional study of the impact of income taxes on behavior, the approach is made difficult by the fact that the marginal tax rate is a (complicated, non-linear) function of income, making it hard to separately identify the tax rate and income effects without making strong functional form assumptions. Second, a key right-hand-side variable of the Allingham-Sandmo model, p—more precisely, perceived p—is very difficult to measure.[17] We suppose that the appropriate p varies sharply depending on the type of evasion and the ability of the relevant tax authority to detect it. It is probably close to one for wages and salaries in countries with comprehensive employer information reporting and adequate computer facilities available to the tax authority, and much closer to zero for self-employment and small business income in all countries. To partially address this issue, Clotfelter (1983) estimates equation (5) separately for non-business returns, business non-farm returns, and farm returns. For the remaining

[15] Harvey Galper, past president of the National Tax Association.

[16] In practice, a Tobit model is used because E frequently takes on a value of zero. Clotfelter (1983) examines both taxable income understatement and adjusted gross income understatement.

[17] The expected punishment for detected evasion f is also difficult to measure, but arguably less so.

variation within filing class, the controls used in B are meant to partially proxy for p. The imperfect proxies both necessarily preclude obtaining the exact magnitude of the effect of p on E, and also make the marginal tax rate coefficient a function of p to the extent that p is correlated with income level. Third, this approach assumes that E measures evasion completely; that is, the auditor finds all evaded income and only evaded income, which is certainly not true in practice, and is especially problematic for unreported income and, in particular, for income tax non-filers .

Feinstein (1991) addresses both the first and third issue directly. To mitigate the first problem, he pools data from two separate years in which the tax rate as a function of income differs. He finds that this is important; the coefficient on the marginal tax rate becomes negative when two years of data are pooled, whereas both Clotfelter (1983) and he find a positive relationship when a single year of data is used.[18] To address the third, Feinstein (1991) develops a model of fractional detection, which allows E to be imperfectly detected by the auditor, where that fraction can depend on observed characteristics, such as the experience of the auditor.[19] Beron et al. (1992) attempt to address the second issue, by constructing the audit probability as the district level audit-rate. However, the district-level audit rate is not exogenous, perhaps reflecting something about the compliance-related characteristics of the population.[20]

Surveys of taxpayers are another potential way to obtain data about tax evasion directly, and such micro data exist, but despite some creative methods developed by psychologists and sociologists to address the problem of untruthful responses to delicate questions, the data are often not highly credible. Lemieux et al. (1994) argue that their survey is an exception and seek to verify that their survey obtains reasonable measures of noncompliance by examining the aggregate income-expenditure gap within the informal economy. Examining this gap is useful in understanding the magnitude of underreporting, but does not necessarily say much about the bias this would induce in the estimates, where the relevant metric is the degree of correlation between income underreporting and the independent variables of interest. Note that, while income underreporting is a substantive concern, the authors point out that such a survey may actually do better at ascertaining the true income level of individuals whose income source is such that the probability of evasion detection in an audit is very low. They analyze individuals' decisions to operate in the formal or informal economy using data from a survey of 2,134 adults in Quebec City. Measuring p remains a problem in this context; the authors include no controls to proxy for p, and note that their estimate of the effect of the marginal tax rate is a function of p. The authors take advantage of the fact that their survey includes demographic variables such as age, sex, and marital status and include these as independent variables in their

[18] However, note that year fixed effects are not used, which means that the estimates could be inappropriately attributing yearly variation in income levels to the independent variables of interest if the variables and yearly variation are coincidentally correlated.

[19] Note, though, that a complete estimate of tax evasion is obtained by this method only if some auditors are able to find all evasion, which is likely not the case.

[20] The authors instrument for the audit rate with the level of IRS resources relative to the number of returns, arguing that the IRS has not been able to allocate its resources so as to achieve its goals, but this approach is invalid to the extent that the IRS succeeds in targeting its resources toward areas believed to be particularly noncompliant or, more precisely, toward areas where enforcement is likely to be relatively more effective.

analysis. DePaula and Scheinkman (2010) make use of a survey of entrepreneurs in Brazil to examine the decision to operate in the formal or informal sector; a major focus of this study is the joint choice of firm size and whether to register with the tax authorities. Firm size in this context is an element in the vector B—it allows for endogenous selection of the firm's audit probability. The choice to register with the tax authorities is a binary measure of E—based on this choice, firms are either in the formal economy or not. Their regression puts B on the left-hand-side and E on the right-hand-side but, because both are jointly chosen, a causal relationship cannot be determined.

Research has also focused on survey evidence about attitudes toward tax evasion, sometimes called tax morale, particularly from the World Values Survey. The correlates and determinants of tax morale have been extensively studied.[21] However, without a direct measure of evasion in the survey, it is difficult to analyze how tax morale affects evasion levels. Thus, for surveys of this sort to provide quantitative evidence regarding tax evasion, in particular regarding determinants that may not be measurable in data sources such as the TCMP, it is important that they be linked to credible measures of evasion.[22]

3.2 Using Traces of True Income

For the reasons mentioned above, the brute-force approach of directly measuring tax noncompliance is rare. In its absence, the primary challenge of empirically examining the magnitude and determinants of noncompliance is that the variable of interest ($E = Y - R$) is not observed. All is not lost, however. In some data sets R (but not Y) is observed, and in other data sets, Y (but not R) is observed. In conjunction with observations of aspects of X, B, N, and H, it is possible to make inferences about the magnitude and determinants of E. But, as we emphasize below, one must proceed with caution.

We begin by considering what can be learned from micro-economic observations of R and X by focusing on a methodology used first by Pissarides and Weber (1989) and, subsequently by Feldman and Slemrod (2007), among others.[23] Consider the following functional forms for equations (1) and (2) :[24]

$$R = \frac{Y}{\alpha_0 + \alpha_1 B_R + \alpha_2 N_R + \alpha_3 H_R} + u_R \tag{6}$$

$$ln(X) = \beta_0 + \beta_1 ln(Y) + \beta_2 N_X + \beta_3 H_X + u_X. \tag{7}$$

[21] For example, see Torgler (2007).

[22] Although see Cummings et al. (2009), which links, and finds internally consistent, survey-based evidence and artefactual field-based evidence.

[23] In both cases, we use stripped-down versions of the models and ignore details that are considered more richly in the papers. For example, in the context of Pissarides and Weber (1989) we ignore that individuals may base their consumption decisions on permanent, rather than current, income.

[24] N_R and H_R refer to elements of N and H that are determinants of R, and the same is true for X.

The parameters of equation (6)—our primary interest—cannot be estimated directly because Y is unknown. Instead, by using (6) to substitute for $ln(Y)$ in equation (7), equation (7) can be written as a function of known quantities, as follows:[25]

$$ln(X) = \beta_0 + \beta_1 ln(\alpha_0 R + \alpha_1 B_R(R) + \alpha_2 N_R(R) + \alpha_3 H_R(R)) + \beta_2 N_X + \beta_3 H_X + u_X. \tag{8}$$

In Pissarides and Weber (1989), X is expenditures on food; in Feldman and Slemrod (2007), X is charitable contributions. Estimating equation (8) would not produce estimates of the α parameters, but rather a mixture of something about evasion (the α's) and something about the relationship between X and Y (the β's). If one were certain about how X varies with Y—and this assumption is made in some studies of the informal economy discussed later—the value of β_1 *implies* certain values for the α parameters. In contrast, these studies put restrictions on the α parameters instead, which would rely on the fact that B and N may not be completely observable, but are different in predictable ways across different sub-groups of earners.

Pissarides and Weber (1989) use data from the Family Expenditure Survey in the U.K.[26] Given that the data are from a survey, they rely on all individuals reporting the same value of income to the interviewers as they would to the tax authority. They categorize individuals as being either self-employed or employees and assume that 1) employees do not evade at all, 2) self-employed status (SE) is a sufficient measure for capturing the ways in which N, B, and H determine evasion for self-employed people,[27] and 3) that the food equation (8) does not differ by employment status. With these assumptions, expression (8) becomes:[28]

$$ln(X) = \beta_0 + \beta_1 ln(R) + \beta_2 SE + \beta_3 N_X + \beta_4 H_X + u_x. \tag{9}$$

The parameter of interest is $k = \frac{E[Y]}{E[R]} = \frac{1}{1-e}$, where e is the evasion rate.[29] With the assumptions noted above, $k = 1$ for employees. For self-employed individuals, $\hat{k} = exp(\hat{\beta_2}/\hat{\beta_1})$.[30]

[25] The error term in equation (6) is suppressed in this specification, but it should be noted that to the extent that it exists, β_1 would be biased due to measurement error if no instrument is used. This error would also affect the α coefficients if the bias or u_R is correlated with B_R, N_R, or H_R.

[26] Recently, Hurst et al. (2010) have applied this methodology to examine income underreporting of self-employed individuals on surveys in the U.S.

[27] The model also imposes a functional form assumption that self-employed individuals choose not to report a fraction of their income (as opposed to a fixed amount of income, for example).

[28] As Pissarides and Weber (1989) observed, R is endogenous, and therefore instruments for R must be used when equation (9) is estimated. Pissarides and Weber are concerned that SE is also endogenous, and therefore instrument for it, too. Likely, the endogeneity concerns for SE arise due to unobservable covariates that reside in the error term. They use a number of instruments for both R and E, including the number of cars per house, house value, and an indicator for whether the spouse is self-employed.

[29] For future reference, note that this procedure yields an estimated true income, and therefore an estimated value of evasion, for each individual. However, this estimate would come with very high standard errors, so it is not very helpful for assessing evasion for any particular individual. Note, though, that some tax authorities employ life-style audits, where an assessment of the taxpayers' consumption patterns and level can shed doubt on the reported income that presumably support this consumption.

[30] It follows that evasion as a percentage of true GDP is

$$E_{agg} = (\hat{k} - 1)\left(\frac{n_{SE}\bar{R}}{n_{SE}\hat{k}\bar{R} + n_{EE}\bar{Y}}\right),$$

Feldman and Slemrod (2007) avoid issues raised by survey data by examining IRS income tax return data instead, and selecting an X that appears on many tax returns in the U.S.—charitable giving. They categorize individuals by the sources of income they receive; however, unlike in Pissarides and Weber (1989), individuals can belong to more than one category if they have multiple types of income. By assuming that 1) there is no evasion on wage and salary income,[31] 2) income sources are a sufficient measure for capturing the ways in which N, B, and H determine evasion for each of these sources,[32] and 3) that the charitable giving equation (8) does not differ by tax return form, equation (8) becomes:

$$ln(X) = \beta_0 + \beta_1 ln\left(R^W + \sum_{i=1}^{n} c_i R_i^E\right) + \beta_2 N_X + \beta_3 H_X + u_X, \qquad (10)$$

where R^W is reported wage income and R_i^E is reported income from another source i, in which positive evasion is assumed possible. Under these assumptions, the c_i's are informative as to the absolute rates of evasion for different income types. In this context and in Pissarides and Weber (1989), if the evasion rate of those that earn wage and salary income is not assumed to be zero, the values of c_i and k, respectively, are informative regarding the *relative* evasion rate between wage and salary income and another form of income considered.

Both of these approaches assume that individuals do not alter X to change their probability of being audited. But, if X is observed by the econometrician, presumably it can be observed by the tax authority; in Feldman and Slemrod (2007), it is actually reported to the tax authority. This provides some motivation for individuals to alter X to lower their chance of being audited. To see this, suppose that a high ratio of charitable giving to reported income increases (or is suspected to increase) the probability of audit (an element of N). Then individuals will likely report (and donate) lower values of X than otherwise. In this case, regression analysis of the system above would suggest low values of evasion, when in fact there may be substantial evasion accompanied by substantial underreporting of X.[33]

The parameters c and k are reduced-form, reflecting the cumulative effect of all elements of N, B, and H that affect evasion levels and differ between income or employment type. By continuing to split income into more and more categories, one could learn about a more narrowly defined reduced-form effect. A possible advantage of this approach is that the researcher need not measure the policy determinants of interest, as long as one knows what they are and how they vary by income source.

---

where n_{SE} is the number of self-employed, n_{EE} is the number of employees, $\bar{R}$ is average reported income of the self-employed, and $\bar{Y}$ is average true income of employees.

[31] The premise of this assumption is that individuals likely do not evade on wage and salary income, because the withholding and informational reporting requirements on wage and salary income in the U.S. make it extremely difficult to successfully evade.

[32] The argument of footnote 27 applies here, too.

[33] Glazer and Konrad (1996) suggest another possible concern, namely that self-employed people will, conditional on other factors, donate relatively more to charities, so as to appear civic-minded. This would bias estimates of self-employed noncompliance upward. Additionally, this procedure would yield biased estimates if the measures of income for tax purposes are *intended* to be different as between employees and self-employed individuals.

This may be particularly useful in the context of audit and detection probabilities, which, as we discussed in section 3.1, are extremely difficult to measure directly. The drawback of this approach is that it is impossible to differentiate the effect of potential determinants from their actual change across income sources. Neither of these approaches address the fact that individuals may select into being self-employed or receiving income in a particular form because they have a higher preference for evasion. If these unobserved preferences are correlated with charitable giving, then the third assumption above is violated and the estimates would be biased. If not, the parameters can be interpreted as the effect conditional on this selection, but it would still be difficult to justify extrapolating the results across income groups.

3.3 Measuring Traces of Noncompliance

Another approach to learning about tax noncompliance when no direct measure exists relies on traces of noncompliance. We introduce this method with a non-tax example.[34] Suppose one were interested in learning about the effect of policy variables on the extent to which drivers exceed legally posted speed limits ("speeding"). Speeding (E) is not observed, but we do observe the deployment of radar detectors. Owning a radar detector makes sense only for those who intend to speed; on the other hand, many speeders do not own a radar detector. Let there be two policies that potentially influence speeding, the speed limit and the extent of police monitoring, that vary by time and jurisdiction (both are elements of N).

One might like to run the following simple regression:

$$E = \lambda_0 + \lambda_1 N + \lambda_2 H + u_E, \tag{11}$$

but this is infeasible because E is unknown. Instead, one might make use of the relationship between unobserved E and the observed element of B, as given by equation (3). Taking derivatives of both sides yields:

$$\frac{dB}{dN} = \frac{\partial B}{\partial E}\frac{dE}{dN} + \frac{\partial B}{\partial N}. \tag{12}$$

By estimating the following reduced-form regression:[35]

$$B = \pi_0 + \pi_1 N + \pi_2 H + u_B, \tag{13}$$

one would like to learn about the sign and the relative magnitude of the reduced-form effect of two different policies on E.[36] If one also knew the sign of $\frac{\partial B}{\partial E}$ and

[34] See Stanley (1995) for an elaboration of this analogy. In the context of noncompliance, traces studied include such things as cash holdings or bank accounts in tax haven countries.

[35] Ideally, one would include jurisdiction fixed effects, so as to identify the parameters off of changes in jurisdiction enforcement or speed limit policies. Otherwise, it is extremely difficult to draw causal inference, because jurisdictions likely have varying preferences for evasion, speed limits, and police enforcement that are correlated with a variety of unobservable characteristics (that is, one cannot control for all important elements of H).

[36] We do not even consider the possibility of obtaining a clean estimate of the reduced-form impact of N on E, since it would only be possible if $\frac{\partial B}{\partial E} = 1$ and $\frac{\partial B}{\partial N} = 0$.

that $\frac{\partial B}{\partial N} = 0$, one could proceed to use the coefficients estimated in equation (13) to learn about the sign of λ_1 and the relative magnitude of the effect of the two different policies.[37] Using the example of police enforcement, it is straightforward to assume that $\frac{\partial B}{\partial E} > 0$—as speeding increases, so does the use of radar detectors. However, the element of N considered in this example is clearly one for which $\frac{\partial B}{\partial N} \neq 0$: as police enforcement increases, the incentive to use a radar detector increases holding speeding fixed. The direct effect on E is to decrease the incentive to speed. It is possible that the latter effect, when multiplied by $\frac{\partial B}{\partial E}$, is actually larger than the former and, if so, the sign of π_1 would be incorrect. Therefore, drawing causal inference should not be done, in general, when the dependent variable is a trace of evasion and $\frac{\partial B}{\partial N} \neq 0$.

An additional problem with this approach is that if some component of B is a good indicator of evasion, and it is observable, then presumably the tax authority could use it to detect evasion, have sanctions against B, or investigate B as part of evasion investigations.[38] One can even imagine a presumptive tax on B or, as in Gordon and Li (2009), a subsidy on substitutes for B. To the extent that components of B trigger tax liability, they will be used less frequently in equilibrium. As a result, it is unlikely that a component of B is both a good indicator of evasion and not too difficult for the researcher to measure accurately. As we discuss in section 4, in some cases it is possible to obtain a reliable macro estimate of a variable even though measuring it at the individual or firm level is difficult.

Another "traces of evasion" strategy is to examine data patterns that can be explained only by evasion.[39] One example is to apply Benford's Law, a property that provides expected frequencies of the digits in tabulated data, as done by Nigrini (1996). Slemrod (1985) uses the behavioral response to the $50 notches in the U.S. individual income tax rate schedule as an indicator of tax evasion. These methods can be informative with regard to identifying the existence or determinants of noncompliance, but are not likely to be helpful in quantifying the magnitude of either of these.

3.4 The Promise of Randomized Field Experiments

Randomized field experiments offer another source of evidence, and are especially promising for identifying the causal effect of potential evasion determinants. Indeed, this approach to inference has been prominent in the "credibility revolution" referred to by Angrist and Pischke (2010). Due to the randomization of the causal variables, a well-constructed field experiment promises a high degree of internal validity for the policy treatments that are varied. Conducting an experiment in this context also

[37] Even though we are only interested in the relative magnitude of two policies, all elements of N and H that are correlated with the policies of interest need to be included as independent variables, or else the relative magnitude of the effect of the two policies would be biased. The only exception is an extreme case in which the omitted variables generate a bias that rescales both coefficients of interest by the same proportional constant.

[38] Radar detectors are illegal in many countries as well as in Virginia and the District of Columbia.

[39] This approach is related to the broader topic of "anomaly detection," used to assist, for example, in the detection of credit card fraud; see Chandola et al. (2009).

allows researchers to side-step the problem of an unknown dependent variable; given random assignment, a difference-in-differences strategy can be employed, such that the average change in reported income before and after treatment of the treatment relative to the control group is the change in evasion due to the policy change.

An early example of the use of field experiments to study noncompliance is Slemrod et al. (2001), who analyze the results of a randomized, controlled experiment conducted by the State of Minnesota Department of Revenue. The treatment group received a letter which informed them that the returns they filed would be "closely examined." Kleven et al. (2010) designed a recent experiment in collaboration with the Danish tax collection agency. They implement a two-year experiment, where in the first year, individuals were randomly selected for a comprehensive audit, and in the second year, individuals were randomly selected to receive "threat-of-audit" letters. In both of these studies, there is at least one treatment that is designed such that the audit probability post-treatment should be believed to be one. If detection conditional on being audited is also one, then this treatment will provide a measure of total evasion in the absence of treatment, as complete compliance is optimal in a deterrence model when detection is certain. Fellner et al. (2009) set up a randomized field study in collaboration with the Fee Information Service in Austria to study a different type of evasion, namely whether individuals register and thus pay for receiving public broadcasting, which is required by law. The main results measure the effect of legal threat letters on the decision to register among individuals who were suspected of evading.

A key concern for field experiments is external validity, both for an economy-wide implementation of the piloted treatment and for variations on the precise treatment studied. For the former, general equilibrium effects will likely cause the economy-wide effect to differ. For the latter, extrapolating the impact of the pilot to a somewhat different policy requires some knowledge of *why* the policy change matters. Furthermore, the search for explanations of the observed responses often leads to examinations of determinants that were not randomized. For example, Kleven et al. (2010) examine the differential effect of the change in audit probability on types of income that do, and do not, face third-party reporting. Part of the differential effect that they find is likely due to different perceived initial probabilities of audit and may also be affected by different underlying preferences for evasion across the two groups. As noted by Kleven et al. (2010), valid causal inference regarding the effect of third-party reporting would exist only if it were also randomized. We also note field experiments are only conducive to examining certain types of policy questions; for example, it is difficult to measure the effect of varying levels of confidence in the government.

Although field experiments promise an unrivaled degree of internal validity, they are costly to run, both in terms of time and money. One way to obtain further insight from a field experiment is to use it to evaluate the internal validity of standard non-experimental results. The seminal paper on this topic, written by LaLonde (1986), examines non-experimental estimators using the National Supported Work Demonstration (NSW) experiment. The method compares the control group from the experiment to a reasonably constructed comparison group (which would be used if the experiment had not been conducted). If the estimated coefficient on an indicator variable for being in the control group is reasonably close to zero, this provides evidence that unobservable differences between the two groups are not inducing omitted variable biases. This, in turn, suggests that the non-experimental method examined could be used to obtain valid inference in settings similar to that examined. A similar approach could be implemented in the context of tax evasion. For example, suppose one wanted to know whether the method used by Feldman and Slemrod (2007), which was discussed in subsection 3.2, produced valid estimates. One could ask whether individuals who chooose to earn income from sources where evasion is more or less feasible are the same in terms of unobservables that affect their marginal propensity to give to charity.[40] Using the data generated in the experiment conducted by Slemrod et al. (2001), one could investigate this question by examining all taxpayers who were either in the experimental treatment group (received a letter informing them that they would receive an audit) or a slightly restricted version of the non-experimental comparison group (individuals who earned only wage and salary income and were not included in the experiment) and running the following regression:

$$ln(X) = \beta_0 + \beta_1 ln\left(R^W + \sum_{i=1}^{n} c_i R_i^A\right) + \beta_2 N_X + \beta_3 H_X + u, \tag{14}$$

where R^W now includes all individuals in the restricted non-experimental comparison group and R_i^A is income split by source of all individuals in the experimental control group.[41] Then, one would fail to reject the null hypothesis that the model was valid if c_i was not statistically different from one for each income source. There are several caveats regarding the application of this method in this context. Most importantly, the reason why wage and salary income is truthfully reported is a function of both the probability of audit conditional on evasion and the likelihood that all the evaded income will be detected if an audit is conducted (both of which are likely close to one for wage earners). The experiment effectively alters the probability of audit, but leaves the probability of detection unchanged. For this reason, the marginal propensity to give to charity out of R^A and R^W may systematically differ, even if the non-experimental method is valid. Additionally, this method is not effective at detecting whether or not the non-experimental estimates are biased because individuals lower their reported charitable giving contributions when they evade to decrease the probability of audit. When individuals in the experiment receive a letter telling them that they will be audited with a probability of one, they may well decide to truthfully report both income and charitable donations.[42] In this case, one may fail to reject

---

[40] Note that this is a slightly different question than that being asked in the labor context. In the latter case, the question is whether individuals who selected into treatment (by signing up for the experiment) are the same on all unobservables as those who did not. This makes it natural to compare the experimental control group (which has selection but no treatment) and the non-experimental comparison group. In the tax evasion context, everyone in the experimental and non-experimental control groups are the same—both include individuals that have selected into each income source. The closest analog is to examine individuals who receive the experimental treatment (now they have incentive to report their true income, but still have the unobservables associated with selection into income sources where evasion is more feasible) and the non-experimental control group (who have selected into income sources with a high cost of evasion).

[41] Wages and salaries are an income source in this context, and the estimated coefficient on this source provides evidence regarding the assumption that there is no evasion present in wage and salary income.

[42] This description is likely consistent with reality if the charitable giving report was false and its purpose was solely to decrease the probability of audit and not the detection rate once the audit was conducted.

the null that the non-experimental method is valid even though it is in fact biased. Thus, while comparisons between field experiments and non-experimental methods should be a part of the credibility revolution going forward (with an experiment that is designed to more closely replicate the non-experimental setting), it remains true that there are some things one would want to study in a non-experimental setting that one finds it very difficult to examine in a non-experimental setting (e.g., a direct examination of the effect of a change in audit probability), and this provides an additional justification for field experiments even if non-experimental studies provide valid estimates of evasion and some causal determinants.

3.5 Lab Experiments

A distinct advantage of lab experiments is that researchers have more latitude to adjust the environment the taxpayer faces relative to a field experiment. This enables researchers to study whatever is of interest, whether it be something concrete like an audit probability or a more attitudinal variable like institutional uncertainty (e.g. Alm et al. (1992)), and at the same time hold everything else constant. However, there are also potential disadvantages of this method. While internal validity is high, external validity is often problematic in this context, because it is not immediately obvious that individuals would respond the same way in a controlled lab experiment that they would if a similar policy were implemented in practice. This is both due to the fact that individuals may behave differently in a lab setting and that there may be additional unobservables that play an important role in the effect of a given policy. While the lab experiment may give a cleaner estimate of the effect of a given policy by itself, it may not do much in terms of explaining how such a change would actually alter evasion in practice. Therefore, a potentially valuable approach is to design a lab experiment in such a way that it has an out-of-the-lab counterpart, which would facilitate an evaluation of the credibility and information value of lab experiments. For example, Alm et al. (2010) compare tax compliance estimates from TCMP data to those generated by a lab experiment, and Cummings et al. (2009) compare lab experiment results to survey responses. Of course, the caveat of any such comparison is that both may be subject to substantive concerns. Finding similar estimates across the two methods may suggest their validity or may suggest that the biases in each are such that the estimates turn out to be roughly the same. However, given the inherent difficulty of observing exogenous variation of certain potentially important aspects of the relevant environment outside of a lab, careful lab experiments remain the best method to gaining broad insights, if not policy-relevant parameters, about the impact of such factors as taxpayers' trust in government.

4 Macro-Based Inference about Noncompliance

Suppose instead one would like to draw inferences about the magnitude of aggregate tax evasion or the informal economy, either across countries, or within a country or countries over time. Perhaps one is also interested in understanding the effect of policy variations on these variables, especially if there are policy variables that vary across time or countries, but not across individuals at a given time within a given country. There are special challenges for macro-based inference. The sample size is inevitably much smaller than in micro-based studies and there are certain to be country-specific variables that affect noncompliance but are difficult, if not impossible, to measure and which are correlated with the measurable determinants. Unlike in micro-based inference of tax evasion, we do not observe an analogue to "reported income."

4.1 Aggregating Micro Data to get Macro Estimates

One approach is to make use of the TCMP/NRP style studies, and simply add up the micro estimates to get aggregate estimates, as the U.S. tax gap studies do. The problem with this approach is that, given the scale of the program required (and perhaps its political sensitivity), no other tax authority has done such a comprehensive study–nor allowed scholars' access to similar data–and so it is impossible to draw inferences from cross-country studies of this nature. Even for the U.S., where these studies were done regularly for many years, it is difficult to draw inferences over time, because the methodologies used have varied. The limited usefulness of these estimates for these purposes is not really a surprise, because the main objective of the TCMP/NRP is not to come up with aggregate "tax gap" measures of the magnitude of evasion or of the nature of evasion. It is rather to help identify returns that are more likely to feature evasion, so as to guide the allocation of enforcement, usually in the form of audit resources. The International Labour Organization takes a different form of micro data—surveys—and aggregates the results by country to focus on a particular aspect of the informal economy, namely employment.[43] A focus of their work is to obtain estimates that are comparable across countries, which is crucial if these estimates are to be used for any meaningful analysis.

4.2 Traces of True Income

A large literature has analyzed what we call traces of true income to shed light on the informal economy rather than tax evasion. A standard focus of attention is electricity use.[44] Electricity use by firms is an input to production, and as such is a derived demand from value added. Electricity use by final consumers is arguably a function of true income, as were food and charity in section 3.2. Thus, we may consider it to be an example of an element of X in equation (2), and expect it to be positively correlated with true income or output across countries. Moreover, electricity is a good example

[43] See Hussmanns (2004). The OECD has also done a recent analysis, building on Hussman's work, in Jütting and de Laiglesia (2009).

[44] For example, see Johnson et al. (1997) and Lackó (2000). Also, Henderson et al. (2009) develop a new trace of true income–satellite data on lights at night. While the primary purpose of that paper is to develop a statistical framework that uses lights data to augment existing measures of true income, it could also be valuable in the framework discussed here if the assumptions outlined below in the context of electricity also hold for lights at night.

of something where aggregate use might be measurable with some accuracy, even while use by an individual or firm is difficult to ascertain. Intuitively, a high ratio of aggregate electricity use to formal income is an indication of a relatively large informal sector, just as a high ratio of expenditure on food or charitable giving to reported individual income is an indication of relatively high underreporting of true income. Thus, the insights and caveats drawn from our earlier discussion using micro data apply. Note, though, that aggregate studies often seek to estimate the extent of the informal economy for each country, which is not generally the objective of micro analysis of tax noncompliance, except by the tax and regulatory authorities themselves.

For now, suppose one is interested in the determinants of the size of the informal economy across countries. Estimation and inference could be conducted in the same manner as done in Pissarides and Weber (1989) discussed above, except that observations are aggregated to the country level, GDP is used instead of reported income, and *SE* becomes an indicator for countries that the researcher expects have high levels of informal economic activity. The same assumptions need to hold in order for the estimation method to be valid. Now, k is interpreted as the relative rate of informal economic activity between these two types of countries. As discussed in subsection 3.2, this method is good for uncovering the rate of noncompliance when there is a variable that clearly delineates evading individuals (or in this case countries) from non-evading individuals. It is not clear what this variable would be in a cross-country context and, possibly for this reason, this method has never been implemented in the macro literature to our knowledge.

As a possible alternative, suppose one posits the following functional forms for equations (1) and (2):

$$E = \lambda_0 + \lambda_1 N_E + \lambda_2 H_E + u_E \tag{15}$$

$$ln(X) = \beta_0 + \beta_1 ln(R+E) + \beta_2 N_X + \beta_3 H_X + u_x, \tag{16}$$

where X is electricity usage, R is official income (likely GDP), E is informal income, N_E and H_E are the elements of N and H that are causes of noncompliance, and N_X and H_X are elements of N and H that are determinants of electricity demand. As E is not observed, one could substitute for E in equation (16), which yields:[45]

$$ln(X) = \beta_0 + \beta_1 ln(R + \lambda_0 + \lambda_1 N_E + \lambda_2 H_E) + \beta_2 N_X + \beta_3 H_X + u_X. \tag{17}$$

Several key assumptions are needed in order for this method to provide consistent estimates of the causal effects of N_E and H_E on E. Just as with other trace of true income strategies, it is important that β_1 is the same, on average, for both official and informal income. In a cross-country context, this means that if the underlying relationship between electricity demand and GDP varies, it cannot vary systematically with noncompliance levels in these countries. Similarly, in a cross-time context within a country, the relationship cannot vary in a systematic way over time that is

[45] The error term in equation (15) is suppressed in this specification, but it should be noted that to the extent that it exists, β_1 would be biased due to measurement error. This error would also affect the λ coefficients if the bias or u_E is correlated with N_E or H_E.

also correlated with noncompliance levels. Further, the elements of N_E and H_E cannot have a direct effect on X beyond their role as a determinant of noncompliance unless they also show up as control variables in equation (17) (e.g., holding noncompliance constant, a country's tax burden level cannot causally alter the demand for X). Additionally, while the vectors N_X and N_E could technically have some overlap (and the same is true for H_X and H_E), separate identification of these two effects would rely on strong functional form assumptions.

One might also be interested in the predicted level of noncompliance for each country at each point in time, which is given by:[46]

$$\hat{E} = \lambda_0 + \hat{\lambda}_1 N_E + \hat{\lambda}_2 H_E. \tag{18}$$

Note that if, after obtaining these predicted estimates, the researcher was interested in examining an additional determinant of noncompliance, the original regression specification should be altered to include this determinant. This new specification would provide evidence regarding the causal effect of this determinant and new predicted values that could be used to create a more precise measure of noncompliance. An alternative approach would be to use the noncompliance estimates obtained from the original regression as a dependent variable to analyze causes of noncompliance. However, there is a basic conceptual problem with this method. Let N_2 include all the elements of N that were *not* included in the original model, but would now be examined in this regression. First, note that the original model could be improved by including the elements of N_2 directly as long as they do not violate the assumptions of the method. The exclusion of variables that cause noncompliance would, in general, lead to incorrect inference and incorrect predicted values of noncompliance in the original model. Due to the latter problem, correct inference cannot be carried out in this setting either. To see why, consider the regression one would be running:

$$\hat{E} = \gamma_0 + \gamma_1 N_2 + \gamma_2 H + u, \tag{19}$$

where the dependent variable is calculated using equation (18). Since the elements of N_2 were not included in the original regression, the dependent variable does not include the predicted value of these elements, except to the extent that the elements of N_2 are correlated with the variables in N_E and H_E that were included in the original model. Therefore, this regression would simply pick up this correlation, which is not very informative, and only provides valid inference in the extreme case where the variables originally included form a perfect proxy for N_2. Such estimates are useful neither for confirming ex-ante hypotheses nor for learning additional information about what factors cause the size of the informal economy to differ across countries.

The literature on this topic has not employed the estimating equation given by equation (17). Lackó (2000) pursues a related approach but, instead of estimating

[46] Note that the standard errors on each country's noncompliance level for each time period would be $\sqrt{n}$ times as large as the standard errors of the average estimated noncompliance level for all countries and time periods, where n is the number of country-time observations in the sample.

(17), estimates an equation akin to the following:[47]

$$ln(X) = \beta_0 + \beta_1 ln(C) + \beta_2 N_X + \beta_3 H_X + \beta_4 E + u_X, \tag{20}$$

where C is per capita consumption of households at purchasing power parity. Substituting for E in equation (20) yields:

$$ln(X) = \beta_0 + \beta_1 ln(C) + \beta_2 N_X + \beta_3 H_X + \beta_4 \lambda_0 + \beta_4 \lambda_1 N_E + \beta_4 \lambda_2 H_E + u_X. \tag{21}$$

It is immediately clear that the λ coefficients of interest cannot be separately identified from β_4, a problem that was not encountered when estimating equation (17).[48]

Johnson et al. (1997), building on work by Kaufmann and Kaliberda (1996), pursue a simpler approach in their examination of post-communism transition countries. They use electricity use as a proxy for total economic activity and GDP as a measure of formal activity, which allows them to back out a measure of informal economic activity directly.[49] They then estimate a regression of the following sort to analyze the determinants of noncompliance:

$$E = \lambda_0 + \lambda_1 N_E + \lambda_2 H_E + \lambda_3 H_X + u_E. \tag{22}$$

While Lackó (2000) and others have pointed out that the Johnson et al. (1997) measure of noncompliance is imprecise due to a variety of other factors that may play a role in determining the gap between electricity usage and official GDP, it is not necessarily flawed as a dependent variable in a regression that studies the causal determinants of informal economic activity if additional controls H_X are used to control for all factors that may provide alternative explanations for the gap which are correlated with other causal variables of interest.[50]

Of course, there is a substantial possibility that unobserved covariates would bias the results, regardless of which method is used. Robinson and Slemrod (2010) provide evidence that such a bias likely exists by showing the correlation between covariates that were previously unobserved—a variety of tax system measures—and previously observed, studied covariates. Although some empirical work recognized the joint importance of tax rates and tax administration, the dearth of empirical proxies for most aspects of tax systems has limited the ability of researchers to pursue these issues. This data vacuum has now been filled by the Organisation for Economic Co-operation and Development (OECD (2006), OECD (2008)) publication of

[47] It is not clear what underlying model would be consistent with a specification where C, which is a function of both R and E, is in logs, but E is not. Also, note that Lackó (2000) is only examining household electricity use.

[48] Lackó (2000) attempts to back out noncompliance from equation (21), but this method faces additional challenges due to the fact that the λ and β coefficients cannot be separately identified.

[49] Johnson et al. (1997) classify the countries into several types, and for each type an electricity-to-output elasticity is exogenously assigned to convert electricity use to a proxy for total economic activity. The elasticities are all near or equal to one.

[50] In practice, Johnson et al. (1997) use only two H_x variables: the level of informal economic activity in 1989 and an indicator for whether or not the country belonged to the Soviet Union. These controls are likely not enough and, if not, some omitted variable bias remains. Note that the same need for including appropriate independent variables H_X applies to all the methods discussed in this section.

a careful cataloguing of scores of tax system aspects of up to 47 countries' tax administration. Robinson and Slemrod (2010) construct summary measures for a subset of the OECD-assembled information. As expected, they find that tax rate measures are also correlated with other tax system aspects, so that including only tax rates as an independent variable presents the danger of assigning to tax rates what is really the impact of other, unrelated tax system aspects, such as the extent of enforcement. The lack of such measures also precludes any investigation of non-rate, non-base effects on outcomes.

For covariates that remain unobserved and do not change over time within a country, a natural approach is to estimate the relationship in first differences or with country fixed effects. With country fixed effects, identification comes only from policy variation within countries over time, although the starkest policy variations are likely to be across countries. Once the OECD data have been collected for a long time period, researchers will be able to pursue cross-country longitudinal analysis of the effects of a wide array of tax system aspects subject to the measurement-related caveats discussed elsewhere in this paper. If characteristics and policies within a country change together, then within-country longitudinal analysis may provide a relatively small gain relative to cross-country analysis. However, if country characteristics evolve slowly while policy changes abruptly, treating a policy change within a country as a natural experiment is a reasonably good approximation, conditional on being able to control for any other factors that affect the dependent variable that occur in the same year. Another way of addressing remaining endogeneity of the policy-relevant variables is to employ instruments; however, finding convincing, strong instruments for each tax policy variable is not an easy task.

Up to this point, we have assumed that R was measurable. When the focus of analysis is income tax evasion, R has the natural interpretation of being reported income, based on tax returns as filed. When the subject of analysis is the informal economy, R is better thought of as a measure of the formal economy. The researcher is usually constrained to use figures from national income accounts, such as GDP. But GDP from national income accounts is not an appropriate measure of the size of the formal economy, because most countries have some procedure for adjusting GDP to include an estimate of the informal economy.[51] If national income accounts statisticians already accurately estimate both the formal and informal economy (and separate them in published data), then the kind of procedure described here is obviously unnecessary, as the size of each country's informal economy is known and published, and these data could be used directly by researchers to assess the determinants, policy and otherwise, of the informal economy. More realistically, the success of official GDP in measuring the informal economy varies significantly across countries and may also vary within countries over time as countries revise and improve their national accounts procedures to account for the informal economy.[52] Thus, GDP is neither true income, a measure of the formal economy, nor a trace of true income, and is better

[51] See Charmes (2006).

[52] Note that in practice countries revise their procedures relatively infrequently so as to maintain a substantial degree of comparability over time Charmes (2006)

described as

$$GDP = R + m_{it}(Y - R), \tag{23}$$

where m_{it} varies across countries and time. In the context of the electricity example above, this mismeasurement would yield incorrect estimates of the relative rate of informal economic activity across countries. Instead, the estimates would measure, in part, the relative ability of the two types of countries to correctly adjust GDP to account for the informal economic sector.

4.3 Traces of Noncompliance

A substantial literature focuses on what can be learned from a macroeconomic trace of evasion or informal economic activity. Currency is the most prominent example, based on the plausible argument that currency, and especially large-denomination bills, is of particular value to launder illegal transactions or evade regulations. Feige (1990) estimates the size of the underground economy by assuming that most unreported economic activity takes place in cash, and that there is a "base year" when the underground economy did not exist. Similarly, Tanzi (1980) (1980, 1983) pursues a methodology that defines B as the ratio of currency to money more broadly defined as M_2. He interprets the portion of this ratio explained by changes in the tax level as an indication of changes in the size of the underground economy. In particular, a regression of the following form is estimated:

$$ln(B) = \beta_0 + \beta_1 ln(N) + \beta_2 ln(H) + u. \tag{24}$$

If one were interested in learning about the causal effect on the informal economy of policy determinants N from this regression, the same considerations as outlined in subsection 3.3 apply. Recall that even under the best of circumstances, it is usually possible to obtain at most the sign and relative magnitude of causal variables when using a trace of noncompliance, while the absolute magnitude of the effect of these variables may be obtained if traces of true income are used instead.

To obtain estimates of the informal economy, Tanzi and others assume that the currency demand in the formal sector C^R is given by:

$$C^R = exp(\hat{\beta}_0 + \hat{\beta}_2 ln(H)) * M_2. \tag{25}$$

That is, they assume that the preference determinants H only affect formal currency demand (and have no effect on currency demand in the informal sector) and the policy determinants N have no direct effect on formal currency demand, both of which are quite strong and easily violated in practice. For example, Tanzi (1983) uses tax burden as a measure of N, which likely affects the size of both the formal and informal economies, and uses the ratio of wage and salary income to total income as a measure of H, which is indicative of the efficacy of noncompliance enforcement policies such as informational reporting. Furthermore, this method assumes that all determinants of currency demand in the formal sector are included and that there are none from the informal sector which are excluded and correlated with the included determinants of formal sector demand. Now, $C^I = C - C^R$ and the velocity of money V is calculated as $V = \frac{R}{C^R+D}$. V is assumed to be the same across the formal and informal sectors.[53] Given these assumptions, E is derived as $E = V * C^I$.

Note that if there is an additional element of N that a researcher wishes to analyze but was not included in the original estimation, the researcher should return to the original regression and include this causal factor in the analysis. Regressing the estimates of E obtained above on causal variables N is not advisable in general. The new analysis cannot do better than the original regression: that is, sign and relative magnitude are the most information that could possibly be gained from these regressions, unless one believes literally each of the assumptions made in deriving E. Furthermore, if the elements of N included in this regression were not included in the original analysis, the coefficients on these elements would be biased if these elements of N were correlated with the elements of H included in the original regression.

4.4 Using Traces of Noncompliance, Traces of True Income, and GDP

The most complex empirical approach to measuring the informal economy and its determinants at a country level makes use of information about traces of true income, traces of noncompliance, measures of official GDP, and estimates generated by studies in subsections 4.2 and 4.3. This is a latent variable (LV) approach, also known as MIMIC (multiple-indicators, multiple-causes) modeling, developed by Zellner (1970) and first applied to the informal economy by Frey and Weck-Hanneman (1984). We examine MIMIC as a system of simultaneous equations, which can be estimated using iterated generalized least squares for seemingly unrelated regressions.[54]

The MIMIC literature distinguishes between "causes" and "indicators" of the informal economy. Causes line up fairly well with the vectors H and N in our model, where H refers to aspects of the non-policy environment and N refers to aspects of the tax and regulatory systems. Indicators comprise a conceptually fuzzier set of variables, although as we discuss below each must behave as a trace of noncompliance due to the estimation strategy employed.

One would like to estimate the following regression:

$$E = \lambda_0 + \lambda_1 N_E + \lambda_2 H_E + u_E, \tag{26}$$

but E is not observed. Instead, suppose one observes two traces of noncompliance given by the vector B. Additionally, suppose all possible variables from the vectors N and H, that are expected to causally affect noncompliance, denoted N_E and H_E, are included as independent variables.[55] One could then write down a system of linear

[53] These assumptions have been extensively criticized in the literature (e.g. Thomas (1999) and Schneider (2005)).

[54] Breusch (2005) shows that this specification is equivalent to MIMIC, when there are two or fewer indicator variables. When there are more than two indicator variables, there are additional covariance restrictions that cannot be imposed using a system of simultaneous regressions. Then, MIMIC must be estimated using maximum likelihood. We choose to focus on the simultaneous equations example because the intuition, when expressed in this way, is more in line with the discussion in the rest of our paper. Moreover, the points we make in this section apply to the more general maximum likelihood case as well.

[55] Given the discussion in a previous subsection, one would likely estimate this regression in differences, in order to eliminate the potential for bias arising from country-specific unobservable variables that do not change over time.

equations for these traces as follows:

$$B = \beta_0 + \beta_1 E + \beta_2 N_B + \beta_3 H_B + u_B. \tag{27}$$

It is in this sense that each element of B must *behave* as a trace of noncompliance; if it does not, $\beta_1 = 0$, which negates any further inference regarding E. Note that in practice B often includes GDP or the growth rate of GDP, which requires formal GDP levels to be directly influenced by noncompliance. Note that if one substituted for E from (26) into equation (27) one could not identify the β parameters if there was any overlap between N_B and N_E or H_B and H_E because these variables appear twice in the same linear regression. To resolve this problem, the MIMIC literature narrows its focus to a subset of causal variables, $N_E^{'}$ and $H_E^{'}$ that the researcher can plausibly claim affect B only through E. One could substitute for E from (26) into (27) and estimate:

$$B = \beta_0 + \beta_1\lambda_0 + \beta_1\lambda_1 N_E^{'} + \beta_1\lambda_2 H_E^{'} + \beta_2 N_B + \beta_3 H_B + \beta_1 u_E + u_B. \tag{28}$$

Note that this estimating equation is very similar to equation (24), which was discussed in subsection 4.3. The reduced-form of (28) is given by:[56]

$$B = \pi_0 + \pi_1 N_E^{'} + \pi_2 H_E^{'} + \pi_3 N_B + \pi_4 H_B + u', \tag{29}$$

where π_1 and π_2 are estimates of $\beta_1\lambda_1$ and $\beta_1\lambda_2$ in equation (28), respectively. While one cannot empirically determine the relationship between B and E, it does seem reasonable to assume that the sign of β_1 is known. With this assumption, one could learn the sign of λ_1 and λ_2 and the t-statistics for π_1 and π_2 would provide a lower bound on the t-statistics for λ_1 and λ_2. However, these estimates are only informative to the extent that other possible determinants of B and E have not been excluded, either because they cannot be measured or because they violate the exclusion assumptions mentioned above.

Note that, in practice, there are usually no elements of N_B and H_B employed in estimation.[57] This presumes that any determinants of B that do not cause noncompliance are uncorrelated with determinants that do cause noncompliance. This is a very strong assumption that need not be imposed, and is not by the studies considered in either subsection 4.2 or subsection 4.3. It is particularly strong in the case of Schneider et al. (2010) because the equations are estimated in levels, not differences. The assumption that variables included in $N_E^{'}$ and $H_E^{'}$ include only those that could plausibly affect B exclusively through their effect on E is frequently violated in practice.[58] These violations have real consequences. By the same logic as given in subsection

[56] In practice, the variables in equation (28) are often standardized or transformed into deviations from the mean. For a discussion of the additional issues that may arise depending on the exact implementation, see Breusch (2005). For a response, see Dell'Anno and Schneider (2006). Throughout this paper, we assume no such transformation has taken place.

[57] For example, see Schneider (2005) and Schneider et al. (2010).

[58] Three examples in Schneider et al. (2010) are: 1) an element of B is the labor force participation rate and an element of H is the unemployment rate (these two variables have a direct, mechanical relationship), 2) an element of B is M_0/M_1 and an element of H is the inflation rate, and 3) an element of B is GDP per capita and an element of N is the unemployment rate.

3.3, whenever elements from the vector N (or H) are included in a regression using a trace of noncompliance, neither the sign nor the relative magnitudes of two variables can necessarily be inferred when $\frac{\partial B}{\partial N} \neq 0$. Furthermore, if these estimates were subsequently used to back out noncompliance, the predicted values would include these variables that in fact only partially, if at all, predict noncompliance.

As with most of the approaches considered in this section, the most daunting endeavor is to attempt to use the regression estimates to back out an estimate of the size of the informal economy. The way in which these estimates are obtained varies by study and we will not discuss these issues extensively here.[59] For our purposes, it is sufficient to note that the estimates of noncompliance combine predicted estimates of noncompliance implied by estimating equation (29) and estimates from studies discussed in sections 4.2 and 4.3.[60] This makes the estimates nearly impossible to interpret, since the estimates for each country are a function of other estimates, where the exact model used to obtain these other estimates varies by country. Furthermore, assumptions used (by other researchers) to obtain these other estimates are often directly violated in MIMIC.

Beyond inferring insights about the causes of noncompliance from estimating a MIMIC model, some researchers employ the MIMIC model estimates of the shadow economy as the dependent variable in regression analyses in order to determine what causes noncompliance. This approach is concerning because neither the predicted estimates obtained from equation (29) nor the predicted estimates from subsections 4.2 and 4.3 are valid dependent variables.[61] Furthermore, the combination of these methods makes the endeavor even more concerning, since it is possible that the varying ways in which the estimates are calculated by country may be correlated with the causal variables of interest. While estimates obtained from such an analysis may appear reasonable ex-post, they are not interpretable as estimates of any causal effect. They are useful neither for confirming ex-ante hypotheses nor for learning additional information about what factors cause the size of the informal economy to differ across countries.

5 Closing Thoughts

In their contribution to a symposium on applied econometrics in the *Journal of Econometric Perspectives*, Angrist and Pischke (2010) assert that, in the last twenty-five years, empirical microeconomics has experienced a "credibility revolution" due largely to a more careful focus on the quality of research design and an emphasis on randomized trials and quasi-experimental studies. In this paper, we argue that, with regard to the empirical analysis of tax evasion and the informal economy, the credibility revolution has, for the most part, not yet arrived. The late arrival of the credibility

[59] For an extensive discussion regarding the details of the methods employed to obtain informal economy estimates, the reader is referred to Breusch (2005), who provides a trenchant criticism of three applications of the MIMIC technique. Dell'Anno and Schneider (2006) respond.

[60] For example, the baseline estimates from Schneider (2005) utilize both currency-demand estimates as well as electricity demand-based estimates, such as Alexeev and Pyle (1972).

[61] The predicted estimates obtained from equation (29) are not valid according to the same argument made earlier regarding the use of predicted estimates given by (18) in subsection 4.2.

revolution is not because of inattention by creative empirical researchers. Rather, it is because severe measurement problems plague empirical analysis in this context, problems that arise not by chance, but because of the nature of the subject matter. We believe, though, that these problems do not imply that the standards of inference and identification should be lower. Instead, we suggest four steps that will incite this literature toward its own credibility revolution.

The first is a renewed call for creativity. Given the absence of direct observability of the phenomenon of interest, social scientists must be inventive in their search for traces of true income and noncompliance, and apply the appropriate econometric techniques to these variables. In natural science, the search for evidence of the invisible rewards creativity. Indeed, the 1960 Nobel Prize in Physics was awarded to Donald Glaser for his invention of the bubble chamber, which reveals the tracks of directly unobservable subatomic particles as trails of bubbles in a superheated liquid, generally liquid hydrogen. Not only can the path of the particles be observed, but bubble density around a track is proportional to a particle's energy loss, and the response of a track's path to a magnetic field can determine the particle's charge and momentum.

Second, we believe that empirical analysis should be guided by the theory of tax noncompliance. Theory provides caveats about identification problems due to endogeneity and left-out independent variables, and should guide the econometric analysis of the traces of true income and traces of noncompliance that creative researchers will still want to examine. Additionally, the use of lab experiments can facilitate testing of theoretical predictions that may otherwise be untestable due to the environment and institutions in place.

Third, a credibility revolution in this field will require more transparency about assumptions, methods, and what can and cannot be inferred with confidence from observed traces of true income and noncompliance, particularly with regards to the macro literature. Substantial advances in the credibility of the macro analysis of causal determinants could be obtained by abandoning the use of dependent variables which are a function of predicted values from previous regression analyses. Instead, traces of true income or noncompliance should be used directly. Researchers should acknowledge the difficulty of gathering and interpreting evidence of the invisible, but should not be dissuaded from investigating this crucial aspect of public economics. However, producing estimates that cannot be interpreted in a meaningful way does not facilitate the progressive accumulation of useful knowledge.

Finally, we encourage the use of randomized field experiments, as these promise to provide the most credible estimates of the effect of policy interventions addressed toward tax evasion and the informal economy. We would ultimately like to see the treatments for these experiments designed such that the randomized treatment provides direct evidence on one or more feasible policy changes and an examination of non-experimental methods using these field experiments as a benchmark.

Acknowledgements This paper was presented as a keynote address to the 66th Annual Congress of the International Institute of Public Finance held in Uppsala, Sweden on August 23-26, 2010. We thank Congress participants and, in particular, Friedrich Schneider for helpful comments. We also thank two anonymous referees.

References

Alexeev, M. and Pyle, W. (1972). A note on measuring the unofficial economy in the former Soviet Republics. *Economics of Transition*, 11(1):153–175.

Allingham, M. and Sandmo, A. (1972). Income tax evasion: A theoretical analysis. *Journal of Public Economics*, 1(3-4):323–338.

Alm, J., Bloomquist, K., and McKee, M. (2010). On the external validity of tax compliance experiments. In process.

Alm, J., Jackson, B., and McKee, M. (1992). Institutional uncertainty and taxpayer compliance. *The American Economic Review*, 82(4):1018–1026.

Andreoni, J., Erard, B., and Feinstein, J. (1998). Tax compliance. *Journal of Economic Literature*, 36(2):818–860.

Angrist, J. D. and Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives*, 4(2):3–30.

Beron, K. J., Tauchen, H. V., and Witte, A. D. (1992). The effect of audits and socioeconomic variables on compliance. In Slemrod, J., editor, *Why People Pay Taxes*, pages 67–89. Ann Arbor: University of Michigan Press.

Breusch, T. (2005). Estimating the underground economy using MIMIC models. http://129.3.20.41/eps/em/papers/0507/0507003.pdf.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3).

Charmes, J. (2006). Measurement of the contribution of informal sector / informal employment to GDP in developing countries: Some conceptual and methodological issues. Expert Group on Informal Sector Statistics (Delhi Group) Paper No. 1.

Christensen, K., Cline, R. J., and Neubig, T. J. (2001). Total corporate taxation: Hidden, above-the-line, non-income taxes. *State Tax Notes*, November 21:525–531.

Clotfelter, C. (1983). Tax evasion and tax rates: An analysis of individual returns. *Review of Economics and Statistics*, 65(3):363–373.

Cummings, R. G., Martinez-Vasquez, J., McKee, M., and Torgler, B. (2009). Tax morale affects tax compliance: Evidence from surveys and an artefactual field experiment. *Journal of Economic Behavior and Organization*, 70:447–457.

Dell'Anno, R. and Schneider, F. (2006). Estimating the underground economy by using MIMIC models: A response to T. Breusch's critique. Department of Economics, Johannes Kepler University of Linz Working Paper No. 0607.

DePaula, A. and Scheinkman, J. (2010). The informal sector: An equilibrium model and some evidence from Brazil. Working Paper 10-024. Penn Institute for Economic Research.

Feige, E. L. (1990). Defining and estimating underground and informal economies: The new institutional economics approach. *World Development*, 18(7):989–1002.

Feinstein, J. (1991). An econometric analysis of income tax evasion and its detection. *RAND Journal of Economics*, 22(1):14–35.

Feldman, N. and Slemrod, J. (2007). Estimating tax noncompliance with evidence from unaudited tax returns. *The Economic Journal*, 117(518):327–352.

Fellner, G., Sausgruber, R., and Traxler, C. (2009). Testing enforcement strategies in the field: Legal threat, moral appeal and social information. CESIFO Working Paper No. 2787.

Frey, B. and Weck-Hanneman, H. (1984). The hidden economy as an 'unobserved' variable. *European Economic Review*, 26(1-2):33–53.

Gërxhani, K. (2004). The informal sector in developed and less developed countries: A literature survey. *Public Choice*, 120:267–300.

Glazer, A. and Konrad, K. (1996). A signaling explanation for charity. *American Economic Review*, 86(4):1019–1028.

Gordon, R. and Li, W. (2009). Tax structures in developing countries: Many puzzles and a possible explanation. *Journal of Public Economics*, 93(7-8):855–866.

Henderson, J. V., Storeygard, A., and Weil, D. N. (2009). Measuring economic growth from outer space. NBER Working Paper 15199.

Hurst, E., Li, G., and Pugsley, B. (2010). Using expenditures to estimate missing self employed income. In process.

Hussmanns, R. (2004). Measuring the informal economy: From employment in the informal sector to informal employment. International Labour Office Policy Integration Department Working Paper No. 53.

Johnson, S., Kaufmann, D., and Shleifer, A. (1997). The unofficial economy in transition. *Brookings Papers on Economic Activity*, 1997(2):159–239.

Jütting, J. P. and de Laiglesia, J. R. (2009). Is informal normal? Towards more and better jobs in developing countries. An OECD Development Centre Perspective: *www.oecd.org/document/*32/0,3343,*en*_2649_33935_42024438_1_1_1_1,00.*html*.

Kaufmann, D. and Kaliberda, A. (1996). Integrating the unofficial economy into the dynamics of post-socialist economies: A framework of analysis and evidence. World Bank Policy Research Working Paper No. 1691.

Kleven, H. J., Knudsen, M. B., Kreiner, K. T., Pedersen, S., and Saez, E. (2010). Unwilling or unable to cheat? Evidence from a randomized tax audit experiment in Denmark. NBER Working Paper 15769.

Lackó, M. (2000). Hidden economy–an unknown quantity? Comparative analysis of hidden economies in transition countries, 1989-95. *Economics of Transition*, 8(1):117–149.

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *The American Economic Review*, 76(4):604–620.

Lemieux, T., Fortin, B., and Frchette, P. (1994). The effect of taxes on labor supply in the underground economy. *The American Economic Review*, 84(1):231–254.

Loayza, N. V. (1996). The economics of the informal sector: A simple model and some empirical evidence from Latin America. *Carnegie-Rochester Conference Series on Public Policy*, 45:129–162.

Nigrini, M. J. (1996). A taxpayer compliance application of Benford's Law. *The Journal of the American Taxation Association*, 18:72–91.

OECD (2006). Tax administration in OECD and selected non-OECD countries: Comparative information series,. Organisation for Economic Co-operation and Development: http://www.oecd.org/dataoecd/37/56/38093382.pdf.

OECD (2008). Tax administration in OECD and selected non-OECD countries: Comparative information series,. Organisation for Economic Co-operation and Development: http://www.oecd.org/dataoecd/57/23/42012907.pdf.

Pissarides, C. A. and Weber, G. (1989). An expenditure-based estimate of britain's black economy. *Journal of Public Economics*, 39(1):17–32.

Robinson, L. and Slemrod, J. (2010). Understanding multidimensional tax systems. Tuck School of Business Working Paper 2010-81.

Schneider, F. (2005). Shadow economies around the world: What do we really know? *European Journal of Political Economy*, 21:598–642.

Schneider, F., Buehn, A., and Montenegro, C. E. (2010). Shadow economies all over the world: New estimates for 162 countries from 1999 to 2007. World Bank Policy Research Working Paper No. 5356.

Shaw, J., Slemrod, J., and Whiting, J. (2010). Administration and compliance. In Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., and Poterba, J., editors, *Dimensions of Tax Design: The Mirrlees Review*, pages 1100–1162. Institute of Fiscal Studies and Oxford University Press.

Slemrod, J. (1985). An empirical test for tax evasion. *Review of Economics and Statistics*, 67(2):232–238.

Slemrod, J. (2001). A general model of the behavioral response to taxation. *International Tax and Public Finance*, 8(2):119–128.

Slemrod, J. (2004). The economics of corporate tax selfishness. *National Tax Journal*, 57(4):877–899.

Slemrod, J. (2007). Cheating ourselves: The economics of tax evasion. *Journal of Economic Perspectives*, 21(1):25–48.

Slemrod, J., Blumenthal, M., and Christian, C. (2001). Taxpayer response to an increased probability of audit: evidence from a controlled experiment in Minnesota. *Journal of Public Economics*, 79(3):455–483.

Stanley, T. J. (1995). Radar detectors, fixed and variable costs of crime. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=219.

Tanzi, V. (1980). The underground economy in the United States: Estimates and implications. *Banco Nazionale del Lavoro Quarterly Review*, 135(2):427–453.

Tanzi, V. (1983). The underground economy in the United States: Annual estimates. *IMF Staff Papers*, 30(2):283–305.

Thomas, J. (1999). Quantifying the black economy: 'Measurement without theory' yet again? *The Economic Journal*, 109(456):F381–F389.

Torgler, B. (2007). *Tax compliance and tax morale: A theoretical and empirical analysis*. Edward Elgar Publishing.

Yitzhaki, S. (1974). A note on income tax evasion: A theoretical analysis. *Journal of Public Economics*, 3(2):201–202.

Zellner, A. (1970). Estimation of regression relationships containing unobservable variables. *International Economic Review*, 11(3):441–454.