1. Introduction

During the late 1970s researchers, reporters and legislators discovered a large and growing underground economy in the United States. Reactions to this phenomenon were diverse with legislators passing a series of tax laws (in 1982, 1984 and 1986) that sought to stem the rising tide of noncompliance and reporters informing the public about the prevalence of tax cheating in this country. Researchers were somewhat slower to react and it is only now that we see substantial research interest in tax administration and tax compliance. With the growing interest and the increasing availability of data for research on tax compliance, it seems wise to step back and assess the state of this research and suggest possible directions for future research. In making this assessment and suggestions, we will rely heavily on our recent experience in conducting compliance research and serving as chair of the National Academy of Sciences' Panel on Research on Taxpayer Compliance.

To be more specific, in the next section we discuss theoretical work on tax compliance. In the section that follows, we consider sources of data for tax compliance research in the United States. In section 4, we briefly review empirical work and in section 5 we outline a research program. The final section contains our conclusions.

2. Theoretical Models of Tax Compliance

Most theoretical work on tax compliance has been done by economists. Other researchers have suggested interesting conceptualizations, but have not, as far as we are aware, offered fully developed theories. Accordingly, we limit this brief survey to economic models of tax compliance.

As far as we are aware, the first economic model of tax compliance was developed by Allingham and Sandmo (1972) who extended the expected utility model of criminal activity originated by Becker (1968) to the tax arena. In the Allingham and Sandmo model, the taxpayer decides upon the amount of taxes to report to the taxing agency. When making this decision the taxpayer seeks to maximize expected utility which is defined to be the sum of the utility value of each outcome weighted by the probability that the particular outcome occurs. The taxpayer faces an extremely simple world in which there is a single, proportional tax rate, taxable income that is known perfectly and costlessly, no tax preparers or advisors, a probability of audit that is unaffected by the taxpayers' reporting behavior, a single penalty based on the amount of income that is underreported, and two possible states of the world (completely successful or completely unsuccessful tax evasion). With this model, Allingham and Sandmo show that higher probabilities of audit deter underreporting and that a higher rate for the proportional tax leads to lower levels of reported income.

Theoretical work during the 1970s sought to extend the Allingham and Sandmo model with major interest being in making income endogenous to the model (e.g., Weiss (1976), Sandmo (1981), Cowell (1985)). Specifically, the extensions of the basic model assume that both leisure time and after tax income affect well being and that the labor/leisure choice and the reporting decision are made jointly. With these assumptions, one can not show that increasing the probability of audit deters underreporting or that increasing the tax rate stimulates it without placing rather unappealing restrictions on the nature of the utility function.

Most recent theoretical work has focused on the interaction between the taxpayer and the taxing agency. This work often marries the expected utility paradigm with game theoretic approaches. The work concentrates on determining the optimal audit policy. Even in very simple single period models of this type, with risk neutral taxpayers and lump sum taxes, it has not been possible to discern the optimal audit strategy. However, it has been shown that random audits (the type of audits assumed in early models) are not necessarily optimal. Random audit strategies may be dominated by cut-off audit strategies under which the taxing agency audits all people who report an income below a certain level (Reiganum and Wilde, 1985).

Researchers have begun to develop multiperiod game theoretic models (e.g., Greenberg (1984), Landabeger and Meilijson (1982)). Landabeger and Meilijson demonstrate that an audit strategy based on the taxpayer's past
compliance behavior can achieve higher levels of compliance—at the same cost—than a random audit strategy. Greenberg has shown that even when audit resources are limited, the extent of evasion can be made arbitrarily small by the appropriate design of an audit strategy in which the results of previous audits are used in selecting returns for audit.

Deductive modeling of the type described above has alerted compliance researchers to a number of important issues—the importance of providing proper “incentives” for compliance, the importance of thinking carefully about the way in which audit policy is modeled, the centrality of the interaction between the taxing agency and the taxpayer. However, this modeling has, to date, been relatively abstract and has not considered the institutional structure (e.g., the tax code and actual policies pursued when administering the code) surrounding tax compliance and tax administration.

In our recent work (Tauchen and Witte, 1986, 1987), we have sought to incorporate some important aspects of the institutional structure while developing a model to guide empirical work on data from the Internal Revenue Services' (IRS) Tax Compliance Measurement Program (TCMP). In our model we see the taxing agency as using a rule to select returns for audit. The rule depends upon characteristics of the taxpayer's filed return and a limited set of socioeconomic variables that are both observable by the IRS and that may legally be used to select returns for audit (e.g., source of income).

The IRS selects the majority (well over half, Wilt, 1986) of returns for audit by using information on current returns from the IRS's Master File (this file contains selected information from filed tax returns) and parameters derived from a previous year's TCMP data. With this type of audit rule, the probability that a taxpayer is audited is not exogenous (as in the random audit models of the 1970s). Further, it is the nature of the rule itself and not simply the probability of audit that affects taxpayer behavior. To be more specific, the parameters of the audit rule not the probability of audit determine the level of income and deductions that a taxpayer reports.

3. Data

In this section, we discuss three types of data that might be used to study compliance behavior: (1) survey data (2) tax return data and (3) audit data. Each type of data has been used for empirical research and all have advantages and disadvantages.

There have been many surveys, including national surveys sponsored by the IRS, that have attempted to estimate taxpayer compliance. Kinsey (1984a) and Scholz, Roth and Witte (1987) provide comprehensive bibliographies of this work. As do all survey data, these data suffer from both response and nonresponse bias. However, we believe that survey data related to compliance research has a number of unusual features that require additional consideration. First, these data relate to an illegal act and thus are sensitive and subject to far greater concerns related to privacy and response bias than are data on non-sensitive topics. As far as we are aware, there has been only one published attempt (Aitken and Bonneville, 1980) to obtain tax compliance information using survey techniques (e.g., random response) that are specifically designed to elicit information on sensitive subjects. Second, compliance questions relate to income and other detailed information on income and personal finances. It is well known that people are reluctant to report such information. Thus, nonresponse may be a more serious problem in compliance surveys than in many other types of surveys. Finally, accurate response requires a substantial recall of the detailed information recorded on the tax return. As far as we are aware, existing survey research has not come to grips with the difficulties involved in obtaining such information. These difficulties are aggravated by the fact that almost half of all returns are joint (the return combines information for a husband and wife) and that nearly half of all returns are prepared by someone other than the taxpayer.

While there are substantial difficulties collecting meaningful compliance information through surveys, this type of data does have some unique advantages. First, the data that has been collected from past survey instruments have reflected only intentional noncompliance with the tax codes. Most other type of data (e.g., TCMP data) combine noncompliance due to intent with errors due to ignorance. Such composite measures of underreporting are difficult to analyze because the factors associated with ignorance of the tax laws may be quite different from those associated with intentional evasion. Second, it is possible in surveys to collect a wide range of sociodemographic and attitudinal variables. Such information is not generally available
on tax return or TCMP data bases.

Tax return data are available on IRS's Master File and a random sample of individual return information is made available to researchers through the IRS's Statistics of Income (SOI) program. Recently, Arthur Young has established a center at the School of Business Administration at the University of Michigan to make these data more readily accessible and to construct a panel of linked returns. These data contain only taxpayer reports and contain no estimate of compliance. However, these data could be quite useful for experimental and quasi-experimental research. For example, if the IRS or external researchers were to carry out an innovative program, say a program designed to "sensitize" taxpayers about tax issues, in one state, the effect of the program could, using SOI data, be estimated by looking at the way in which changes in reporting in that state differ from changes in states that received no such sensitizing.

Perhaps, the premier data for compliance research is data collected by the IRS through its Tax Compliance Measurement Program (TCMP). Under this program, the IRS randomly selects approximately 50,000 individual returns and subjects them to full audits by the IRS's most experienced auditors. For every line item on the return, the auditors record both the taxpayer's report and the amount they determine should have been reported. The difference between the taxpayer's report and the auditor's findings provides a natural measure of the extent to which "taxpayers" are complying with the tax laws. Indeed, these data have been used by the IRS to estimate the total amount of noncompliance in the US (U.S. Department of the Treasury, 1983) and by researchers to estimate models of tax compliance (Clotfelter, 1983). However, there are a number of difficulties involved in using these data in this way. First, auditors do not uncover all underreporting. In general, auditors are more successful in uncovering overstated adjustments, deductions and exemptions than they are in uncovering unreported income. The reason is simple, for adjustments, deductions and exemptions the burden of proof rests with the taxpayer while for unreported income it is up to the auditor to make the discovery.

On the basis of a 1976 study, the IRS estimated that auditors uncover only $1 of every $3.50 that is not recorded on the tax return (U.S. Department of the Treasury, 1983). Beginning in 1979, the IRS began supplying TCMP auditors with all information documents (e.g., all 1099s) related to the returns they were auditing. This should mean that recent TCMP data uncover a far greater proportion of unreported income than was possible prior to 1979 when these documents were not available to the auditors.

4. Empirical Work

Survey research (see Kinsey (1984a) for a review) has generally focused on the effect of attitudes, beliefs and socio-demographic factors on compliance. This research finds, in general, that individuals who place greater trust in government, believe that the tax system is fair and equitable, and believe that others generally comply with tax laws are more compliant. Women and older individuals are found to be more compliant than others.

There has, to date, been relatively little research that has used revenue department data to study tax compliance. In a very early study, Schwartz and Orleans (1967) test the relative effectiveness of moral appeals and reminders of penalties in obtaining compliance using an experimental design and aggregate information on reporting behavior provided by the IRS. Their results indicate that both moral appeals and reminders served to increase reported adjusted gross income (AGI). Appeals to conscience raised reported AGI significantly and the effect of appeals to conscience on reported AGI is larger than the effect of sanctions and threats, although not significantly so.

Clotfelter (1983) uses TCMP data to estimate a reduced form model of compliance. His dependent variable is the logarithm of underreported income (AGI and taxable income) as indicated by the difference between auditors' findings and taxpayers' reports. Clotfelter assumes that the audit rules are the same for all taxpayers and excludes the probability of audit from his model. He finds after tax income, the marginal tax rate, the proportion of wages in AGI and region of the country to be significantly related to underreporting of AGI.

Witte and Woodbury (1985) use a data set compiled by IRS which includes information derived from the TCMP, detailed IRS records of administrative activities and socio-demographic variable from the Census. The data set contains only a current measure of tax compliance constructed by the IRS. Using this measure as their dependent variable Witte and Woodbury find that both audits and computerized notices
serve to increase compliance. They also find that opportunities for noncompliance, attitudes, a broad range of socio-demographic and some community characteristics are significantly related to compliance.

5. A Proposed Research Agenda

Our own assessment of existing research and data leads us to believe that four distinct types of research could be very valuable at this juncture: (1) research that develops theoretical models that are grounded firmly in the legal and administrative realities that surround tax compliance and tax administration; (2) a set of experiments and quasi-experiments designed to improve our understanding of tax compliance behavior and improve administration of the tax laws; (3) careful work that uses the audit data generated by the IRS's Tax Compliance Measurement Program in conjunction with other data to estimate models of compliance behavior; (4) work that assesses the ability of various approaches for obtaining self-report measures of compliance. We discuss each of these areas of research in turn.

For reasons of mathematical tractability, existing theory tends to be relatively narrowly focused. Most theoretical research abstracts from important taxpayer practices (e.g., the use of preparers, the prevalence of joint returns, the fact that the taxpayer chooses a "portfolio" of line items related to the level of taxable income to report). This is a promising area for research. Theoretical work that extends the recent dynamic and game theory models of compliance may help explain the joint determination of taxpayer and tax administrator behavior. In addition, theoretical work that better incorporates the nature of tax law and administrative procedures and of a taxpayer's decision problem would be useful for guiding empirical work. At a minimum, these models could provide some direction for the selection of explanatory variables and the specification of the error structure.

As the theoretical work on tax compliance proceeds, it would also be worthwhile to test the hypothesis of the more abstract models. Laboratory experiments seem particularly promising in this regard. Such experiments have recently been used to test abstract economic theories related to efficient markets, Coase's Theorem and the expected utility paradigm.

Field experiments have been used for various types of research related to tax administration and tax compliance. See Boruch (1986) for an excellent discussion of experiments and quasi-experiments in the tax area. We have already discussed the study by Schwartz and Orleans (1967) which used a field experiment to discern the relative effectiveness of different types of taxpayer appeals on compliance. More recently the IRS has employed contractors who have used experimental methods to assess the relative merits of alternative methods of eliciting compliance information in surveys (Aitken and Bonneville, 1980) and has itself conducted a field experiment to assess the effectiveness of various methods of handling accounts receivable (Pengo, 1985). We believe that there are a number of interesting areas (e.g., the examination of returns, the development of tax forms and instructions, taxpayer services, taxpayer appeals) in which experiments could provide valuable policy and research insights. We will discuss only one such area—the effect of the examination of returns and the resulting contact with taxpayers on compliance. Schwartz and White (1987) contains a discussion of a number of other promising experiments and quasi-experiments in the tax area.

For many years, the IRS has stated repeatedly that examinations of tax returns are its major instrument for obtaining compliance. Yet the studies to date, whether conducted by external researchers or by the IRS, have produced decidedly mixed and controversial results regarding the effect of examinations on compliance. Further, budget stringency, the development of more efficient and effective computer checking and matching routines, and the demands of other tasks (e.g., campaigns against drug dealers, collection of child support obligations) have meant that audit coverage has declined (from 2.9 percent in 1969 to 2.0 percent in 1980 and 1.3 percent in 1985) while the number of computerized notices sent to taxpayers has increased. We do not know the relative effectiveness of various methods of examining returns—in-person audits, correspondence audits and computerized notices. A carefully designed program of field experiments could help us answer this very important question. An even less expensive and sensitive set of experiments could discern the relative effectiveness of various methods for communicating with taxpayers in the course of examining their returns.

Researchers seeking to carry out experiments in the tax compliance area must be sensitive to a number of issues
which do not arise in other areas of research. First, the IRS is legally prohibited from releasing any return information that could be tied to an individual. Thus, researchers can only expect to receive data as did Schwartz and Orleans in experimental research on designated groups in a form that prevents any individual being identified. Second, the IRS is legally required to "treat all taxpayers equally". Court cases have upheld the IRS's right to conduct the TCMP audits. However, the equal treatment issue may well arise when experiments are conducted in such sensitive areas as taxpayer audits and punishment. Finally, the IRS is an arm of the federal government and as such is sensitive to political sentiment.

Members of Congress do not like the IRS to "harass" their constituents and experiments that are geographically concentrated may well be politically infeasible.

Field experiments can provide us with robust estimates of the effect of particular interventions on compliance. However, such research can not, at least in the short run, provide us with a broad understanding of the relative effect of various activities and characteristics on compliance. Such insights will, we believe, come more quickly from estimating multivariate models of compliance using the TCMP and other data bases.

The IRS has recently developed a number of methods for giving researchers access to this important but highly sensitive (the data are used to develop the IRS's audit formulas) data. For example, the agency has provided researchers with aggregate TCMP results at the national and the IRS district level, and computer output of work that researchers request on the individual TCMP files. Further, the agency has encouraged researchers to spend time at the IRS under the Intergovernmental Personnel Act (IPA) and other arrangements.

In availing themselves of the new opportunities for research on the TCMP data, researchers should carefully consider the way in which these data can be used for compliance research. Our own consideration of these data have led us to the following conclusions. First, the dependent variable used in empirical models of taxpayer compliance behavior should be information reflecting the taxpayers' reports not the auditors' estimates of the amount of income underreported (e.g., Clotfelter (1983)) or a surrogate estimate of compliance (e.g., Witte and Woodbury (1985)). The advantages relate mainly to the interpretation of the coefficients of the empirical model.

With the taxpayer report as the dependent variable these coefficients clearly relate to the taxpayers' reporting behavior. However, with the auditors' estimate of the amount of income underreported as the dependent variable the coefficients of the empirical model reflect the auditors' as well as the taxpayers' behavior. Since, as noted earlier, auditors are estimated to uncover only $1 of every $3.50 that is not reported (U.S. Department of the Treasury, 1983), the effect of auditors' behavior can be quite important. For research work, in which the taxpayers' compliance behavior is of interest, the compliance measures based on auditors' estimates are clearly of limited use. (See Ben, Tauchen and Witte, 1987, for a more extended discussion of the advantages of using the items reported by taxpayers as the dependent variables in empirical work based on TCMP data).

Based on previous work with the TCMP data, we have also concluded that additional data sources will have to be used in conjunction with the TCMP data. The TCMP data contain only information available on the tax return and the auditors' findings relating to the accuracy of the taxpayers' reports. To estimate meaningful models of compliance behavior, additional information must be added to the TCMP data bases. Detailed knowledge of the IRS's administrative activities will be required if the effect of such activities on compliance is to be estimated. Information on the socioeconomic characteristics and attitudes of taxpayers will also be required. Linking these various types of data to the TCMP files will be both difficult and sensitive. As far as we are aware, there is only one data base, the 1969 three digit zipcode data base constructed by the IRS and used by Witte and Woodbury (1985), that attempts to link a broad range of compliance, sociodemographic and administrative information. Under an NSF grant, Tauchen, Witte and Beron are currently trying to link TCMP, SOI, Census and administrative information at the district level for 1979. Linkage with TCMP data at the individual level will prove much more sensitive than linkage of aggregated data. The recent report of the NAS Panel on Research on Taxpayer Compliance (Scholz, Roth, and Witte, 1987) suggests a number of possible institutional structures that might allow TCMP data to be linked to other sources of information at the individual level. However, we feel that it will be some time before such a
linked individual data base will be available to researchers.

This brings us to the topic of survey research and the role that such research can play in studying tax compliance. As it is currently structured, we believe that survey research is useful mainly for providing estimates of the prevalence of intentional noncompliance with tax laws. Previous survey research has examined issues related to the intent to evade taxes. As such, the data from these studies provide somewhat limited information about the extent of noncompliance with tax laws since noncompliance can result from both intent and ignorance. Further, doubts about the validity of current survey measures of compliance are raised by a study of 155 Dutch taxpayers (Hessing, et al., 1985). This study finds only a negligible correlation (i.e., r=.06) between self reports and administrative records of noncompliance. Clearly more research is called for to improve survey measures. While there has been some research on methods designed to elicit information on tax compliance (see Aitken and Bonneville, 1980), there has been, as far as we are aware, no research related to the timing of interviews (interview dates close to filing time would seem more likely to elicit accurate information), selection of interviewee, (the individual who accumulated and processed the tax information might be able to provide more accurate information), or review of tax documents (the prospective interviewee might be able to provide more accurate information by reviewing tax documents before the interview).

6. Conclusions

We believe that tax reporting behavior is one of the most interesting and fruitful areas for compliance research. Consider the situation. Most of the adult population is "at risk" of failing to comply with tax laws annually. Further, the tax paying population is asked to provide a detailed report of relevant activities to the IRS each year. In addition, employers and others are asked to provide detailed information regarding their payments to taxpayers. Virtually all of these records are computerized and can provide a valuable source of research data. If this was not enough to stimulate research, we have the IRS's TCMP program which provides detailed findings regarding the compliance of a random sample of approximately 50,000 returns every three years. With an issue as important as tax compliance and data sets as promising as those available at the IRS and other data centers, one might well ask why there has been relatively little research on tax compliance to date. The answer is twofold. First, the IRS was not until quite recently receptive to researchers' proposals. Second, researchers were reluctant to delve into the mass of details involved in understanding tax return data (once a year is bad enough) and the way in which the tax system is administered. We believe that all of this is changing and are hopeful that a research agenda that combines careful theoretical and empirical research will enhance our understanding not only of tax compliance but also of broader compliance issues. For example, studies of tax compliance may well enhance our understanding of compliance with requests, rules and laws in general. Further such research may provide insights regarding the way in which individuals cope with complex tasks since for many the annual struggle with a tax return is one of the most complex tasks undertaken.

NOTES

1. The report of the Panel is contained in Scholz, Roth and Witte (1987).
3. There have been some exceptions in this regard. For example, see Yitzhaki (1974) and Witte and Woodbury (1982).
4. Scotchmer (1986) has recently written an interesting paper that models the taxpayer's decision to use a tax advisor.
5. See Hoffman and Spitzer (1985) for a survey of this literature. Lattimore and Witte (1977) have recently used computerized laboratory experiments to test various models of decision making under uncertainty.
References


