1-1-1992

Analyzing Income Tax Evasion Using Amnesty Data with Self-Selection Correction: The Case of the Michigan Tax Amnesty Program

Steven E. Crane  
*Marquette University, steven.crane@marquette.edu*

Farrokh Nourzad  
*Marquette University*

*Publisher link*. © 1992 University of Michigan Press. Used with permission.
Analyzing Income Tax Evasion Using Amnesty Data with Self-Selection Correction: The Case of the Michigan Tax Amnesty Program

*Steven E. Crane and Farrokh Nourzad*

The issue of income tax evasion has been subjected to a great deal of scrutiny by economists. Since the pioneering work by Allingham and Sandmo (1972), the standard approach to analyzing the individual’s evasion decision has been to employ a portfolio-choice framework in which the optimal level of evasion is obtained from maximizing the expected utility of income after taxes and penalties.

Using this approach, four factors have been commonly found to affect the decision to evade. These are the individual’s true income, the tax rate, the probability that the evader will be detected, and the penalty rate to which detected evaders are subjected. In most cases, a positive relationship between the level of evasion and the individual’s true income, and negative relationships with both of the compliance policy tools are obtained. With respect to the tax rate, however, most models have been unable to determine an unambiguous relationship due to potentially opposing income and substitution effects.\(^1\)

---

1. A tax rate increase produces a substitution effect that leads to greater evasion because it increases the marginal return to successful evasion. On the other hand, by reducing disposable income, a higher tax rate generates an income effect that may lead to more or less evasion depending on the individual’s attitude toward risk. As long as risk aversion is a decreasing function of income, the income effect is negative. Therefore, unless risk aversion increases with income, or the substitution effect is strong enough to dominate the income effect, one obtains the counterintuitive result that higher tax rates lead to reduced evasion, or that the effect is indeterminate. A qualification to this was pointed out by Yitzhaki (1974), who showed that if taxes are proportional and fines are levied on evaded taxes rather than evaded income, there would be no substitution effect.
Considerable additional work has been done within this basic analytical framework. Perhaps most important have been attempts to endogenize various critical parameters, such as the penalty rate and the probability of detection (e.g., Koskela 1983). The model has also been made more complicated by introducing an endogenous labor supply decision (e.g., Pencavel 1979; Sandmo 1981) and by adding richness to various aspects of the tax system, such as its degree of progressivity (e.g., Marchon 1979; Pencavel 1979). The consequences of these extensions typically have been that few, if any, unambiguous predictions can be extracted from the model.

Recently, a number of authors have approached the tax evasion question from a different perspective, using the tools of game theory (e.g., Graetz, Reinganum, and Wilde 1986; Reinganum and Wilde 1986 and 1988). In principle, this approach holds much promise, because it is perhaps the most effective way to truly endogenize the compliance policy variables in such a way as to adequately address the strategic aspects of the interaction between the taxpayer and the tax authority. However, it is a bit too soon to tell how successful this emerging literature will be in generating usable general conclusions regarding the evasion decision.²

Empirically, a fairly extensive literature has emerged. Most of this empirical work has its roots in the traditional portfolio-choice approach to modeling the decision to evade. What distinguishes these studies is how they cope with the obvious measurement difficulties associated with analyzing an illegal activity such as tax evasion. Indeed, a variety of innovative approaches to measuring evasion have been employed.

Some researchers (e.g., Friedland, Maital, and Rutenberg 1978; Geeroms and Wilmots 1985; Spicer and Becker 1980) have designed experiments or conducted surveys in order to generate relevant data. Others (e.g., Crane and Nourzad 1986; Tanzi 1983) have approached the problem from a macroeconomic perspective. An attempt has even been made to develop an evasion index from the distribution of tax returns within tax brackets (Slémrod 1985). Only a few authors (e.g., Clotfelter 1983; Dubin and Wilde 1988; Klepper and Nagin 1989; Witte and Woodbury 1985) have been able to develop direct measures that are representative of evasion behavior under actual tax systems. Of these, only Clotfelter has been able to examine the issue at the individual level.

This essay presents an empirical investigation that is based on the tradi-

². One interesting finding thus far is that a tax rate change generates an additional effect through its impact on the marginal return to auditing. It has been shown that, under some simplifying assumptions and for certain audit classes, this effect dominates the conventional tax rate effect leading to a negative overall impact (see Graetz, Reinganum, and Wilde 1986). This result holds independently of the taxpayer's attitude toward risk.
Using Amnesty Data with Self-Selection Correction

169

tional theoretical framework but uses a new approach to measuring and analyzing income tax evasion. It explains how data from state income tax amnesty programs can be used to construct direct measures of income tax evasion at the individual level. We then demonstrate how these measures can be examined empirically, despite complications due to both the self-selected nature of the data and the possible interaction between the decisions to evade at the federal and state levels.

The rest of our essay is organized as follows. The next section makes the case for using amnesty data to analyze evasion and discusses the methodological procedures needed to address the complications that arise due to self-selection bias and the presence of multiple taxing jurisdictions. The third section provides a brief overview of the income tax structure and the tax amnesty program in the state of Michigan, the source of the data used in this study. In the fourth section we present our empirical model of Michigan state income tax evasion and our estimation results. This is followed by some concluding comments in the fifth section.

Using Amnesty Data to Study Evasion

Amnesty Data as Evidence of Evasion

State income tax amnesty programs represent what we believe is a new source of microdata that allows the construction of direct measures of tax evasion. Participants in these programs are effectively admitting that they cheated and are indicating the magnitude of their noncompliance. Thus, the variable most difficult to quantify in an empirical evasion study, the level of noncompliance, is readily available. This is especially true if one focuses on amnesty filers who amended a return that was previously filed, rather than individuals who filed for the first time under amnesty. Assuming information from these amended filers' original returns is available, construction of an evasion measure is straightforward. One simply assumes that the amended returns represent the "truth" and compares the figures on these returns with their counterparts on the original returns.

In many cases, amnesty returns provide considerable information useful

3. It is not clear what to do with the nonfilers, either theoretically or empirically. This is explained in n. 15, relating to our discussion of the data used here.

4. This seems to be a plausible assumption because it is unlikely that one who has voluntarily admitted to evading on a particular tax return would file a false amended return. This is especially true if it is announced that the amnesty returns may be audited as well. However, it may be argued that an individual might admit only a portion of his or her evasion in the hope of reducing the chances of additional auditing, which could find the remaining evasion.
for constructing independent variables as well. Given the assumption that the taxpayer is truthful when filing under amnesty, the information on the amended return will reveal such important taxpayer attributes as true income, and perhaps a usable marginal tax rate or certain demographic characteristics.

What will most likely be lacking, however, are compliance policy variables. Because amnesty data are primarily cross-sectional and penalty provisions are usually uniform across individuals, no measure of the monetary cost of evasion is likely to be available. By the same token, reliable measures of either a subjective probability of detection or an objective audit rate proxy are not likely to be readily available at the state level. Thus, at best, some indirect control for detection probability is all that is likely to be possible.

The Issue of Selectivity Bias

A major drawback of using amnesty data for the purpose of analyzing evasion behavior is the probable self-selection bias due to the fact that amnesty filers themselves determined whether or not they would participate in the program. Clearly, those evaders who voluntarily chose to participate in a particular state’s amnesty program may not be representative of the population of income tax evaders as a whole in that state, much less evaders in general. Fortunately, while complicated, it is possible to deal with this type of self-selection bias econometrically. This involves using a maximum likelihood technique that incorporates not only the variables influencing the evasion decision, but also those influencing the subsequent decision to participate in the amnesty program.

The procedure can be described as follows. We wish to estimate an evasion function such as

\[ y_i = X_i \beta + u_i, \quad i = 1, 2, \ldots, n, \]  

where \( y_i \) is a measure of evasion, \( X_i \) is a vector of the determinants of evasion, \( \beta \) is a vector of unknown parameters, and \( u_i \) is a random error term with mean zero and variance \( \sigma^2 \). But because evaders who are also amnesty participants are a self-selected group, estimating equation 1 using ordinary least squares (OLS) would result in biased estimates. This is because the distribution from which the sample comes is truncated in that it does not include nonparticipating evaders.

5. Usually many years are open for filing. However, given bad memory and the expiration of the statute of limitations, one suspects most returns are for the most recent years.

6. For a survey of the literature on the selectivity bias that arises from different types of self-selected samples along with remedial procedures, see Maddala 1983, chap. 9. Also see
Correcting the selectivity bias in amnesty data requires one to incorporate knowledge of factors that influenced the decision of amnesty-taking evaders to participate in the amnesty program. If such factors can be identified, one can obtain unbiased maximum likelihood (ML) estimates of the parameters of equation 1 using the following likelihood function (Maddala 1983, 266-67).

\[
\prod_i \left[ \frac{\Phi\left(\frac{Z_i \delta - (\zeta/\sigma)(y_i - X_i \beta)}{(1 - \zeta^2)^{1/2}}\right)}{\Phi(Z_i \delta)} \right] \\
(2 \pi \sigma^2)^{-1/2} \exp \left[ -\frac{1}{2 \sigma^2} (y_i - X_i \beta)^2 \right]
\]

(2)

where \(Z_i\) is a vector of factors influencing the participation decision, \(\delta\) is a vector of unknown parameters, \(\Phi(\cdot)\) is the distribution function of the standard normal, \(\zeta\) is the correlation coefficient between \(u_i\) and the error term of the participation function, and all other notations are as defined previously.\(^7\)

The term in the large bracket is the ratio of the conditional probability of participation in the amnesty program, given \((y_i - X_i \beta)\), to the unconditional probability of participation. The term outside of this bracket is the density function of \((y_i - X_i \beta)\). Thus, the bias correction procedure involves scaling the density function of \((y_i - X_i \beta)\) using the ratio of the two probabilities as weights. This procedure yields unbiased estimates for the parameters of the evasion model (the \(\beta\)'s), even though the estimates of the parameters of the participation function (the \(\delta\)'s) are unreliable. However, given that we are interested in the former set of estimates, the unreliability of the estimates of \(\delta\) is no cause for concern.

The Multiple Jurisdiction Issue

An additional complication with the use of state amnesty data for analyzing evasion arises from an important institutional aspect of the overall income tax system in the United States. The income of most U.S. taxpayers is subject to

---

7. In choosing participation variables, one need not be concerned with possible correlation with the evasion variables. In fact, some variables may appear in both the evasion and the participation functions. All that is necessary for identification purposes is that the participation function contains more than just a constant term (see Bloom and Killingsworth 1985). With this procedure, the critical factor is the correlation between the error terms of the evasion and the participation functions.
both federal and state income taxes. In this situation, there is likely to be overlap among the income tax returns that must be filed. Thus, we would expect that the decision regarding what to report on a state income tax return is likely to be affected by what was reported on a federal income tax return, and vice versa.

This interconnection may be due to several factors, some of which can be traced to the structure of the state return. Many state income taxes are piggybacked on the federal tax system. A typical way of doing this is to use Federal Adjusted Gross Income (AGI), with selected modifications, as the income subject to state taxation. Some states taking this approach often require that copies of the federal return be submitted with the state return for verification purposes. In this context, an evader would most likely carry his or her understatement of federal AGI forward to the state income tax return. This can result in state income tax evasion that is strictly caused by the federal evasion decision.

The interconnection between federal and state income tax returns can be more subtle, however. In truth, it may be that the two returns are completed simultaneously as the individual weighs potential costs or benefits associated with compliance at the state level with the corresponding costs or benefits at the federal level. If there is a perceived interdependency between either the costs or benefits of the two systems, the amount of evasion on one tax return may be affected by and affect what is reported on the other return. Such a perceived interdependency could arise from the well-known fact that state and federal tax enforcement agencies have a variety of information-sharing arrangements.

These possibilities suggest that an analysis of the behavior reflected on state tax returns should not be undertaken in isolation from the behavior reflected on the federal return. Exactly how to deal with this, of course, depends on the details of the tax code of the particular state whose amnesty data are employed.

**Michigan Income Tax System and Amnesty Program**

**The Michigan Income Tax**

Michigan's income tax system is fairly straightforward and closely linked to the federal system. In fact, the Michigan system is piggybacked on the federal

8. The taxpayer may also be subject to local income taxes as well. However, these taxes are generally less relevant in the present context because they are usually simply payroll taxes where the full amount is withheld. As a result, there is not much to the income reporting issue, and there may not even be a tax return to file.

9. The interdependency is also relevant to the analysis of federal tax evasion, although the effect may be weaker if the direction of causality is primarily from the federal return to the state return.
return such that Michigan residents begin completing their state return by reporting the AGI figure found on their federal returns. Several adjustments to federal AGI are then made to allow for both income that is taxable at the state but not at the federal level and for income that is not taxable in the state of Michigan but is taxable at the federal level.\textsuperscript{10} Making these adjustments results in the income subject to tax, from which personal exemptions ($1,500 each), but not itemized deductions, are subtracted to obtain state taxable income.

Michigan Taxable Income was subject to a flat tax rate that varied from 4.60 percent to 6.35 percent during the different tax years covered by this study (1980–85). Two different sets of credits were then deducted from the resulting gross tax figure.\textsuperscript{11} Once these credits were removed, any tax payments through withholding or quarterly estimated payments were deducted to determine the balance due or amount of refund.

Given this tax structure, evasion could occur on a return in a number of ways. First, the level of AGI may have been misstated. However, given that this figure was to be taken from the federal return, it seems likely that any evasion here was driven by decisions made when filling out the federal return.\textsuperscript{12} Second, the additions to or subtractions from AGI could be misstated. As with AGI, most of this information appears on the federal tax return and its associated schedules. However, since copies of the supplemental federal schedules were not required by the state authorities, it is possible that some independent, state-driven evasion may have occurred at this point.

The other primary possibilities for evading on Michigan income tax returns include overstating the number of exemptions or either type of credit. Of course, one could also have overstated withholding or estimated payment levels, but, given the clear paper trail associated with these items, this seems less likely. Thus, the type of evasion at the state level that seems most likely to

\textsuperscript{10} The primary adjustments that are added to AGI are interest income received from other states, certain capital gains and other gains, and losses attributable to activity in other states. The primary subtractions from AGI are for interest income from the federal government, military pay or benefits, retirement or pension benefits, certain capital gains, and income attributable to another state.

\textsuperscript{11} The first set of credits were for income taxes paid to Michigan cities or other states and contributions to Michigan institutions such as universities and libraries. The second set of credits were special allowances for such items as property tax payments, home heating expenditures, and investments in solar energy equipment.

\textsuperscript{12} The original decision to evade on the federal return could be influenced by opportunities on the state return. We do not consider this possible feedback. Of course, just because this type of evasion is federally driven does not mean that it is not state evasion. In fact, it may be that some state evasion occurs because the individual wants his or her state return to be consistent with his or her federal return. This does suggest, however, that one must look elsewhere for explanations of this form of state evasion behavior.
have occurred is a misstatement of adjustments to AGI, an overstatement of credits from the gross tax bill, or some combination of the two. What makes this activity potentially observable is the existence of a tax amnesty program.

The Michigan Tax Amnesty Program

The Michigan Tax Amnesty Program was fairly conventional. Since it is described in detail elsewhere (see Bowman and Martin 1987 and 1988b; Fisher, Goddeeris, and Young 1989), only a brief summary will be provided here. Michigan Amnesty ran for a limited time, from May 12 through June 30, 1986. Under this program, which was part of a comprehensive overhaul of the state's tax enforcement system, individuals were invited to come forward and report tax deficiencies for any state tax. The carrot used to encourage amnesty filings was that all criminal and civil penalties were waived. The stick was that penalties were to be raised significantly after amnesty expired, and there was to be a dramatic increase in the enforcement effort.

Amnesty was viewed as an unqualified success by the Michigan authorities, as almost $110 million was collected, representing a payback of $55 for every $1 expended on the program. It was recognized that these figures overstated the effectiveness of the program, however, because they included collections of receivables from those who were already known to the tax authorities and from those who would have been identified using normal procedures. As with many other states, Michigan both permitted and encouraged those with known, but currently outstanding, deficiencies to participate. Nevertheless, the Michigan Department of Treasury estimates that about 37 percent of the returns and 41 percent of the revenue came from "original cases" that were not in receivables or otherwise previously known to the department.

Because our essay focuses on income tax evasion, only those who filed for income tax amnesty are of current interest. After eliminating the returns based on known receivables, there were 32,614 returns claiming amnesty for income tax deficiencies, which represented 69 percent of the returns filed, and

13. For a comparative survey of the general provisions of various states' amnesty programs, see Mikesell (1986). Summary material also can be found in Dubin, Graetz, and Wilde (1990).

14. The taxes included were income, sales/use, single business, intangibles, inheritance, excise, severance, and public utility property tax. However, almost four-fifths of the revenue collected came from the individual income, single business, and use tax.

15. For details on the increased compliance effort at the state level, see Bowman and Martin (1988a). It should also be pointed out that all amnesty information was forwarded to the IRS.
31 percent of the nonreceivable amnesty revenue. This made income tax filers one of the largest groups of amnesty participants.

To file for income tax amnesty, individuals used the state’s standard amended individual income tax form. This form closely follows the Michigan individual income tax return itself. In fact, it requests figures as originally reported on any previously filed return, the changes in any of those figures, and the new, “correct” figure. This makes it quite simple to measure non-compliance. We simply compared the newly reported figures with the originally reported figures.

We have claimed that Michigan’s Tax Amnesty Program was conventional in most respects. However, there is one very notable difference from most other states, which has to do with the attention given to advance planning for subsequent research. The Michigan Department of Treasury was more interested in extensively investigating amnesty participants than most other state treasury departments. Toward this end, they created the Michigan Amnesty Data Base, which represents a 10 percent random sample of all amnesty participants. This involved expending over 1,250 hours of labor accumulating, entering, and verifying information submitted by participants.

The original one-in-ten sampling included accounts receivable returns, but these were subsequently discarded, as were a number of returns containing problems that made them unusable. The Amnesty Data Base contained 4,203 nonreceivable returns overall, of which 2,985 or 71 percent pertained to income taxes. Of these, 588 were filed by individuals who were amending a return filed previously, while 2,397 were individuals who had not filed previously. The Michigan Department of Treasury believes that this sample is representative of the population of Michigan State Tax Amnesty filers, while at the same time recognizing that it is not likely to be representative of the population of Michigan income tax evaders as a whole. It is this latter problem we hope to neutralize through the self-selection bias correction procedure described previously.

The Empirical Model and Estimation Results

Specifying an Empirical Model of State Evasion

In our empirical analysis, we are concerned only with income tax amnesty filers. Further, we limit our attention to those who had previously filed and were, therefore, filing amended returns under amnesty.16 This means that the

16. There are two reasons for limiting attention to the amended return filers. One is theoretical. Nonfilers are examples of corner solution individuals who are normally explicitly ruled out prior to conducting the comparative static analysis that provides the foundation for any
largest possible sample for our analysis would contain 588 observations. However, the sample actually used in this study is considerably smaller.

Prior to carrying out our econometric analysis, we examined the data for internal consistency. This involved recalculting the tax bill and reconciling this with the net balance due on both the original and amended returns of each of these 588 individuals. In the process, we discovered a number of problem observations. For some observations, all or most of the data were missing from one or both returns. In other cases, obvious taxpayer or data entry errors were found. In still others, we were simply unable to replicate the figures on the tax returns by making the required calculations. In a few cases, the figures suggested negative evasion of one form or another.

After removing the observations with these problems, the sample size was reduced to 213 observations. While this loss of observations is regrettable, we believe that by trimming the sample we are reducing the measurement error and thus may be increasing the signal-to-noise ratio in the data. We are unaware of any systematic bias that these omissions might have introduced into the sample, although we recognize that a large sample size would be preferable given our use of the maximum likelihood estimation procedure.

In analyzing the data in this sample, we must address the two issues raised previously, namely, the self-selection bias inherent in the amnesty data and the interdependency between the federal and state tax returns. Regarding the latter, as a first approximation, we assume the process is recursive in that the decision to evade state income taxes follows the decision to evade federal income taxes.

The individual is assumed to decide how much federal income tax he or she wishes to evade, and then to complete a federal tax return accordingly. Then he or she completes a state tax return conditional on what has been reported at the federal level. Thus, we allow for unidirectional causality from the federal evasion decision to the state tax evasion decision, while neglecting the possibility of full simultaneity between the two. In light of both the relative magnitudes of the two tax bills and the way people typically fill out their returns, this is not unreasonable. In any case, this is an improvement over simply ignoring any interdependency on the grounds that we have little or no theoretical guidance as to what to expect.17

sign expectations. The other reason is tied to measurement and econometrics. One measure of evasion is the amount of income not reported, which in this case is exactly equal to true income. Thus, one might be in the position of wanting to regress income on income. If the measure of evasion is evaded taxes, then one could be regressing this on income and taxes, which is, of course, how evaded taxes was calculated.

17. See, for example, Crane and Nourzad (1990).
Viewing the problem in this way affects how state income tax evasion is measured. Because the Michigan income tax system is so closely tied to the federal system, a significant portion of the evasion that appears on the state return may simply be a carry-through from the federal evasion decision. In other words, a misstatement of federal AGI may be carried forward to the state return for consistency and result in a reduced state tax liability. However, this incidental state level evasion is the result of the federal tax evasion decision and not the decision to evade state income taxes per se and should be netted out. 18

In order to net out the federally driven state evasion we assume that the AGI figure reported on the state return accurately represents what was filed at the federal level. If this is true, any difference between the amended AGI figure and the AGI figure originally reported must be due to evasion traceable to the federal tax return. This difference is then subtracted from the difference between the taxable income reported on the amended and original state income tax returns. The result is a measure of the misstatement of the additions and subtractions to federal AGI when adjusting it for Michigan tax purposes. This is an income-based measure that is, arguably, purely Michigan-based evasion.

Given the extensive tax credit system that is part of the Michigan tax code, we must also take possible credit overstatement into account. Once again, this is determined by comparing the originally reported figures with those on the amended return. To get a measure of overall evasion, we multiply the misstatement of income by the appropriate tax rate and add the result to the misstated credit figure. This gives an overall measure of Michigan income tax evasion not directly carried forward from the federal tax return. 19

This evasion measure is to be regressed on a series of independent variables chosen by drawing on the traditional theoretical evasion models, the results of previously empirical work on evasion, and data availability. In principle, we should control for true income, marginal tax rate, probability of detection, and penalty rate. Unfortunately, data availability and institutional considerations limit our options.

To reflect the individual’s true income level, we use AGI as reported on the amended return. While a broader measure of income would be preferable, none is available to us. The fact that Michigan is a flat rate tax state means

18. In fact, if this is not done, it would amount to building in a spurious positive relationship between Michigan evasion and true income, one of our independent variables. This will become evident as the empirical model is developed.

19. The measure of credit misstatement is the difference between total credits claimed on the amended and the original returns.
that, in any given year, the same tax rate applies to all individuals. Therefore, no marginal tax rate variable can be included in our essentially cross-sectional econometric model of state tax evasion.

The problems identified in the previous section with the compliance policy variables were encountered as expected. The fact that the sample is primarily cross-sectional, coupled with the uniformity of Michigan’s penalty rate across individuals during our sample period, means that no penalty rate can be included in the model. In principle, a similar problem should not be present for the probability of detection; subjective assessments of detection probability certainly vary across individuals. Therefore, it is at least conceptually possible to have a different probability for each individual.

In practice, however, reliable measures of subjective probabilities are difficult to obtain. A common alternative in empirical evasion studies has been to use some measure of the objective audit probability as a proxy. Unfortunately, detailed audit figures are very sensitive information from the perspective of the tax authorities and are often not available.

Because of the difficulty with obtaining direct measures of audit activities, we have resorted to controlling for differential probability of detection effects in an indirect manner. Michigan Treasury documents indicate that, prior to the reorganization that followed amnesty, compliance efforts were divided among Treasury offices in nine geographic regions. Since each office had considerable discretion over enforcement efforts in its area, it seems plausible that Michigan audit rates varied by these geographic regions. Therefore, using zip codes, we associate each amnesty filer with one of these nine geographic regions and construct a series of dummy variables that reflect these regions.

In addition to the variables identified by theory, previous empirical evidence suggests that we should also control for demographic characteristics and differences in opportunities for underreporting certain types of income. In this vein, we include dummy variables to reflect married filers and male filers.\textsuperscript{20} We also include a dummy variable indicating returns that were prepared by professional tax preparers.\textsuperscript{21} Further, in order to reflect differences in the opportunity to underreport, we include a variable measuring the percentage of the overall tax bill that was withheld.

We complete our basic state evasion equation by including one additional

\textsuperscript{20} The latter, of course, only applies to single filers, because married filers must include an individual of each sex.

\textsuperscript{21} The expected sign on the professional tax preparer variable is controversial. In fact, in 1987, the IRS held a conference regarding the role of the tax practitioner in promoting tax compliance at which the topic was debated at great length. Complicating the matter further is the fact that this variable may be endogenous.
variable to control for evasion behavior at the federal level. Given our contention that there may be a link between the federal and the state evasion decisions, we believe that some measure of federal evasion must be a part of an analysis of evasion on the state tax return. Ideally, we would use a measure of the full amount of evasion found on the federal tax return. Of course, doing this would require access to the federal tax returns of the individuals in our sample, which, unfortunately, was not available to us. As an alternative, for our measure of evasion on the federal return, we use the difference between the AGI reported on the amnesty return and that on the original state return.\footnote{\textsuperscript{22}}

We incorporate federal evasion into our model using a two-step procedure, where federal evasion is first regressed on a series of its determinants, and the predicted value is then included in the state evasion equation. Because our measure of federal evasion contains a disproportionate number of zero values (44 out of 213 observations), the tobit estimation procedure is employed.\footnote{\textsuperscript{23}}

It is difficult to form an a priori sign expectation for the coefficient of this variable. The effect will likely depend on whether cheating at the state level is perceived by the evader as a substitute for cheating at the federal level or as a complement to it. In the latter case, one would expect a positive sign for this coefficient, while one would expect a negative sign in the former case.

As with the state evasion model, selection of explanatory variables used in the federal evasion equation is based on a combination of theoretical guidance, past empirical results, and data availability. Two primary variables are true income and the marginal tax rate. To capture the effect of true income, we include measures of AGI and AGI-squared based on the amended return. The quadratic term is included to allow for possible nonlinearities due to both risk aversion and the progressivity of the federal tax system. To capture the direct federal tax rate effect, we include the marginal tax rate that applied to the average taxable income of each individual’s AGI class.\footnote{\textsuperscript{24}}

Reasonable federal tax compliance variables were more difficult to obtain. In fact, we could come up with no penalty rate variable that was Michigan specific, yet had the necessary variance across individuals. The same held

\textsuperscript{22} Obviously this will not reflect all federal evasion, because it misses any federal evasion taking the form of overstated exemptions, deductions, or credits. However, it is the only measure currently available. In addition, if for some reason there are discrepancies between the AGI on the federal return and the state return, our evasion variable will be measured with error.

\textsuperscript{23} In principle, a self-selection correction should be applied at this stage as well. To date, we have not been able to obtain estimates using this more complicated procedure.

\textsuperscript{24} Specifically, we determined the federal AGI class that each individual belonged to using the IRS Statistics of Income—Individual Returns. Next, we calculated the average taxable income for those falling in this particular AGI class. Finally, the marginal tax rate on this average taxable income was determined by consulting the appropriate tax table.
true for a direct measure of the audit rate. However, we do incorporate indirect audit controls by making use of the fact that the IRS categorizes taxpayers into different audit classes based upon return characteristics and subjects these classes to different audit regimes. Based on this, we include dummy variables that loosely categorize our sample individuals into federal audit classes. While it is not clear exactly what we should expect of these variables in terms of sign and significance, it is hoped that their inclusion will be a sufficient control for the differential audit probabilities.

Since the federal evasion decision is also likely to be affected by the opportunities available, we need an additional control for this influence. However, given that we do not have access to the federal returns, no information on federal tax withholding is available. As an alternative, we construct a dummy variable identifying returns of individuals belonging to occupations that tend to have greater opportunities to evade. These include returns of the self-employed, farmers, and salespeople. Finally, we include the two dummies for demographic characteristics discussed earlier—male and married, as well as the dummy reflecting the use of a paid tax preparer. The logic follows that of our earlier discussion.

As discussed previously, our Michigan evasion model cannot be estimated using OLS because the sample is self-selected. In order to apply the maximum likelihood estimation procedure discussed previously, we need to incorporate the participation decision into the likelihood function. This requires identifying factors influencing this decision. There is little formal work to draw on regarding the determinants of amnesty participation. However, recent work using this same Michigan Amnesty Data File by Fisher, Goddeeris, and Young (1989) is helpful. They argue that the decision to participate in amnesty programs is influenced by the perceived increase in the postamnesty penalty and the probability of detection and, to some extent, by nonpecuniary factors such as personal guilt.

As with our evasion equation, there is little we can do with the postamnesty penalty rate because penalties are generally uniform. Therefore, we concentrate on the other factors. The perceived increase in the probability of detection is likely to depend upon what one thinks or expects about the

25. The audit characteristics and classes are as follows. CLASS3: returns with AGI below $10,000, with farm or self-employment income; CLASS4: returns with AGI between $10,000 and $50,000, without farm or self-employment income; CLASS5: returns with AGI between $10,000 and $30,000, with farm or self-employment income; CLASS6: returns with AGI greater than $50,000, but no farm or self-employment income; and CLASS7: returns with AGI greater than $30,000, with farm or self-employment income. In our model, the default category is returns with AGI less than $10,000 and no farm or self-employment income.

26. For additional work on amnesty participation, see Andreoni (1988); Leonard and Zeckhauser (1987); Malik and Schwab (1988).
postamnesty compliance effort. We attempt to capture this influence in several ways.

It is widely known that certain types of income such as rents, royalties, and business income are typically underreported with greater frequency than other income. Thus, we might reasonably expect the tax authorities to concentrate their enforcement efforts in these areas. In this case, returns with these types of income will likely receive increased postamnesty scrutiny, and the evaders who filed returns containing these items might reasonably expect their probability of detection to rise. Based on this reasoning, we construct a dummy variable to identify returns with income from these sources.

Another way to partially capture the effects of a perceived increase in the postamnesty compliance effort is suggested by Michigan Department of Treasury documents. These indicate that special attention was to be focused on the returns of professionals who were licensed by the state. Therefore, we define another dummy variable that identifies returns of individuals in occupations that are typically licensed by states.

We also make a crude effort at allowing for possible nonpecuniary influences using an interesting piece of information contained in the Amnesty Data File. The Data Base identifies all returns in the sample that were accompanied by a special letter of explanation concerning amnesty. Including such a letter is curious behavior given that the idea of amnesty is to induce people to come forward in part because there will be no questions asked. It seems somewhat plausible that those feeling the need to provide a special explanation for their participation are individuals who may be experiencing some type of remorse over their behavior. With this in mind, we specify a dummy variable reflecting the presence of such a letter.

On the other hand, the factor that perhaps has the greatest effect on the perceived increase in postamnesty detection probability reflects an influence that is not directly related to the Michigan amnesty program. It is quite likely that the perceived probability of detection at the state level would rise significantly following an IRS audit. Therefore, such individuals would probably be much more likely to participate in a state amnesty program. To control for this effect, we use information in the Amnesty Data Base to construct a dummy variable indicating that the participant was under IRS audit.

To summarize, the participation decision is incorporated into the likeli-
hood function using dummy variables representing (1) sources of income that are likely to attract attention, (2) occupational categories that had been targeted for additional scrutiny, (3) unusual behavior that may reflect remorse at having evaded, and (4) the fact that a return was under review by the IRS. Obviously, this specification is rather ad hoc and too simple to completely capture the complex participation decision. However, given the data limitations, we believe this is a reasonable first attempt that we hope can adequately neutralize the bias in our sample.

TABLE 1. Tobit Estimates of Unreported Federal Adjusted Gross Income on Original Returns of Michigan Income Tax Amnesty Participants (FEDEVA)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Normalized Coefficient</th>
<th>Asymptotic T-Ratio</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGI</td>
<td>0.89769E-05</td>
<td>3.0916</td>
<td>0.00185</td>
</tr>
<tr>
<td>AGISQ</td>
<td>-0.37268E-06</td>
<td>-3.1909</td>
<td>0.00130</td>
</tr>
<tr>
<td>MTR</td>
<td>0.30711E-01</td>
<td>2.3781</td>
<td>0.01660</td>
</tr>
<tr>
<td>OCCUPTN</td>
<td>0.49835</td>
<td>1.9517</td>
<td>0.05048</td>
</tr>
<tr>
<td>PREPARE</td>
<td>0.76215</td>
<td>4.2921</td>
<td>0.00002</td>
</tr>
<tr>
<td>MALE</td>
<td>0.44382E-01</td>
<td>0.16595</td>
<td>0.86820</td>
</tr>
<tr>
<td>MARRIED</td>
<td>0.88319E-01</td>
<td>0.46735</td>
<td>0.64037</td>
</tr>
<tr>
<td>CLASS3</td>
<td>-0.28912</td>
<td>-0.34919</td>
<td>0.72693</td>
</tr>
<tr>
<td>CLASS4</td>
<td>-0.31495</td>
<td>-0.80268</td>
<td>0.42179</td>
</tr>
<tr>
<td>CLASS5</td>
<td>-0.89862</td>
<td>-1.0774</td>
<td>0.28070</td>
</tr>
<tr>
<td>CLASS6</td>
<td>-1.2412</td>
<td>-2.2636</td>
<td>0.02339</td>
</tr>
<tr>
<td>CLASS7</td>
<td>0.35106</td>
<td>0.46854</td>
<td>0.63915</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.1791</td>
<td>-2.7107</td>
<td>0.00673</td>
</tr>
</tbody>
</table>

Squared correlation between observed and predicted values: 0.2270
Log likelihood: -1815.0489

*Variable definitions are as follows:
AGI = Adjusted gross income on amended return
AGISQ = AGI squared
FEDEVA = Underreported federal AGI
MTR = Federal marginal tax rate
OCCUPTN = 1 if occupation = sales, self-employed, or farming; otherwise = 0
MALE = 1 for male taxpayers; otherwise = 0
MARRIED = 1 for married taxpayers; otherwise = 0
PREPARE = 1 if return prepared by a professional preparer; otherwise = 0
CLASS3 = 1 if AGI ≤ $10,000 with farm or self-employment income; otherwise = 0
CLASS4 = 1 if $10,000 ≤ AGI ≤ $50,000 without farm or self-employment income; otherwise = 0
CLASS5 = 1 if $10,000 ≤ AGI ≤ $50,000 with farm or self-employment income; otherwise = 0
CLASS6 = 1 if AGI ≥ $50,000 without farm or self-employment income; otherwise = 0
CLASS7 = 1 if AGI ≥ $30,000 with farm or self-employment income; otherwise = 0

*Marginal significance level, prob |t| < x.
The Estimation Results

We begin our review of the estimation results with the tobit estimates for the first stage, which explain the magnitude of the understatement of federal AGI (results shown in table 1). All things considered, reasonable results are obtained. Both AGI and AGI squared are statistically significant, and have a positive-negative sign pattern, which suggests that, other things being equal, evasion rises with increases in true AGI, but at a decreasing rate.

The federal tax rate (MTR) has a statistically significant positive effect on evasion. This is consistent in terms of sign and significance with some previous findings (e.g., Clotfelter 1983; Crane and Nourzad 1986). The coefficient of the occupation dummy variable (OCCUPTN) is also positive and statistically significant. This is as expected, given the widely held view that individuals in these occupations have greater opportunities to evade. In contrast, neither of the demographic variables appears to have any relevance for explaining the level of unreported federal AGI.

The remaining parameter estimates are more difficult to interpret. The significant positive coefficient for the professional tax preparer dummy (PREPARE) could spark some controversy. However, given such complications as the blurred distinction between tax evasion and avoidance and the ability of the taxpayer to shop around for a sympathetic preparer, it is difficult to know what to make of this finding. Along these same lines, a reasonable interpretation for the audit class coefficients is not readily apparent. Thus, we make no attempt to draw compliance policy inferences from these estimated coefficients and remain content to treat them as nothing more than necessary control variables.

Of course, the results shown in table 1, however interesting, are a means to an end rather than an end in themselves. This equation has been estimated primarily to obtain the predicted values of FEDEVA that are then fed into the Michigan evasion estimation that is the main focus of our analysis. The maximum likelihood estimates for Michigan, which allow for the effect of the federal evasion decision and correct for self-selection bias, are reported in table 2.29

The predicted value of FEDEVA has a significant positive effect on pure Michigan income tax evasion. This would suggest that state-level evasion is complementary to federal evasion, even when the carry-through effect is

29. Prior to estimation, all continuous variables entering the likelihood function were standardized using the sample means and standard deviations. This was done to improve the convergence properties of the model. OLS results using the same standardized variables are reported in appendix table A1 for comparison purposes.
## TABLE 2. Maximum Likelihood Estimates (with Self-Selection Correction) of Michigan Individual Income Taxes Evaded (MICHEVA)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-Ratio</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEDEVA</td>
<td>0.284086</td>
<td>3.677</td>
<td>0.00024</td>
</tr>
<tr>
<td>AGI</td>
<td>0.410593</td>
<td>2.478</td>
<td>0.01320</td>
</tr>
<tr>
<td>PCTWHD</td>
<td>-0.159595</td>
<td>-2.165</td>
<td>0.03041</td>
</tr>
<tr>
<td>MALE</td>
<td>-0.410599</td>
<td>-1.242</td>
<td>0.21437</td>
</tr>
<tr>
<td>MARRIED</td>
<td>-0.117273</td>
<td>-0.633</td>
<td>0.52676</td>
</tr>
<tr>
<td>PREPARE</td>
<td>-0.482466</td>
<td>-2.526</td>
<td>0.01154</td>
</tr>
<tr>
<td>MTCLEMENS</td>
<td>-0.016751</td>
<td>-0.091</td>
<td>0.92772</td>
</tr>
<tr>
<td>FLINT</td>
<td>-0.338720</td>
<td>-0.616</td>
<td>0.53769</td>
</tr>
<tr>
<td>SAGINAW</td>
<td>0.056257</td>
<td>0.131</td>
<td>0.89605</td>
</tr>
<tr>
<td>LANSING</td>
<td>-0.125622</td>
<td>-0.320</td>
<td>0.74905</td>
</tr>
<tr>
<td>KALAMAZOO</td>
<td>-0.124401</td>
<td>-0.216</td>
<td>0.82882</td>
</tr>
<tr>
<td>GRANDRAPI</td>
<td>0.301049</td>
<td>1.250</td>
<td>0.21138</td>
</tr>
<tr>
<td>TRAVERSE</td>
<td>0.268905</td>
<td>1.170</td>
<td>0.24216</td>
</tr>
<tr>
<td>UP</td>
<td>-0.099677</td>
<td>-0.333</td>
<td>0.73942</td>
</tr>
<tr>
<td>Constant1</td>
<td>0.954183</td>
<td>2.707</td>
<td>0.00678</td>
</tr>
<tr>
<td>IRS-AUDIT</td>
<td>0.321658</td>
<td>0.815</td>
<td>0.41486</td>
</tr>
<tr>
<td>LICENSED</td>
<td>7.609160</td>
<td>0.000</td>
<td>0.99988</td>
</tr>
<tr>
<td>GUilt</td>
<td>-0.188707</td>
<td>-0.513</td>
<td>0.60762</td>
</tr>
<tr>
<td>RENTROYL</td>
<td>-0.338990</td>
<td>-0.487</td>
<td>0.62603</td>
</tr>
<tr>
<td>Constant2</td>
<td>-0.172743</td>
<td>-0.252</td>
<td>0.80118</td>
</tr>
<tr>
<td>$\xi$</td>
<td>0.859215</td>
<td>6.734</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Squared correlation between observed and predicted values 0.4356  
Log likelihood -233.7700  

*Variable definitions are as follows.
- **FEDEVA** = Predicted value of FEDEVA (from table 1)
- **MICHEVA** = Michigan income taxes evaded
- **PCTWHD** = Percentage of Michigan income tax withheld
- **MTCLEMENS** = 1 if taxpayer’s zip code in Mount Clemens area; otherwise = 0
- **FLINT** = 1 if taxpayer’s zip code in Flint area; otherwise = 0
- **SAGINAW** = 1 if taxpayer’s zip code in Saginaw area; otherwise = 0
- **LANSING** = 1 if taxpayer’s zip code in Lansing area; otherwise = 0
- **KALAMAZOO** = 1 if taxpayer’s zip code in Kalamazoo area; otherwise = 0
- **GRANDRAPI** = 1 if taxpayer’s zip code in Grand Rapids area; otherwise = 0
- **TRAVERSE** = 1 if taxpayer’s zip code in Traverse City area; otherwise = 0
- **UP** = 1 if taxpayer’s zip code in Upper Peninsula; otherwise = 0
- **IRS-AUDIT** = 1 if taxpayer under IRS audit; otherwise = 0
- **LICENSED** = 1 if taxpayer’s occupation = architecture, medicine, health, law, personal services, transportation; otherwise = 0
- **RENTROYL** = 1 if rents, royalties, or business income present; otherwise = 0
- **GUilt** = 1 if a letter of explanation accompanied amended returns; otherwise = 0

All other variables are defined in table 1.
eliminated. In other words, on average, pure state-level evasion increases as federal evasion rises. This also suggests that a careful matching of federal and state tax return data may be an effective compliance policy tactic. A similar conclusion might be drawn regarding back-up withholding provisions, given the significant negative relationship between state-level evasion and the proportion of the tax bill withheld (PCTWHD).

According to the data in table 2, true income (AGI) has a significant positive effect on evasion, as one would expect. On the other hand, the demographic variables perform poorly in the Michigan evasion equation, as neither coefficient achieves statistical significance at conventional levels. As for the geographic region dummy variables, none is individually statistically significant. However, it is hoped that, together, they control for differences in audit rates.

In contrast, the professional tax preparer variable (PREPARE) produces an interesting result; its coefficient is statistically significant but has the opposite sign as in table 1. In other words, it appears that, on average, the use of a professional tax preparer is associated with more evasion at the federal level, but less evasion at the state level. This is a bit difficult to explain. One possibility is that tax preparers recognize a close link between the funding of state government and both their own welfare and that of their clients. Of course, it could also be an indication of shortcomings in model specification or variable measurement. This is certainly a topic for further investigation.

In addition to the evasion function coefficients, the maximum likelihood procedure also provides estimates for the participation parameters (as shown in table 2). But little should be made of these estimates themselves. Recall that the participation variables are included to correct for selection bias, and reliable estimates of the corresponding coefficients are not to be expected (Maddala 1983, 267). It is worth noting, however, that \( \xi \), the correlation coefficient between the error terms of the evasion and participation functions, is quite large and is statistically significant, as would be expected given the self-selected nature of a sample of tax amnesty participants.

Concluding Comments

We have employed a new approach to measuring and empirically analyzing income tax evasion. Using data from the Michigan State Income Tax Amnesty Program, we demonstrated how state amnesty data can be utilized to construct a direct measure of evasion at the individual level. Proper empirical analysis of the evasion behavior embodied in this measure required us to address two problems. First, because the data pertained only to voluntary participants in the state amnesty program, we had to employ self-selection bias correction procedures that take the amnesty participation decision into account. Second,
because these data are related to a state that closely links its income tax code to the federal tax system, we had to examine evasion at both levels in order to take the possible interjurisdictional influences into account.

Our analysis indicates that there is a positive linkage from evasion at the federal level to evasion at the state level. In addition, we find that higher income and greater noncompliance opportunity lead to greater evasion at both levels. We also find some support for the proposition that federal marginal tax rates are positively related to federal evasion. Another result, which is somewhat curious, is that the use of a professional tax preparer is associated with more evasion at the federal level, but less evasion at the state level. On the other hand, our results provide no insights into the effects on evasion of either deterrence efforts or demographic characteristics.

The research described here should be of interest for several reasons. Some of the results provide additional evidence regarding the effects of such traditional theoretical determinants of evasion as income and the marginal tax rate. Others confirm the importance of such institutional arrangements as tax withholding as effective compliance policy tools. This latter finding is reinforced by the evidence reported here that state and federal evasion are complementary activities. This suggests that further information sharing and return matching would be a fruitful activity. On the other hand, the conflicting results regarding the role of the tax practitioner are likely to be of value more for the debate they spark than for the results themselves.

Another interesting implication of our research is what it implies about the design and implementation of amnesty programs. We have demonstrated that amnesty data can be useful for analyzing income tax evasion, despite some of the inherent limitations of these data. This would suggest that tax authorities in any jurisdiction who may be contemplating a tax amnesty should give serious consideration to this type of analysis when designing the structure and administration of the program. In hindsight, it is unfortunate that more states did not plan for follow-up analysis and, as a result, much potentially useful data have been destroyed or are not available in a usable form.

This research can be extended and improved in a variety of ways. At the state level, a good place to start is the participation function. Ideally, this decision would be formally modeled in order to identify the relevant explanatory factors more accurately. At the very least, it would be desirable to obtain better measures of the postamnesty penalty and probability variables that are used for our ad hoc participation function. Along the same lines, better treatment of compliance policy instruments in the evasion function itself is needed.

Improvements can also be made in the analysis of evasion at the federal level. Ideally, data from federal returns should be used to get a more complete measure of the extent of federal evasion. Furthermore, this estimation should
be done in a way that allows for both the truncation and self-selection bias problems.

Despite this ample room for improvement, we believe that we have produced some interesting results. Equally important, in our opinion, is that we have demonstrated how amnesty data can be used to gain insights into the issue of income tax compliance. We believe this represents a new and potentially fruitful direction for tax compliance analysis that should be given serious attention.

APPENDIX


<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-Ratio</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEDEVA</td>
<td>0.335475</td>
<td>5.012</td>
<td>0.00000</td>
</tr>
<tr>
<td>AGI</td>
<td>0.427782</td>
<td>7.310</td>
<td>0.00000</td>
</tr>
<tr>
<td>PCTWHD</td>
<td>-0.133573</td>
<td>-2.324</td>
<td>0.02112</td>
</tr>
<tr>
<td>MALE</td>
<td>-0.338994</td>
<td>-1.720</td>
<td>0.08696</td>
</tr>
<tr>
<td>MARRIED</td>
<td>-0.104177</td>
<td>-0.727</td>
<td>0.46830</td>
</tr>
<tr>
<td>PREPARE</td>
<td>-0.538374</td>
<td>-3.791</td>
<td>0.00020</td>
</tr>
<tr>
<td>MTCLEMSENS</td>
<td>-0.107951</td>
<td>-0.721</td>
<td>0.47190</td>
</tr>
<tr>
<td>FLINT</td>
<td>-0.384031</td>
<td>-1.388</td>
<td>0.16679</td>
</tr>
<tr>
<td>SAGINAW</td>
<td>-0.006079</td>
<td>-0.022</td>
<td>0.98248</td>
</tr>
<tr>
<td>LANSING</td>
<td>-0.207721</td>
<td>-0.884</td>
<td>0.37779</td>
</tr>
<tr>
<td>KALAMAZOO</td>
<td>-0.202568</td>
<td>-0.630</td>
<td>0.52911</td>
</tr>
<tr>
<td>GRANDRAPI</td>
<td>0.260335</td>
<td>1.303</td>
<td>0.19900</td>
</tr>
<tr>
<td>TRAVERSE</td>
<td>0.173662</td>
<td>0.646</td>
<td>0.51918</td>
</tr>
<tr>
<td>UP</td>
<td>-0.220065</td>
<td>-0.830</td>
<td>0.40779</td>
</tr>
<tr>
<td>Constant</td>
<td>0.557466</td>
<td>2.696</td>
<td>0.00763</td>
</tr>
</tbody>
</table>

Adjusted $R^2$ | 0.39446  |
SER            | 0.77813  |
$F(14,198)$    | 10.86420 |
Log likelihood | $-241.02306$ |

Note: See Table 2 for variable definitions.

REFERENCES


