Searching for Ghosts: Who Are the Nonfilers and How Much Tax Do They Owe?

by

Brian Erard
Department of Economics
Carleton University
and
Chih-Chin Ho
Internal Revenue Service
Research Division

January 1999

Abstract. This paper is about "ghosts" – individuals who fail to comply with their income tax filing requirements. As their name suggests, the identities and characteristics of these individuals are shrouded in mystery. In this paper we attempt to de-mystify the issues surrounding ghosts and examine their role in the compliance process. We begin by extending a standard model of tax evasion to account for the existence of ghosts. We then examine the empirical significance and policy relevance of our extension using a unique data set containing detailed tax and audit information for both filers and non-filers of U.S. federal income tax returns.

Searching for Ghosts: Who Are the Nonfilers and How Much Tax Do They Owe?

Brian Erard
Department of Economics
Carleton University
and
Chih-Chin Ho
Internal Revenue Service
Research Division

1. Introduction

Over the past three decades, economists have devoted substantial attention to the issue of tax compliance.[1] Most of this literature is concerned with the taxpayer's decision about how much income to report on his tax return and the tax agency's response to this report. A group that has been neglected by this research has been those individuals who simply choose not to file a return, a group sometimes referred to as "ghosts" by academics and policy-makers.[2]

Based on available evidence from Jamaica (Alm et al. (1991)) and the U.S. (Crane and Nourzad (1993)), it appears that nonfiling poses a significant problem for tax agencies. However, very little is known about this form of evasion. In this paper we extend our understanding of ghosts by describing their characteristics, identifying the factors associated with their decision not to file a tax return, and measuring their unpaid tax liability. We begin in section 2 by developing an extended model of taxpayer reporting behavior

---

We are grateful to the Internal Revenue Service for providing us with access to the data used in this analysis. We would particularly like to acknowledge Jeff Colson, Dennis Cox, Carol Sattler, Arthur Sparrow, and Joel Stubbs for many helpful discussions. We also thank Jim Alm, Matt Murray, and Joel Slemrod for valuable comments on an earlier draft. The first author thanks the Social Sciences and Humanities Research Council of Canada for financial assistance. Any opinions expressed in this document are those of the authors; they do not necessarily represent the views of the Internal Revenue Service

[1] See Andreoni, Erard, and Feinstein (1998) for a recent survey of this literature.

[2] The term "ghosts" is borrowed from Cowell (1990), who notes that it is commonly used by Inland Revenue in the U.K. to refer to individuals for whom no official record exists. Refer to Cowell and Gordon (1993) for a theoretical analysis of the role of ghosts in sales tax evasion.

that includes nonfiling as a strategic option. We then examine the empirical significance and policy relevance of our extension using a unique data set containing detailed line-item tax and audit information for both filers and non-filers of U.S. federal income tax returns. Section 3 lays out the econometric framework we employ to estimate our model. Section 4 contains a description of the data, and section 5 presents the results of our analysis. In section 6 the results are employed to generate tables comparing features of the filer and ghost populations. Section 7 provides a discussion of the net tax liabilities of ghosts, and a brief conclusion if offered in section 8.

2. Model Development

In this section a simple theoretical framework is presented for understanding the decision whether to file an income tax return. We begin by considering a standard model of the taxpayer's choice of how much income to report on his return. We then extend the model to account for nonfiling as a strategic option.

The Reporting Decision

Our specification of the taxpayer's reporting decision is essentially the classic specification given by Allingham and Sandmo (1972), with the following two qualifications: (i) the penalty rate applies to unpaid taxes rather than unpaid income (see Yitzhaki (1974) for a discussion of this modification); and (ii) a measure of the compliance burden associated with filing a return is explicitly accounted for. According to this model, in choosing an optimal reporting strategy a taxpayer takes into account the tradeoff between the tax savings from underreporting an additional dollar of income and the increased risks of detection and penalty. Formally, a taxpayer with true taxable income Y chooses an amount of income to report X to maximize the following expression:

$$(1-p)U\left[Y-tX-c\right]+pU\left[Y-tX-(1+\theta)t(Y-X)-c\right], \tag{1}$$

where $U[\bullet]$ is a standard expected utility function, p is the probability of audit, t is the proportional tax rate, θ is the proportional penalty rate on undeclared taxes, and c is

a dollar measure of the burden of complying with income tax filing requirements.[3] If p exceeds $1/(1+\theta)$ the taxpayer will choose to file a correct report ($X = Y$), because the marginal expected cost associated with the first dollar of underreporting exceeds the marginal benefit. Otherwise, the taxpayer's optimal choice X^* is implicitly determined according to the first order condition:

$$(1-p)tU'\left[Y - tX^* - c\right] = pt\theta U'\left[Y - tX^* - (1+\theta)t(Y - X^*) - c\right]. \tag{2}$$

The left-hand side of equation (2) represents the utility gain from successfully evading taxes by an additional dollar, weighted by the probability of not being audited. Analogously, the right-hand side represents the utility loss from having been caught evading taxes by an additional dollar, weighted by the probability of audit. At the optimal level of evasion, the marginal expected benefit of evasion just equals the marginal expected cost.

Accounting for Ghosts

Although the above model of taxpayer reporting behavior is fairly standard in the literature, it fails to account for the possibility that an individual simply may choose not to file a return. According to the results presented in this paper, a substantial fraction of U.S. residents who are legally required to file a federal income tax return fail to do so. In order to account for such ghosts, it is necessary to extend the above model to describe the incentives for not complying with one's filing obligation. Consider an individual who has prepaid an amount W of his tax liability.[4] Suppose that if he fails to file his return, this individual perceives that he will be apprehended with probability q and forced to pay his back taxes plus a penalty. He computes the expected utility associated with not filing a return according to the following formula:

$$(1-q)U\left[Y - W\right] + qU\left[Y - W - (1+f)(tY - W)\right], \tag{3}$$

[3] See Blumenthal and Slemrod (1992) for evidence on the magnitude of the compliance burden. Note that this model could be extended to allow c to be a function of the amount of effort that goes into legal and illegal tax avoidance schemes. See, for example, Cross and Shaw (1982) and Slemrod (1994).

[4] Prepayments include taxes withheld, estimated tax installments, and other payments of potential tax liability prior to the return submission date.

where f is the nonfiler penalty rate that applies to the outstanding tax balance.[5]

Assuming that the level of tax prepayments W is a free choice variable, the individual will compute the value of W that maximizes equation (3).[6] He will then base his decision to file on whether the value of equation (1), evaluated at the optimal level of reported income X^*, exceeds the value of equation (3), evaluated at the optimal level of tax prepayments W^*. Observe that in the absence of a filing burden c, the forms of the objective functions described by equations (1) and (3) are equivalent. Thus if $c = 0$, $p = q$, and $\theta = f$, the optimal choice of tax prepayments W^* under the nonfiling option will be precisely equal to t times the optimal choice of reported income X^* under the filing option, and the individual will be indifferent between filing and not filing.

It follows that an individual will be relatively more likely to become a ghost the greater the filing burden c, the lower the perceived chances for successful underreporting $(1 - p)$, the higher the penalty rate for underreporting θ, and the lower the probability q and rate of penalty f associated with not filing.

Dynamic Considerations

An issue not generally taken into account in models of tax compliance is the dynamic nature of an individual's reporting decisions. However, as discussed by Engel and Hines (1994) and Erard (1992), a taxpayer's past audit and reporting history is likely to have important implications for his current reporting choice. What also should be recognized is that an individual's filing history will tend to have important consequences for his current filing decision. Consider, for example, an individual who failed to comply with his filing requirement for the previous tax year. If this individual were to file a return for the current year, he may perceive that this would increase the probability that his past filing violation

[5] This rate incorporates any penalty attributable to the failure to make sufficient estimated tax payments or to have sufficient taxes withheld.

[6] In practice, a high value of W may provide a signal to the tax agency that the individual possesses sufficient income to have a tax filing requirement. A more general model would account for this possibility by allowing the probability of detection q to vary with W. Analogously, a low report X from a filer may serve as a signal to the tax agency of likely tax noncompliance. Again, a more general model would account for this by allowing the audit probability p to vary with X. However, the main factors influencing the choice between filing and not filing are adequately represented by the simpler fixed audit probability specification presented in this paper.

would be uncovered. By filing a return this former ghost also would provide the tax agency with an up-to-date address, which might increase the odds that he would be located and penalized if his past nonfiling violation were uncovered. For similar reasons, a taxpayer who filed a return in a previous tax year may perceive that the chances of detection would be relatively high if he were to fail to file a return for the current year.[7] The above insights are formalized in Appendix A, where an extended two-period model of the filing and reporting decisions is presented. In the econometric analysis presented below we explicitly take into account the past filing history of the individuals in our sample to account for possible persistence in filing behavior.

3. Econometric Specification

In this section we develop an econometric specification for analyzing an individual's decision whether to comply with his income tax filing requirement. We restrict our attention to individuals who were legally obliged to file a 1988 U.S. federal individual income tax return. The members of this population are divided into two categories, *filers* and *ghosts*, according to whether they have complied with their 1988 filing requirements. As discussed in section 4 our data includes detailed line-item tax and occupation information for individuals from each category. The data on filers comes from a stratified random sample of the overall filer population. The data on ghosts comes from a stratified random sample of the "locatable" nonfiler population. The latter population includes all ghosts who could be located through an intensive search by Internal Revenue Service (IRS) agents. Sample weights are available that make the filers and ghosts in our sample broadly representative of the overall filer and locatable nonfiler populations, respectively. The locatable nonfiler population is of considerable policy interest, because it represents the portion of the overall ghost population that the IRS would be able to uncover through an intensive search and audit process. However, it is also desirable to learn about the frequency of unlocatable nonfilers, the amount of taxes that these individuals owe, and the motivations behind their decision not to file an income tax return. The econometric specification presented below make it possible to draw inferences about all ghosts, whether locatable or not.

[7] In fact, in the U.S. the IRS has what it calls a "stop-filer" program designed to identify and investigate prior year taxpayers who have not filed a return for the current year.

Distributional Assumptions

The theoretical framework presented in section 2 indicates that an individual is more likely to file a return when the likelihood of apprehension for not filing is high. One of the factors that will determine the likelihood of apprehension is the ease with which the tax agency can locate the individual. In our data sample, an intensive search by the IRS agents failed to locate a number of potential nonfilers. We therefore model the probability that an individual can be located jointly with the individual's filing decision.

We define the binary outcomes of the filing decision as follows:

$$F = \begin{cases} 1 & \text{if the individual files a return;} \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, we define the marginal outcomes of the nonfiler search process as:

$$L = \begin{cases} 1 & \text{if the nonfiler is located;} \\ 0 & \text{otherwise.} \end{cases}$$

We specify the joint distribution for F and L using a logistic framework.[8] Let $P_{FL}(F = f, L = l)$ denote the joint probability that $F = f$ and $L = l$ (where $f, l \subset \{0, 1\}$). The joint probability distribution is summarized by the following equations:

$$P_{FL}(F = 1, L = 1) = exp\left(\beta_F' X_F + \beta_L' X_L + K\right)/D \tag{4}$$

$$P_{FL}(F = 1, L = 0) = exp\left(\beta_F' X_F\right)/D \tag{5}$$

$$P_{FL}(F = 0, L = 1) = exp\left(\beta_L' X_L\right)/D \tag{6}$$

$$P_{FL}(F = 0, L = 0) = 1/D, \tag{7}$$

where

$$D = 1 + exp\left(\beta_L' X_L\right) + exp\left(\beta_F' X_F\right) + exp\left(\beta_F' X_F + \beta_L' X_L + K\right).$$

[8] See Nerlove and Press (1973), Mantel and Brown (1973), and Morimune (1979) for prior applications based on this framework.

The term X_F represents a vector of explanatory variables for the probability that an individual will choose to file. Similarly, X_L represents a vector of explanatory variables for the probability that an individual will be located if he elects not to file. The term K represents a measure of the strength of the correlation between the likelihood of filing and the probability of being located.

To understand the relationship between the above specification and an ordinary univariate logit framework, consider the implied conditional probability that F equals one given that L equals zero ($P_{F|L}(F = 1|L = 0)$):

$$P_{F|L}(F = 1|L = 0) = \frac{exp\,(\beta'_F X_F)}{1 + exp(\beta'_F X_F)}.$$

This is clearly a univariate logit specification of the filing decision for those individuals who could not be located if they elected not to file. Similarly,

$$P_{F|L}(F = 1|L = 1) = \frac{exp\,(\beta'_F X_F + K)}{1 + exp\,(\beta'_F X_F + K)},$$

which is a univariate logit specification of the filing decision for those individuals who could be located if they did not file. When $K = 0$, we see that the above two conditional probabilities are the same, implying that F and L are independent events. When $K > 0$, an individual who could be located is more likely to file than one who could not be located, while the converse is true when $K < 0$.

Allowing for Simultaneity

The above specification allows for an interdependence between the probability of filing and the probability that an individual can be located. This relationship, however, is an indirect one, analogous to a correlation between the error terms in a seemingly unrelated regression model. It is plausible that the decision to file is directly related to the probability of being located. In particular, we might expect that an increase in the probability that an individual can be located would have a direct positive impact on the probability that he would choose to file. To account for this possibility, we extend our specification to allow for an element of simultaneity between the probability of filing and the chances of being located:

$$P_{FL}(F=1,L=1) = exp\left(\beta'_F X_F + \alpha\beta'_L X_L + \beta'_L X_L + K\right)/D \tag{8}$$

$$P_{FL}(F=1,L=0) = exp\left(\beta'_F X_F + \alpha\beta'_L X_L\right)/D \tag{9}$$

$$P_{FL}(F=0,L=1) = exp\left(\beta'_L X_L\right)/D \tag{10}$$

$$P_{FL}(F=0,L=0) = 1/D, \tag{11}$$

where D is now defined as

$$D = 1 + exp\left(\beta'_L X_L\right) + exp\left(\beta'_F X_F + \alpha\beta'_L X_L\right) + exp\left(\beta'_F X_F + \beta'_L X_L + K\right).$$

Our extended specification includes an index of the likelihood of being located ($\beta'_L X_L$) as an additional explanatory variable for the filing decision. This extension can best be understood by considering the following simultaneous equations specification:

$$F^* = \beta'_F X_F + \alpha L^* + \epsilon_F \tag{12}$$

$$L^* = \beta'_L X_L + \epsilon_L, \tag{13}$$

where F^* and L^* are latent variables for the propensities to file and be located, respectively, and ϵ_F and ϵ_L are random disturbances. A limited information maximum likelihood (LIML) specification for this framework can be derived by substituting the reduced form for L^* into the equation for F^*, yielding:

$$F^* = \beta'_F X_F + \alpha\beta'_L X_L + \epsilon^*_F \tag{14}$$

$$L^* = \beta'_L X_L + \epsilon_L, \tag{15}$$

where $\epsilon^*_F = \epsilon_F + \alpha\epsilon_L$. This is analogous to our extended logistic specification, in which the index of the likelihood of being located enters directly as an explanatory variable for the filing equation. Thus, our specification can best be thought of as a LIML specification of a simultaneous relationship between the probability of filing and the likelihood of being located.

Conditional Likelihood Function

Our data contain detailed information pertaining to the filing decision for two groups of individuals: filers and located nonfilers. This information is not available, however, for the remaining group (unlocated nonfilers). Given the truncated nature of our sample, it is necessary to condition our analysis of the filing decision on the probabilities of an individual belonging to one of the first two groups.

The conditional likelihood function will involve different expressions for filers and located nonfilers. If we define L_1 as the conditional likelihood associated with the former group,L_1 may be expressed as

$$L_1 = \frac{exp\left(\beta'_F X_F + \alpha\beta'_L X_L\right) + exp\left(\beta'_F X_F + \alpha\beta'_L X_L + \beta'_L X_L + K\right)}{exp\left(\beta'_F X_F + \alpha\beta'_L X_L\right) + exp(\beta'_L X_L) + exp\left(\beta'_F X_F + \alpha\beta'_L X_L + \beta'_L X_L + K\right)}. \quad (16)$$

Similarly, defining L_2 as the conditional likelihood associated with located nonfilers, we have

$$L_2 = \frac{exp(\beta'_L X_L)}{exp\left(\beta'_F X_F + \alpha\beta'_L X_L\right) + exp(\beta'_L X_L) + exp\left(\beta'_F X_F + \alpha\beta'_L X_L + \beta'_L X_L + K\right)}. \quad (17)$$

Two-Stage Estimation Strategy

Since the conditional likelihood function excludes all unlocated nonfilers from the analysis, it can be expected to generate poor estimates of the likelihood that a given nonfiler can be located. This is a common problem in truncated regression specifications. To get around this difficulty, we take advantage of the fact that although details pertaining to the filing decision (X_F) are not available for unlocated nonfilers, we do have details pertaining the chances of being located (X_L) for this group. Observe from equations (10) and (11) that the conditional probability that an individual will be located given that he does not file is of the logistic form:

$$P_{L|F}(L = 1|F = 0) = \frac{exp\left(\beta'_L X_L\right)}{1 + exp\left(\beta'_L X_L\right)}. \quad (18)$$

This observation leads us to estimate the parameters of our model in two stages. First, we estimate β_L by performing a univariate logit analysis of equation (18) using our sample of located and unlocated individuals who did not file. We then substitute the estimated value of β_L into the conditional likelihood function defined by equations (16) and (17) and estimate the remaining parameters (β_F, K, and α). The standard errors for the second stage parameter estimates are adjusted to account for first-stage sampling error using the procedure described in Murphy and Topel (1985).

In section 5, we compare the results of our two-stage approach with the results obtained by maximizing the conditional likelihood function simultaneously over all of the model's parameters.

Choice-Based Sampling

A minor complication to our analysis is that different sampling rates were used to select the filers and nonfilers in our study, resulting in a choice-based sample. Manski and Lerman (1977) have shown that weighting the likelihood function by the inverse of the sampling rates will generate consistent estimates for choice-based samples. We therefore apply this weighting strategy in both of the stages of our analysis.[9]

4. Summary of Data

The data used for filers of 1988 federal income tax returns is based on a 25 percent random subsample of the IRS TCMP Phase III Survey. This survey contains the results of intensive line-by-line audits of a stratified random sample of approximately 54,000 individual income tax returns for tax year 1988. For most line items both the amount that was reported by the taxpayer and the amount that the examiner determined should have been reported are available. In addition, information is recorded about the prior filing history of the taxpayer, and a code is available for the taxpayer's occupational category.[10] A set of sample weights is included to make the data representative of the national return

[9] We adjust the standard errors of our parameter estimates to account for the weighted estimation procedure using the formula presented in Manski and Lerman (1977).

[10] This code is recorded by the IRS examiner based on his assessment of the taxpayer's occupation.

population.[11] Selection into the 25 percent subsample was restricted to taxpayers who were required to file a 1988 return.[12]

The data on potential nonfilers is from the collection-based segment of the IRS TCMP Phase IX Nonfiler Survey for tax year 1988. This survey includes information for a stratified random sample of approximately 23,000 cases from a population of 83 million individuals for whom there was no record of a 1988 individual income tax return. These individuals were identified through a social security number match of IRS tax records with the Social Security Administration Date of Birth/Date of Death Master File, which lists all individuals with valid social security numbers.[13] The potential nonfilers identified through this match include actual ghosts, late filers, and individuals who were not required to file a return.[14] An intensive effort was made by IRS agents to locate each of the individuals in the sample. Information that was known about each individual prior to the search is available, including the individual's age, whether a return had been filed for the previous tax year, and whether information return documents were available for the 1988 tax year.

A total of 18,689 of the 23,286 potential nonfilers in the sample were successfully located through the search process. The sample weights for these 18,689 individuals sum to approximately 57 percent of the potential nonfiler population.[15] Revenue officers had access to information documents and past filing records. Armed with this information they conducted interviews or field visits to determine whether a successfully located individual's

---

[11] These weights do not account for returns that were filed late or for the returns of non-resident taxpayers.

[12] According to our tabulations approximately 9.7 percent of the returns in the TCMP survey, representing 10.1 million households, were not legally required to file a return. In the majority cases these individuals voluntarily filed a return to claim a refund or an Earned Income Credit. In analyzing whether a taxpayer in the sample was required to file, we were unable to identify whether any advance earned credit payments had been received. Therefore, a taxpayer who was required to file solely as a result of having received such payments would have been excluded from the analysis. Since very few individuals received advance earned income credits in tax year 1988, this is not a substantive problem.

[13] Non-residents and individuals without valid social security numbers were excluded from the analysis.

[14] Recall that ghosts (nonfilers) are defined as individuals who fail to file a return in violation of federal filing requirements.

[15] Unlocated individuals in the sample tended to have much larger sample weights as a consequence of the way the sample was stratified.

income was above the filing threshold. Tax returns were secured from 3,549 individuals who were deemed to have been in violation of their tax filing requirements.

A separate segment of the nonfiler survey, the examination-based segment, is used to construct variables for analyzing the filing decision. A random subsample of 2,195 of the 3,549 secured delinquent returns from the collection-based segment were subjected to intensive line-by-line audits. The information recorded in the examination-based segment of the survey is comparable to that recorded in the TCMP Phase III Survey of filers discussed previously. We have adjusted the sample weights for the secured delinquent returns in this file so that they are broadly representative of all located nonfilers from the collection-based segment.[16] An additional adjustment to the sample weights was made to convert the individual-specific sample weights into return-specific weights. This adjustment was necessary to make the data on nonfilers comparable to the data on filers, which are recorded on a return-specific basis. To accomplish this objective, we divided the sample weights for the secured delinquent returns of married joint nonfilers by a factor of two. All else equal a delinquent married couple's return has approximately twice the chance of being included in our sample as a delinquent single individual's return. This is because it would be sufficient for either member of the couple to be included in the sample of located nonfilers for their joint return to be secured.

5. Results of the Analysis

In this section we present the results of our analysis of taxpayer filing behavior. We first present results for the probability that a potential nonfiler can be located, followed by results for the decision whether to file a return.

[16] The collection-based segment identifies a total of 4,563 individuals who failed to comply with their filing requirement, including the 3,549 from whom returns were secured. The collection-based segment divides returns into 23 sampling strata based on factors such as the presence or absence of information returns, the amount of income shown on those returns, the individual's filing history, and age. Within each stratum, all individuals have the same sample weight. For each of the 23 sampling strata employed for sample selection, we adjusted the sample weights for the returns in the examination-based segment upwards so that the sum equaled the stratum total for the nonfilers in the collection-based segment.

Locating Potential Nonfilers

The first stage of the two-stage analysis involves logit estimation of whether an individual who does not file a return can be located. In addition to a constant term, the following variables are used as regressors (X_L) in this stage of the analysis:

1. **PRIOR YR. FILER:** Dummy variable equal to one if the individual filed a 1987 income tax return; zero otherwise.
2. **IRP INCOME DUMMY:** Dummy variable equal to one if there is an information returns program (IRP) record of 1988 income; zero otherwise.
3. **PRIOR YR. FILER * IRP INCOME DUMMY:** Interaction of the above two dummy variables.
4. **AGE 65 OR OVER:** Dummy variable equal to one if the individual's age is sixty-five or greater; zero otherwise.
5. **SPOUSE:** Dummy variable equal to one if available records indicate a spouse; zero otherwise.

Variables pertaining to the presence of a prior tax return are included, because the information from these returns may be useful for identifying the current or previous address of the individual. Variables relating to the presence of information return documents are included, because these documents may contain information about the individual's address, his place of work, or where he holds financial accounts. The age 65 or over dummy is included to account for the possibility that elderly individuals are relatively easy to locate compared to young individuals. Elderly individuals, for example, may be more firmly established in their communities. The dummy variable for a record of a spouse is included to account for the possibility that married taxpayers may be more readily located.

The results of the logit analysis of the probability of being located are presented table 1.[17] Each of the parameter estimates are of the expected sign, and they all are statistically significant. The interaction between the prior year return and IRP income dummies is negative and rather large, indicating that having both a record of a prior year return

[17] The analysis incorporates the sampling weights, which make the observations representative of the overall population of individuals who did not file a return.

and IRP information only slightly increases the odds of locating an individual over having one of these sources of information.

The Decision Whether to File

In addition to the constant term, the following variables were included as regressors (X_F) for the filing decision:

1. **PRIOR YR. FILER:** Dummy variable equal to one if the individual filed a 1987 income tax return; zero otherwise.
2. **FILING BURDEN:** An IRS estimate of the number of hours required to complete the tax return.
3. **FILING THRESHOLD:** A dummy variable equal to one if the individual's gross income is within 5 percent of the filing threshold level for his age and filing status; zero otherwise.
4. **BURDEN*THRESHOLD:** Interaction between the above two variables.
5. **STATE TAX DUMMY:** Dummy variable equal to one for residence in a jurisdiction with a state-level income tax; zero otherwise.
6. **SCH. C Dummy:** Dummy variable equal to one if Schedule C (business) income or loss is present; zero otherwise.
7. **SCH. F Dummy:** Dummy variable equal to one if the chedule F (farm) income or loss is present; zero otherwise.
8. **GEN. MGR/TOP EXEC.:** Dummy variable equal to one if the individual is a general manager or top executive; zero otherwise. (This dummy is excluded from the analysis, making this the omitted occupation category.)
9. **GOVT. OFFIC./ADMIN.:** Dummy variable equal to one if the individual is a government official or administrator; zero otherwise.
10. **ACCOUNTANT:** Dummy variable equal to one if the individual is an accountant; zero otherwise.
11. **ATTORNEY:** Dummy variable equal to one if the individual is an attorney; zero otherwise.

12. **CONSTRUCTION:** Dummy variable equal to one if the individual is a construction trade worker; zero otherwise.
13. **HELPER/HANDLER:** Dummy variable equal to one if the individual is a helper, handler, equipment cleaner, laborer, or parking lot attendant; zero otherwise.
14. **GARAGE WORKER:** Dummy variable equal to one if the individual is a garage or service station worker; zero otherwise.
15. **AUTO MECHANIC:** Dummy variable equal to one if the individual is a motor vehicle mechanic; zero otherwise.
16. **VEHICLE OPERATOR:** Dummy variable equal to one if the individual is a motor vehicle operator; zero otherwise.
17. **INSURANCE AGENT:** Dummy variable equal to one if the individual is an insurance agent; zero otherwise.
18. **OTHER:** Dummy variable equal to one if the individual's occupation is different from those identified by the other occupation dummies; zero otherwise.
19. **AGE 65 OR OVER:** Dummy variable equal to one if the individual's age is 65 or greater; zero otherwise.
20. **MARRIED:** Dummy variable equal to one if the individual's filing status is married joint return; zero otherwise.
21. **# DEPENDENTS:** Number of dependents.
22. **UNEMPL. INC. DUMMY:** Dummy variable equal to one if the individual received unemployment income; zero otherwise.
23. **AGI:** Adjusted gross income divided by $100,000. (If AGI is negative, AGI is set equal to zero.)
24. **LOCATABILITY:** Index of the likelihood of being located (equal to $\beta_L' X_L$ in equation (18)).

The variables related to income, occupation, and filing status were based on the examiner-determined values rather than those reported by the taxpayer. Due to noncompliance, the former are likely to be more representative of the true values of these variables.

Based on the theoretical framework presented in section 2 and Appendix A, the decision whether to file a return should depend on an individual's past filing behavior, the burden

associated with filing, the opportunities for successfully underreporting income, and the chances of being caught and penalized for not filing. The dummy variable for the presence of a 1987 tax return is included to account for the individual's past filing history. The IRS employs a mathematical formula to estimate the number of hours it would take to complete a tax return for an individual with given sources of income and deductions. We apply this formula to measure the filing burden. We also include a dummy variable for whether an individual's income is close to the filing threshold and an interaction between the burden measure and the threshold dummy. Our intuition is that an individual may elect not to file if his income is only marginally above the threshold, particularly if his return is difficult to complete.[18]

The dummy variable for residence in a jurisdiction with a state income tax might be expected to have a positive association with filing a return. To the extent that such states also have nonfiler detection programs and share information with the federal government, an individual from a state with its own tax may perceive a greater risk of penalty for not filing. It is difficult to predict the sign on the business and farm income dummies a priori. An individual with these sources of income may have relatively good opportunities for underreporting income if he files. On the other hand, to the extent that his income from these sources is "off-the-books", he may have relatively good opportunities for not filing as well.[19]

We control for the influence of a variety of occupations on the filing decision. One would expect that a relatively high percentage of those employed in the first four occupation categories would file returns given the visibility of their jobs and the importance of their reputations. Individuals in the next five categories may be disproportionately represented within the nonfiler population given the relatively good opportunities for concealing income from such occupations.

---

[18] Taxpayers may be able to reduce their filing burden by paying a tax practitioner to complete their returns. Refer to Erard (1993) for an analysis of the decision to use a tax preparer and its consequences for reporting compliance.

[19] As discussed by Simon and Witte (1982) it is commonly believed that individuals with substantial "off the books" income are disproportionately represented among the nonfiler population.

We also control for a number of demographic characteristics in our analysis, including age (whether age 65 or over), marital status, number of dependents, receipt of unemployment insurance, and income.

The final explanatory variable is an index of the likelihood that an individual could be located if he were to choose not to file. It is included to allow for the possibility that individuals for whom the potential for being located is high are relatively more likely to perceive a high probability of detection and penalty for not filing. If so, the coefficient on this variable should be positive.

Table 2 presents the results for the analysis of the decision whether to file an income tax return based on our two-stage estimation procedure. In addition to providing the estimated parameter values and associated t-statistics, we have included estimates of the marginal effect for each variable on the probability of filing. These measure the marginal change in the probability of filing a return for an increase in a given variable, holding all other variables fixed.[20]

The marginal effect for a given variable will tend to vary, depending on the values of the variables being held fixed. For this reason, two separate sets of marginal effects are provided. The first set is computed using the weighted mean values of the variables over the entire sample. The second set is computed using the weighted mean values of the variables over the subsample of nonfilers. Thus, the first set will provide an indication of the marginal effect for an individual with the average characteristics of the overall population, while the second will provide an indication of the marginal effect for an individual with the average characteristics of the ghost population.

For a given occupation dummy variable, this marginal effect is computed by taking the difference between the probability of filing when that occupation dummy is equal to one, the remaining occupation dummies are all zero, and the other variables are held at their mean values, and the probability of filing when that occupation dummy is zero and the other occupation dummies and all other variables are held at their mean values. The

[20] The probability of filing is computed as $\frac{exp(\beta_F' X_F + \alpha\beta_L' X_L + \beta_L' X_L + K) + exp(\beta_F' X_F)}{1 + exp\left(\beta_L' X_L\right) + exp\left(\beta_F' X_F\right) + exp\left(\beta_F' X_F + \alpha\beta_L' X_L + \beta_L' X_L + K\right)}$. The value of $\beta_L' X_L$ is held constant in the computation of the marginal effects of all variables other than the index, itself.

marginal effects for the non-occupation dummies are computed as the difference between the probability of filing when the dummy is equal to one and all other variables are held at their mean values and the probability of filing when the dummy is equal to zero and all other variables are held at their mean values.

As expected, their is substantial persistence in filing behavior. An individual who filed in the previous year is very likely to file in the current year. The first set of marginal results (based on the overall sample variable means) indicates that having filed last year increases the probability of filing this year by 34 percent. The second set of marginal results (based on the nonfiler subsample variable means) indicates that having filed previously raises the chances of filing in the current year by 60 percent!

The estimated coefficients on the burden and threshold variables are not statistically significant. However, the coefficient on the interaction between these variables is negative and significant. For an individual whose income is near the filing threshold, the estimated marginal effect of a one hour increase in the time necessary to complete a return is about a 2 percent increase in the probability of not filing

Individuals with business income as well as those employed as helpers, handlers, service station workers, vehicle mechanics, or insurance agents are relatively less likely to file a return than those in other lines of work. Interestingly, the results indicate that attorneys and government officials are somewhat less likely to file than individuals in several other occupations (including chief executives, general managers, and accountants). However, they are more likely to file than those in the occupations listed previously.

The elderly and the unemployed are relatively less likely to file. However, the other demographic controls (marital status, number of dependents, and income) are not significantly related to the filing decision.

The estimated coefficient of the index for the likelihood that an individual can be located is positive and significant. A one unit increase in this index, evaluated weighted mean value of the index for the ghost population, results in an 11.4 percent increase in the likelihood of being located. The marginal effect of 8.9 percent is therefore quite large, suggesting that the filing decision is very responsive to an increase in the odds of being located.

The estimated value of parameter K, which measures the strength of the correlation between the probability of filing and the probability of being located is also positive and significant. This indicates that unobserved factors which make an individual easier to locate also tend to make him likely to file.

Table 3 presents the results obtained from estimating all parameters in a single stage by maximizing the conditional likelihood function over the truncated sample. The results for the filing decision are quite similar to those presented in table 2.

6. A Comparison of Filer and Nonfiler Return Characteristics

In this section we employ the results of our econometric analysis to generate statistics on nonfiler income, adjustment, and deduction characteristics. We compare these statistics with the corresponding values for the filer population.

We provide separate estimates for the "locatable" ghost and overall ghost populations. The former population is defined as the set of ghosts who would be located if an intensive search were performed by IRS for all potential nonfilers. The latter is defined as the entire ghost population, including those ghosts who would not be located through an intensive search. To generalize our located nonfiler results to the overall ghost population, we adjust the sample weights for located nonfilers using the first-stage probability estimates from the two-stage analysis of section 5. Specifically, the original sample weight for each located nonfiler is divided by the logit-based estimate of the probability that the individual would be located. Our statistics for the overall ghost population are then computed based on the adjusted weights. Our statistics for the filer population are based on a weighted analysis of the complete TCMP Phase III Survey data file. We exclude those taxpayers who were not required to file from the analysis. Again, the statistics are computed using the examiner-determined values for the relevant variables.

Table 4 summarizes income and deductions for filers, locatable ghosts, and all ghosts. Relative to ghosts, filers tend to have substantially larger incomes. For example, their total income before adjustments is on average over two and one-half times larger than that of nonfilers. Taxable income for filers represents 68.8 percent of total income before adjustments. For ghosts, taxable income represents 71 percent of total income before

adjustments, indicating that nonfilers have relatively fewer offsets to income. Intuitively, nonfilers have little incentive to participate in tax planning. Ghosts are relatively less likely to have itemized deductions in excess of the standard deduction threshold. Interestingly, though, among those ghosts whose deductions exceed the threshold, the average total deduction is actually larger than that of filers who itemize. Table 4 also indicates that income is on average larger for locatable ghosts than for the overall ghost population. However, their mean income is still only about half that of filers.

Table 5 displays income, adjustment, and itemized deduction amounts as a percentage of total income before adjustments for filers, locatable ghosts, and all ghosts. Wages and salaries, interest, dividends, and pension income make up a much more substantial share of total income for filers than nonfilers, while business income and net capital gains receipts are relatively more important for nonfilers. The findings for wages and salaries and business income reflect the fact that the ghost population includes a disproportionate share of self-employed individuals. The findings for interest, dividends, and pension income may reflect an aversion by nonfilers to leaving a paper trail. A possible explanation for the capital gains finding is that nonfilers have relatively less incentive to offset taxable capital gains with capital losses. Perhaps for similar reasons, discretionary adjustments and itemized deductions tend to be relatively less important as a share of total income for nonfilers than they are for filers.

7. Net Tax Liability

We have used our adjusted sample weights for located nonfilers to generate an estimate of the net tax liability of the overall ghost population. The results indicate that ghosts owed approximately $5 billion in income taxes for tax year 1988 after subtracting tax prepayments such as witholding and estimated tax payments they had made. In addition, they owed approximately $2.8 billion in self-employment taxes. These figures will tend to understate the true net tax liabilities of ghosts, because even experienced examiners are unable to uncover all unpaid taxes. In its most recent tax gap report (Internal Revenue Service, 1996), the IRS has used an approach similar to ours to estimate the nonfiler

tax gap.[21] However, an adjustment was made using the IRS tax model and assumptions about the degree to which income not subject to information reporting can be detected to account for undetected tax liability. The official estimate of nonfiler net income tax liability (excluding self-employment taxes) for tax year 1988 amounts to $11 billion after adjusting for undetected noncompliance. No official figure is available for understated self-employment taxes. The estimated size of the ghost population based on our approach is 7.9 million.[22] The IRS estimate of the tax gap for the 110 million filers of tax year 1988 returns is $73 billion. Thus, while we find that the number of ghosts is only 7/100 as large as the number of filers, the IRS estimate of nonfiler tax gap is approximately 15/100 as large as the filer tax gap.

As discussed previously, even an intensive search by IRS was unable to locate all potential nonfilers. However, as shown in table 4, locatable nonfilers tend to have higher incomes than those who are not locatable. It seems likely, therefore, that locatable nonfilers are responsible for a large share of the overall nonfiler tax gap. In fact, our results (based on detected net tax liabilities) indicate that approximately 82 percent of the overall nonfiler tax gap is attributable to locatable nonfilers.

8. Conclusion

Nonfilers have been a neglected group in theoretical and empirical research on tax compliance. Much of this neglect has been due to the lack of reliable information about their characteristics, a problem so severe that nonfilers are sometimes referred to as "ghosts" by academics and policy-makers. This study provides important evidence on the characteristics of nonfilers and the taxes for which they are liable. We find that nonfilers tend to be concentrated in those occupations where income is less visible to the tax authority. In addition, for taxpayers with income near the filing threshold, the burden associated with

[21] In the preliminary stage of our research, we employed a probit analysis of the probability an individual could be located rather than a logit analysis. The results were quite similar. The IRS employed our probit analysis in generating its tax gap estimates using a somewhat different weighting scheme than that employed in this study.

[22] This is a return-based estimate, meaning that it represents the number of returns that should have been filed but were not.

completing a return serves as a deterrent to filing. Thus, programs that reduce the burden of filing (such as existing taxpayer assistance programs and simplified tax returns) may encourage individuals to file, particularly those who find themselves near the filing threshold. Further, there appears to be substantial persistence in filing behavior; an individual who files in one year is quite likely to file in the subsequent year. Thus, once a ghost is brought into the system, he is likely to continue to file.

Identifying ghosts and encouraging them to file is a challenging task. The results of this study indicate that only 57 percent of the potential nonfiler population could be located through an intensive search. However, locatable nonfilers apparently account for a disproportionate share of all unpaid taxes. Thus, a substantial portion of the nonfiler tax gap is at least potentially collectable. The extent to which it is cost-effective and/or socially desirable to search out nonfilers and recover the gap is an important question for future research.

Appendix A: The Dynamics of Filing Behavior

Here we present a two-period model of the decision whether to file a tax return. This model allows us to address the dynamics of filing behavior in a simple way. The first period of the model has the same structure as the static model presented in section 2. If the individual files a return and reports X_1 in period 1, his expected utility that period will be:

$$(1-p)U\left[Y_1 - tX_1 - c\right] + pU\left[Y_1 - tX_1 - (1+\theta)t(Y_1 - X_1) - c\right], \qquad (A1)$$

where Y_1 is actual period 1 taxable income, c is the filing burden, t is the proportional tax rate, θ is the penalty rate on underreported taxes, and p is the audit probability.
Should the individual elect not to file a return, his expected utility for period 1 is:

$$(1-q)U\left[Y_1 - W_1\right] + qU\left[Y_1 - W_1 - (1+f)(tY_1 - W_1)\right], \qquad (A2)$$

where W_1 is the level of tax prepayments in period 1, f is the penalty rate on unpaid taxes, and q is the probability of detection.

The individual is forward-looking and realizes that his choice for period 1 may impact on his 2nd period utility. Let us first consider the scenario where the individual has filed in period 1. We assume that the probability of audit if he should file in period 2 remains at p, regardless of whether he has been audited in period 1.[23] Further, we assume that if this individual should file in period 2, only his second period evasion will be subject to taxation and penalty if he should be audited in that period. Since our model consists only of two periods, we assume that each period represents a number of years, so that the statute of limitations for unreported taxes in the first period will have expired by the start of the second period.[24] On the other hand, should the individual elect not to file in period 2, we assume that the probability of detection will exceed q – the probability that

[23] In the U.S., it is not publicly known whether taxpayers who have been audited by the IRS and found to be noncompliant face a higher audit risk in subsequent years. For simplicity, we follow Engel and Hines (1994)in treating the audit probability as unchanged.

[24] A three year statute of limitations is in effect in the U.S.

not filing in the first period would be detected. In particular, if the individual's report was audited in period 1, we assume that he will be detected for sure if he fails to file in period 2. If he is not audited in period 1, we assume that the probability that he will be caught if he does not file in period 2 equals ($\tilde{q} \geq q$). The fact that the individual filed in the first period is likely to make the tax agency suspicious if no return is filed in the second period (particularly if his period 1 report has been audited). Moreover, the presence of a first period return will provide the tax agency with details about the taxpayer (home address, period 1 employer, etc.) that are likely to make it easier to locate the taxpayer and confirm that a return should have been filed.

Consider now the scenario where the individual has not filed a period 1 return. Should the individual decide to file in period 2, we assume that the probability of audit ($\tilde{p}$) will be greater than p – the probability of audit he would face if he had also filed in period 1. By filing in period 2 the individual would provide a signal to the tax agency that he may have had a filing obligation in period 1 as well.[25] Moreover, his second period return would provide details that would make it easier for the tax agency to locate the individual and confirm that a period 1 return was required. If the individual files in period 2 and is audited, we assume that he will be penalized not only for any underreported taxes in that period, but also for any unpaid taxes in period 1. In the U.S. there is no statute of limitations for imposing back taxes and civil penalties on nonfilers.[26] Should the individual elect not to file in period 2, we assume that the probability of detection depends on whether his past nonfiling violation was detected in period 1. If it was detected, we assume that the individual will be caught with certainty if he again fails to file in the second period. On the other hand, if his past violation was not detected in the first period, we assume that the probability of detection in period 2 remains at q. If the individual is caught not filing during the second period, we assume that he will be subject to back taxes and penalties for both periods.

[25] If his past nonfiling violation was already uncovered in period 1, it is likely that the tax agency would give his return in period 2 greater scrutiny than that of an habitual filer.

[26] There is, however, a six year statute of limitations for imposing criminal sanctions for nonfiling.

Under the above assumptions, the individual will develop a state-contingent filing strategy. Define F_1 as the outcome of filing in period 1 and $\bar{F}_1$ as the outcome of not filing in period 1. Similarly let $F_{2|H}$ represent the outcome of filing in period 2 conditional on the individual's period 1 filing and audit history (H), and define $\bar{F}_{2|H}$ as the outcome of not filing in period 2 conditional on the individual's period 1 history. This history ($H = \{F, A\}$) includes the period 1 filing outcome (F_1 or $\bar{F}_1$) and audit outcome (A_1 if the individual was audited or $\bar{A}_1$ if he was not audited).[27] The individual will develop a state-contingent filing strategy, which may involve different second period choices depending on whether his first period reporting (or filing) violation has been uncovered. In all the individual has a choice of eight possible strategies:

$$\left\{F_1 \cap \left(F_2|\{A_1, F_1\} \cup F_2|\{\bar{A}_1, F_1\}\right)\right\} \tag{A3}$$

$$\left\{F_1 \cap \left(F_2|\{A_1, F_1\} \cup \bar{F}_2|\{\bar{A}_1, F_1\}\right)\right\} \tag{A4}$$

$$\left\{F_1 \cap \left(\bar{F}_2|\{A_1, F_1\} \cup F_2|\{\bar{A}_1, F_1\}\right)\right\} \tag{A5}$$

$$\left\{F_1 \cap \left(\bar{F}_2|\{A_1, F_1\} \cup \bar{F}_2|\{\bar{A}_1, F_1\}\right)\right\} \tag{A6}$$

$$\left\{\bar{F}_1 \cap \left(F_2|\{A_1, \bar{F}_1\} \cup F_2|\{\bar{A}_1, \bar{F}_1\}\right)\right\} \tag{A7}$$

$$\left\{\bar{F}_1 \cap \left(F_2|\{A_1\, \bar{F}_1\} \cup \bar{F}_2|\{\bar{A}_1, \bar{F}_1\}\right)\right\} \tag{A8}$$

$$\left\{\bar{F}_1 \cap \left(\bar{F}_2|\{A_1, \bar{F}_1\} \cup F_2|\{\bar{A}_1, \bar{F}_1\}\right)\right\} \tag{A9}$$

$$\left\{\bar{F}_1 \cap \left(\bar{F}_2|\{A_1, \bar{F}_1\} \cup \bar{F}_2|\{\bar{A}_1, \bar{F}_1\}\right)\right\} \tag{A10}$$

It is useful to begin an analysis of the optimal filing strategy by considering the case where the individual would just be indifferent between filing and not filing in a one period model. In particular, suppose that $p = q$, $c = 0$, and $\theta = f$. If we also assume that $\tilde{p} = p$ and $\tilde{q} = q$, it is evident that for each of the following four pairs of strategies, the individual will be indifferent among the two choices.

1) $\left\{F_1 \cap \left(F_2|\{A_1, F_1\} \cup F_2|\{\bar{A}_1, F_1\}\right)\right\}$ and $\left\{F_1 \cap \left(F_2|\{A_1, F_1\} \cup \bar{F}_2|\{\bar{A}_1, F_1\}\right)\right\}$

2) $\left\{F_1 \cap \left(\bar{F}_2|\{A_1, F_1\} \cup F_2|\{\bar{A}_1, F_1\}\right)\right\}$ and $\left\{F_1 \cap \left(\bar{F}_2|\{A_1, F_1\} \cup \bar{F}_2|\{\bar{A}_1, F_1\}\right)\right\}$

[27] We use the same notation whether an audit was of a taxpayer or of a nonfiler.

3) $\left\{\bar{F}_1 \cap \left(F_2|\{A_1, \bar{F}_1\} \cup F_2|\{\bar{A}_1, \bar{F}_1\}\right)\right\}$ and $\left\{\bar{F}_1 \cap \left(F_2|\{A_1, \bar{F}_1\} \cup \bar{F}_2|\{\bar{A}_1, \bar{F}_1\}\right)\right\}$

4) $\left\{\bar{F}_1 \cap \left(\bar{F}_2|\{A_1, \bar{F}_1\} \cup F_2|\{\bar{A}_1, \bar{F}_1\}\right)\right\}$ and $\left\{\bar{F}_1 \cap \left(\bar{F}_2|\{A_1, \bar{F}_1\} \cup \bar{F}_2|\{\bar{A}_1, \bar{F}_1\}\right)\right\}$

Intuitively, if the individual has not been audited in period 1, the expected utility is the same under the filing and nonfiling options for a given level of evasion in period 2. It further can be shown that the individual will prefer the first pair of strategies over all the others, and the last pair of strategies will be least preferred. The first pair of strategies involves filing a return in period 1. Filing in period 1 is preferable, because there is a statute of limitations on penalties for evaders who file. In contrast, evaders who do not file face a chance of being penalized in period 2 when their evasion has not been discovered in the preceding period. The first pair of strategies also involves filing in period 2 when the period 1 report has been audited. This is preferable to not filing, because an individual who fails to file in period 2 following an audit in the preceding period will be audited with certainty. Thus, not filing effectively eliminates the scope for evasion in the second period (if the individual does not file, he will find it optimal to prepay his entire period 2 tax liability if he has been audited in period 1). On the other hand, the individual may have an attractive opportunity for evasion if he does file in period 2 following a period 1 audit. The fourth pair of strategies is least preferred, because it involves neither of the desirable actions of filing in period 1 or filing subsequent to being audited in period 1. In general, it is not possible to rank the second and third pairs of strategies in order of preference, although they each clearly are less desirable than the first pair and more desirable than the fourth.

To investigate how the preference ranking of strategies changes as parameters of the model are altered, we have carried out a series of simulations. These simulations are based on a constant relative risk aversion utility function ($U(W) = W^{\alpha}$). Here we describe some of the most salient findings. First, consider increasing the filing burden in each period from zero. This tends to make filing less attractive. For modest values, we find that $\left\{F_1 \cap \left(F_2|\{A_1, F_1\} \cup \bar{F}_2|\{\bar{A}_1, F_1\}\right)\right\}$ is a dominant strategy. For larger values of burden in each period, $\left\{F_1 \cap \left(\bar{F}_2|\{A_1, F_1\} \cup \bar{F}_2|\{\bar{A}_1, F_1\}\right)\right\}$ is dominant, and for very large values, never filing $\left\{\bar{F}_1 \cap \left(\bar{F}_2|\{A_1, \bar{F}_1\} \cup \bar{F}_2|\{\bar{A}_1, \bar{F}_1\}\right)\right\}$ is dominant.

If instead of increasing the filing burden, we increase both $\tilde{q}$ and $\tilde{p}$ so that they are larger than q and p, respectively, we find that always filing $\left\{F_1 \cap \left(F_2|\{A_1, F_1\} \cup F_2|\{\bar{A}_1, F_1\}\right)\right\}$

is dominant. For more general configurations of the other parameters, we find that when both $\tilde{q}$ and $\tilde{p}$ are large relative to q and p, the same filing choice tends to persist from one period to the next. In other words, if an individual prefers to file in period 1, he will also prefer to file in period 2. Similarly, if he does not choose to file in period 1, he will not choose to file in period 2 either, unless his evasion has been detected in period 1.[28]

With our constant relative risk aversion specification (which implies decreasing absolute risk aversion), the individual will tend to evade by less when he does file as the filing burden increases. Similarly, as the tax rate t increases, the optimal level of evasion under both the filing and nonfiling options will tend to decrease, again due to decreasing absolute risk aversion. Finally, as the rate at which period 2 income is discounted increases, nonfiling in period 1 becomes relatively more attractive. Intuitively, the lack of a statute of limitations for nonfiler evasion becomes relatively less important as second period income is more heavily discounted.

[28] In the latter case, he will generally find it desirable to file in period 2. However, if the burden of filing is sufficiently high (and opportunity for underreporting taxes sufficiently low), he may prefer the option of prepaying all tax liability in period 2 and not filing.

References

Andreoni, J., B. Erard, and J.S. Feinstein. (1998) "Tax Compliance," *Journal of Economic Literature*, **36**, 818–860.

Allingham, M.G. and A. Sandmo. (1972) "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, **1:3/4**, 323–338.

Alm, J., R. Bahl, and M.N. Murray. (1991) "Tax Base Erosion in Developing Countries," *Economic Development and Cultural Change*, **39:4**, 849–872.

Blumenthal, M. and J. Slemrod. (1992) "The Compliance Cost of the U.S. Individual Income Tax System: A Second Look After Tax Reform," *National Tax Journal*, **45:2**, 185–202.

Cowell, F.A. (1990) *Cheating the Government: The Economics of Evasion.* Cambridge, MA: MIT Press.

Cowell, F.A. and J.P.F. Gordon. (1993) "Auditing With Ghosts," Paper presented at the 49th Congress of the International Institute of Public Finance, Berlin, August 23-26.

Crane, S.E. and F. Nourzad. (1993) "An Empirical Analysis of Factors that Distinguish Those Who Evade on Their Tax Return From Those Who Choose Not to File a Return," Paper presented at the 49th Congress of the International Institute of Public Finance, Berlin, August 23-26.

Cross, R.B. and G.K. Shaw. (1982) "On the Economics of Tax Aversion," *Public Finance*, **37**, 36–47.

Engel, E.M.R.A. and J.R. Hines, Jr. (1994) "Retrospective Audits and Tax Evasion Dynamics in the United States," Mimeo, Harvard University.

Erard, B. (1992) "The Influence of Tax Audits on Reporting Behavior," in *Why People Pay Taxes: Tax Compliance and Enforcement*, ed. J. Slemrod. Ann Arbor: University of Michigan Press.

———. (1993) "Taxation with Representation: An Analysis of the Role of Tax Practitioners in Tax Compliance," *Journal of Public Economics*, **52**, 163–197.

Graeber, Michael J., Bonnie L. Nichols, and D. Arthur Sparrow. (1992) "Characteristics of Delinquent Returns," U.S. Department of the Treasury, Internal Revenue Service, The IRS Research Bulletin, Publication 1500, pages 38–46, December, Washington, D.C: GPO.

Internal Revenue Service. (1996) "Federal Tax Compliance Research: Individual Income Tax Gap Estimates for 1985, 1988, and 1992," U.S. Department of the Treasury, Internal Revenue Service, Publication 1415 (Rev. 4-96), Washington, D.C..

Manski, C. and S. Lerman. (1977) "The Estimation of Choice Probabilities from Choice-Based Samples," *Econometrica*, **45**, 1977-1988.

Mantel N. and C. Brown. (1973) "A Logistic Reanalysis of Ashford and Sowden's Data on Respiratory Symptoms in British Coal Miners," *Biometrics*, **22**, 649–665.

Morimune, K. (1979) "Comparisons of Normal and Logistic Models in the Bivariate Dichotomous Analysis," *Econometrica*, **47**, 957–975.

Murphy, K.M. and R.H. Topel. (1985) "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics*, **3**, 370–377.

Nerlove, M. and S.J. Press. (1983) "Univariate and Multivariate Log-Linear and Logistic Models," Rand Corporation Working Paper Number R-1306-eda/nih.

Simon, C.P. and A.D. Witte. (1982) *Beating the System: The Underground Economy.* Boston: Auburn House.

Slemrod, J. (1994) "A General Model of the Behavioral Response to Taxation," Mimeo, University of Michigan.

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