Econometric Models for Multi-Stage Audit Processes: An Application to the IRS National Research Program*

Brian Erard
B. Erard & Associates

and

Jonathan S. Feinstein
Yale School of Management

Abstract. We develop an econometric methodology to control for errors in assessments that result from multi-stage audit processes. We then apply our methodology to data from a random sample of individual income tax audits collected under the Internal Revenue Service's National Research Program to assess the extent to which noncompliance is successfully identified on various income line items of the tax return.

*This paper is based on research performed under contract for the Internal Revenue Service. We thank Charles Bennett, Kim Bloomquist, Robert Brown, Janice Hedemann, Drew Johns, Mark Mazur, Alan Plumley, and Eric Toder for their considerable assistance with understanding the National Research Program data as well as various issues surrounding tax compliance and enforcement. Any opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the Internal Revenue Service.

I. INTRODUCTION

In this paper, we develop an econometric methodology to control for errors in assessments that result from multi-stage audit processes. We then apply our methodology to data from a random sample of individual income tax audits collected under the Internal Revenue Service's National Research Program (NRP) to assess the extent to which noncompliance is successfully identified on various income line items of the tax return.

Auditing is a standard and essential tool for assessing the validity and reliability of information and processes. Three of the most common forms of audit are financial, operational, and compliance. Financial audits are used to verify the accuracy of financial statements of governments and businesses. Operational audits are employed to assess managerial performance through an analysis of the effectiveness and efficiency of the operational structure, internal control procedures, and processes. Compliance audits are used to evaluate whether, and to what extent, policies, procedures, and other requirements for individuals, businesses, or organizations are being met. Compliance audits are frequently conducted by governments. Examples include examinations of tax returns; audits to assess compliance with regulatory policies, such as environmental regulations; and audits to evaluate whether reporting, spending, and other requirements are being met with respect to government-funded programs.

A common feature of these various forms of audit is that they normally seek only to provide reasonable assurance. Due to practical constraints, it is often infeasible to exhaustively examine every detail or aspect of an operation, system, or report. Hence, audits normally rely on sampling and testing, either at random, or in areas deemed to be of greatest risk for substantial noncompliance with reporting, procedural, or other requirements. Moreover, even when an issue or process is evaluated, there is often potential for imperfect detection of noncompliance. For example, in tax audits, examiners are not always successful in uncovering certain forms of income that have been understated. Thus, audit findings are frequently subject not only to sampling errors, but also errors in detection. In this paper, we introduce some econometric methods for controlling for such errors when analyzing the results of audits, and we apply these methods to a sample of individual income tax audit results to develop estimates of detected and undetected tax noncompliance. Our approach is based on the detection controlled methodology introduced by Feinstein (1990, 1991), which we have adapted to account for the multi-stage nature of the tax return examination process.

Typically, audit processes involve several stages, and it is important to account for impact of the decisions made during these stages on the outcome of the audit. In our tax audit application, some of the key decisions made during the audit include: which returns to audit; the type of audit to be conducted; the classification of mandatory issues to be examined; and whether additional unclassified issues should be examined. These decisions are made at different stages of the process, and by different individuals. In particular, selection of returns for audit is conducted early in the process according to a stratified random sampling design. Under this design, returns considered at higher risk of noncompliance are subject to a higher sampling rate. Once selected, a return is assigned to a seasoned examiner known as a "classifier" who assesses what type of audit should be conducted (accept as filed; correspondence audit involving only one or a few issues; or a more intensive face-to-face audit). The majority of returns in our data were subjected to a face-to-face audit, and it is these returns that are the focus of our analysis. For such returns, the classifier is responsible for selecting a set of mandatory issues to be audited. At the examination stage, the examiner has discretion to audit additional issues on the return that have not been classified.

Our work builds on an earlier model that we developed (B. Erard & Associates, 2005, 2006, 2007) to assist the IRS in estimating the aggregate tax reporting gap associated with federal individual income tax returns. The current framework extends our earlier work down to a more detailed level of analysis at the level of individual income components, focusing on estimating noncompliance associated with these income components. It is hoped that the resulting estimates from this approach will serve as key inputs for a microsimulation model of individual income tax reporting noncompliance under development by the IRS. The IRS has a sophisticated tax calculator that can be used to combine our estimates of underreporting by income component with separate estimates of noncompliance with respect to deductions, credits, and other offsets for each return, generating detailed estimates of tax noncompliance.

For our analysis, we rely on the Internal Revenue Service's National Research Program. In this important initiative, the IRS gathers data about tax noncompliance through stratified random sample of approximately $45,000$ federal individual income tax returns that have been subjected, in most cases, to rather substantial audits. The initial wave of NRP data is for tax year 2001. (For more background see Brown and Mazur, 2003.)

We develop a pair of models, each tailored to specified set of income items. The NRP uses a classification procedure as a first stage in the examination process, in which a classifier determines whether a given line item or schedule should be intensively reviewed during the audit. For certain items, including those for which income is primarily covered by information reporting (examples include wages, interest, and dividends), this classification screening stage is quite important. For such items, third party information documents often provide a very strong indication of how much should be reported on the return. In many cases the amount reported by the taxpayer for such an item is consistent with what is shown on the information documents, obviating the need to perform a detailed examination of the item. Typically, then, such an item is classified for examination only when the reported amount is inconsistent with what is shown on third party information documents, when those information documents appear to be incomplete or suspicious, or when other available information points to a potential problem with the line item. For income items for which classification is an important screening process, we develop a model that includes an equation describing this process, thus extending the detection controlled model in a new direction that reflects the multi-stage nature of NRP examinations.

In contrast, there are other income items in the NRP that are routinely classified for a careful examination. For instance, in the great majority of cases where a tax return reports income from a nonfarm or farm sole proprietorship, the relevant schedule (Schedule C or Schedule F) is classified for examination. In cases like these where classification is fairly routine, it is not productive to model the classification process. In such cases, we therefore estimate a specification for the line item that does not include a classification equation.

A challenging issue for empirical estimation is those cases in which an examiner audits an income item that was not classified for examination. In the case of income items subject to extensive information reporting, it sometimes happens that an item is not classified for examination, but the examiner nonetheless thoroughly investigates the item, sometimes uncovering significant misreporting. We suspect that this typically happens when the examiner has uncovered some trace or signal that the income item may have been misreported during the course of his audit, which drives him to explore the issue more thoroughly. To address such cases, we model the examiner's decision to audit an income item that has not been classified for examination using a simple exponential model in which his chances of examining the issue are dependent on the level of noncompliance. Our logic is that when noncompliance is present it is more likely the examiner will get a signal indicating the presence of noncompliance, triggering him to examine the item. We are in the process of developing and estimating a more structural model of this process, in which we model the examiner's decision more explicitly as a choice problem. We hope to present this model and results in a subsequent paper.

For income items not subject to extensive information reporting, classification (and hence, examination) is more or less routine when the component is reported on the return. However, when the component is not reported on the return, a detailed examination of issues surrounding the component is relatively uncommon. Typically, some probing is carried out during the course of the examination to assess whether the component should have been reported, but further investigation only occurs if the probe indicates a significant potential for unreported income. We therefore develop separate specifications for this group of income components, one for the case where the component has been reported on the return and another for the case where the component has not been reported. In the former case, we specify the examination as a single stage process. In this specification, the income component is thoroughly examined, and the examiner detects all, some, or none of the noncompliance that is present. In the latter case, where the component has not been reported on the return, we specify the examination as a two-stage process. In the first stage, an initial income probe either uncovers the presence of underreporting or it does not. In the second stage, which applies when the presence of underreporting has been uncovered, the magnitude of underreporting is assessed. Detection errors are accounted for in each stage. In the first stage, a detection error occurs if the initial probe fails to uncover underreporting when it is present. In the second stage, a detection error occurs if the examiner's assessment of the magnitude of underreporting reflects only a portion of the actual amount of noncompliance that is present.

Our specification provides a richer framework for the econometric analysis of audit and compliance systems than previous models, which often overly simplify the steps involved in selecting cases and issues for intensive examination. In particular, we believe many realworld enforcement systems grapple with the issues of different kinds of items being reported and have multiple layers of evaluation. We note that our models do not address the full range of behavioral issues that arise in these systems. Specifically, the models have not been derived from a specific game-theoretic, utility maximization framework. Nor do they account for factors such as social norms and preferences. Still, our framework incorporates a relatively simple semi-structural model of taxpayer behavior that describes the NRP classification and examination process with some care, recognizing that the process is different for different types of income sources and that it has multiple stages. We believe that this framework offers a good foundation for developing a next generation of models that incorporate both the kinds of process level detail we include and a more structural behavioral framework for taxpayer reporting.

Our preliminary estimates document considerable heterogeneity in detection rates across examiners for some income items, a finding consistent with earlier work, for example by Feinstein (1989, 1991), Alexander and Feinstein (1987), and Erard (1993, 1997). In addition, these estimates indicate that the NRP classification process is successful in flagging for examination many of the more substantial cases of underreporting with respect to various income components on the return. However, we find that in those cases where auditors choose to examine income components that have not been classified, they sometimes uncover substantial noncompliance, indicating that the classification process alone cannot always identify every issue on a return where noncompliance is present.

Over the course of our analysis, we have found that the yield on ordinary NRP classification-guided face-to-face examinations is not significantly different from the yield from the more comprehensive examinations of a comparable set of returns in the NRP calibration sample. For the calibration sample, auditors were instructed to perform a very thorough examination (more like those undertaken under the predecessor TCMP). This appears to indicate that the NRP approach of guiding audits through a process of classifying mandatory issues for examination, while still allowing examiners the opportunity to examine unclassified issues, may achieve similar results to comprehensive auditing of all issues on all returns. However, the calibration sample size was rather small (1,642 returns), so the results are merely suggestive, not conclusive. Moreover, as discussed previously, examiners are not always able to detect all noncompliance on a return, even with very thorough examinations. So, it remains important to allow for and assess the extent to which noncompliance goes undetected on NRP examinations.

As our preliminary results are still being evaluated by the IRS, we are only able to present a limited set of results that includes statistics on actual audit adjustment rates and predictions of the degree to which noncompliance has been successfully detected with respect to selected income components. We are unable to present the explanatory variables we include in our noncompliance specifications, the parameter estimates associated with those variables, or the implied magnitudes of detected and undetected noncompliance. Once our results have been carefully reviewed by the IRS, we are hopeful we will able to make a fuller set of results public, in a subsequent paper.

The remainder of our paper is organized as follows. In Section II we describe the NRP.

This is followed by a discussion of modeling considerations in Section III. In Section IV we present our model of taxpayer reporting and the NRP classification and examination processes, derive the likelihood functions we estimate, and discuss estimation issues we encountered and modifications to our base model. In Section V we present a preliminary set of empirical results, and in Section VI we conclude.

II. THE NRP

The NRP database contains a stratified random sample of approximately $45,000$ federal individual income tax returns from tax year 2001 that were subjected to special examination procedures. An important feature of the data acquisition process is that not all cases follow the same pathway for data collection. There are in particular five features of the data acquisition process that are important for analysis, which we discuss in turn.

First, returns are subject to a classification process. In this process a classifier examines the filed return and places the return into one of three categories: (i) accepted – meaning the return is accepted as is or with minor adjustments, and, importantly, there is no further contact with the taxpayer (except if the return is then selected into the calibration sample – see below); (ii) correspondence audit – meaning a correspondence will be initiated with the taxpayer regarding a circumscribed set of issues for which adjustments may be made – but there is no planned face-to-face audit; and (iii) audit – a face-to-face audit. The breakdown of cases into these 3 categories is approximately: 3,400 accepted; 2,600 selected for correspondence audit; and the balance, roughly 39,000, selected for face-to-face audit. As explained in Section I, we focus our analysis on returns in category (iii) – the vast majority of returns in the NRP sample, and the returns for which aggregate noncompliance, on a weighted basis, appears to be far and away the greatest.

Second, for returns in categories (ii) and (iii) the classifier flags a set of issues for either correspondence, in the case of returns falling in category (ii), or examination during audit, for returns falling in category (iii). Classification may be triggered by a variety of factors. For instance, a classifier will generally assign an issue for audit in cases where a third-party information document indicates the presence of income not reported by the taxpayer. As a second example, in those cases where a taxpayer reports self-employment income, the classifier normally will assign various issues on Schedule C or Schedule F to be investigated, because noncompliance is known to be prevalent on these schedules. We discuss how we model this classification process below in the Section III. footnote[1]We note that NRP classifiers had access to "case-building" information when making their decisions. This information was drawn from a variety of sources, both governmental and non-governmental. Some of these sources include third-party information documents, previous year tax returns, IRS activity with respect to the taxpayer over the preceding several years, and a credit history. This information is placed on the NRP data record for the case – though the precise use made of it by the classifier is not recorded. We do make use of the third-party information reports in building our models, but we do not make use of other case-building information. The role of this additional information in the classification and audit process is an area for future research.

Third, a subset of cases is chosen for the "calibration sample". The calibration sample includes approximately 470 returns originally assigned to be accepted as filed from the ordinary NRP sample as well as approximately 1,175 randomly selected returns that were not included in the ordinary NRP sample. Generally, returns assigned to the calibration sample receive a more thorough audit than they were initially assigned to receive. The calibration sample is a stratified random sample covering all three classification categories (returns initially classified as accepted, for a correspondence audit, and for a face-to-face examination). We are in the process of using use the calibration sample to help identify certain parameters in one of our models, but we have not yet completed this work.

Fourth, during a face-to-face audit examiners have the discretion to go beyond the issues that were flagged for examination during the classification process. Indeed, examiners frequently do probe for sources of income not reported on the return, even when these sources have not been classified. Either as a result of such probes or through other information gained during the audit, the examiner may become suspicious of potential noncompliance on an unclassified issue. In such cases, it is not uncommon for the examiner to investigate more deeply and discover significant noncompliance with respect to the issue. Importantly, the NRP data record which issues were examined, which of them were classified, and any adjustments that were made as a result of the examination.

Fifth, with respect to issues examined during a face-to-face audit, the examiner is required to record a zero when he audits an issue but finds no misreporting. This is important in making a clear distinction in the data between issues examined for which no noncompliance is found and issues not examined. By allowing for multiple intensities of interaction with and levels of audit of taxpayers, the NRP sample design deviates significantly from that of the predecessor Taxpayer Compliance Measurement Program (TCMP), which called for uniformly intensive face-to-face examinations. To properly analyze the NRP, it is therefore necessary to develop and apply new models that account for these differences in sample design and examination procedures.

We perform our empirical analysis using only those returns that were subject to face-to-face examination; we exclude those returns that were subject to a correspondence audit or accepted as filed. In restricting our attention to face-to-face examinations, we exclude approximately 6,000 returns. When weighted, these returns represent approximately 43 percent of the overall return population. Nonetheless, our analysis of the calibration sample suggests that this portion of the population is responsible for only a very small share of aggregate noncompliance in the population.

III. ISSUES FOR ESTIMATION

There are three fundamental issues that must be addressed when developing models of tax noncompliance and its detection in the NRP. These are: (i) heterogeneity in reporting behavior, particularly that there are unusually high levels of under-reporting by a small proportion of taxpayers; (ii) the failure of the examination process to completely identify all cases of noncompliance; and (iii) the NRP examination process itself, which has a specific structure.

Heterogeneity in Reporting Behavior

Our models of taxpayer reporting behavior follow our earlier work (see for example Alm, Erard, and Feinstein, 1996) in specifying reporting noncompliance using a log-normal distribution. The log-normal specification allows for a skewed distribution in which there is a "long tail" to the right of the distribution. This captures the empirical fact that there is a small portion of taxpayers with very high levels of noncompliance. In our specifications, we also account for the nontrivial percentage of taxpayers who fully and accurately report their tax liability. In some cases we break the noncompliance decision into two parts:[2] (i) a simple probit model is used to estimate the probability that an income component has been underreported, and (ii) a log-normal regression is used to estimate the magnitude of under-reporting conditional on under-reporting having occurred.

Undetected Noncompliance

Nondetection arises whenever the examiner fails to detect all noncompliance on a return. In the NRP this may happen for two distinct reasons: (i) the examiner fails to detect all noncompliance on an issue he examines; or (ii) the examiner fails to audit an issue on which noncompliance exists.

We model nondetection of the first kind using the detection-controlled methodology developed by Feinstein (1990, 1991). Other researchers have used this methodology to analyze

[2] We do not model the taxpayer's reporting decision as a utility maximization problem nor as an optimal decision under some comparable structural model. Rather, we focus structurally on the enforcement side.

tax compliance (Alexander and Feinstein, 1987; Erard, 1993, 1997) as well as a variety of applications in other domains, including regulation and health care. We employ in particular the fractional detection model developed in Feinstein (1991).

We model the second form of nondetection using a specification that extrapolates from noncompliance found on returns examined for a specific issue to likely noncompliance for the same issue on returns for which the issue was not examined. As we discuss in our results section this extrapolation is empirically challenging to do in a sensible way, in part because when an examiner chooses to examine an issue that was not classified for examination it is often because he has a lead of some kind that points to potential noncompliance. As a consequence, detected noncompliance on unclassified issues that are ultimately examined tends to be relatively high, and we cannot simply assume that the rate of noncompliance is comparable on returns for which examiners do not have a lead and do not examine the issue. We develop specific modeling strategies to address this problem.

An important issue when accounting for undetected noncompliance is how to treat cases where a return has received a negative adjustment for an income component; in other words, cases where the examiner has assessed that the taxpayer has overstated the amount of that income component on the return. While certainly not trivial, negative audit adjustments are less common and tend to be much smaller, on average, than positive adjustments. The NRP includes a set of sample weights that make the returns broadly representative of the overall return population in tax year 2001. In the NRP, approximately 7 percent of the weighted sample of returns received a negative adjustment to their reported total income amount, while approximately one-third of the weighted sample received a positive adjustment. Among returns with a negative adjustment in the weighted sample, the median reduction in total income was $135. The median increase in total income among positive adjustment cases was $917.

In our econometric analysis, we account for the possibility that examiners do not always fully uncover cases of income underreporting during their examinations. Should we also account for the possibility that cases of income overreporting sometimes go undiscovered? We do not do so in this study. Rather, we treat cases with negative adjustments the same as cases with no adjustment; namely, as instances where the examiner has assessed the report to be perfectly compliant. footnote[3]For the purposes of tax gap estimation, the IRS does net out the amount of income overreporting discovered on returns from the estimated amount of income underreporting prior to computing the change in tax liability. However, no adjustment is made to account for possible undiscovered cases of income overreporting. Our reasoning is that, while taxpayers may not have much incentive to reveal instances of underreporting to the examiner, they (and their representative, if any) do have an incentive to reveal instances of overreporting. To the extent that taxpayers and their representatives review their returns and related documentation prior to and during the audit, there is a reasonable chance that cases of overreporting will come to light and be disclosed to the NRP examiner. In addition, the NRP examiners are charged with examining returns with an eye toward finding both instances of underreporting and instances of overreporting; their role in the study is to assess the correct amount of tax, neither too much nor too little. Our working assumption is therefore that any detection errors with regard to cases of overreported income are likely to be modest and do not justify the substantial additional econometric modeling that would be required to assess them. footnote[4]Refer to Alm, Erard, and Feinstein (1996) and Erard (1997, 1999) for examples of econometric studies that control for income overreporting.

It should be noted that our measure of noncompliance for an income component is essentially the difference between the amount of income that the IRS would assess for the component if it were fully aware of all relevant information and circumstances pertaining to the 2001 tax period and the amount actually reported on the return. Various reviews and checks are built into the NRP to help ensure that examiner assessments are consistent with the IRS' interpretation of tax laws, rules, and procedures. In many instances, the correct amount to report when the underlying facts are known is relatively straightforward to determine, and virtually all reasonable individuals proficient in tax law would agree on this amount. However, there do exist some grey areas of the tax law for which the rules for reporting a given type of income are less clear cut. With respect to cases involving these areas of the law, experts might reasonably disagree on what amount should be reported. Our estimates are based on what the IRS would deem the appropriate amount to report is in such cases if it were fully aware of the relevant facts and circumstances.

It should also be noted that our analysis makes no attempt to infer the motive behind misreporting. In particular, we make no attempt to distinguish among deliberate and unintentional errors. footnote[5]Refer to Erard (2003) for a study that does attempt to distinguish deliberate from unintentional reporting violations. For all issues for which noncompliance is discovered in the NRP, the examiner is supposed to record the reason, as far as he can determine it, for the noncompliance, using a supplied list of reasons and the reason code. We do not use this information in our analysis.

IV. MODELS

In this section we present the main models we estimate, which we view as novel and important for analyzing the NRP. First we present our model of noncompliance, classification, and detection (Model 1) that is used to analyze noncompliance on income components subject to extensive information reporting, such as wages, interest, and dividends. As discussed in Section III, this specification allows us to account for the likely possibility that noncompliance with respect to one of these components tends to be larger when the component either has been classified for examination or has not been classified but nonetheless has been examined. Next we present the modified detection-controlled model (Model 2) that we use for income components that are not subject to extensive information reporting. In most cases, such income components are classified for examination whenever they are reported on the return, so it is unnecessary to model the classification decision for these components.

Model 1: Noncompliance, Classification, and Detection

Consider an income component, such as wages. If we allow N to represent the true magnitude of noncompliance with respect to this component, then, for any given return in our sample, the value of N might be zero, signifying perfect compliance, or positive ($N > 0$), signifying a positive magnitude of noncompliance. footnote[6]As discussed in Section II, for purposes of estimation, if a taxpayer overstates his income on the return, we treat this as an instance of perfect compliance ($N = 0$).

The first equation in our econometric model provides a specification of the magnitude of noncompliance:

$$ln(N^* + h) = {\beta_N}' x_N + \epsilon_N. \tag{1}$$

In this expression N^* is a latent variable describing the taxpayer's propensity to commit noncompliance on this particular income component. (Note that we do not subscript the income component for ease of notation.) The symbol h represents a displacement parameter that allows the distribution of N^* to extend below zero – h is required to be greater than or equal to zero; x_N is a set of variables associated with taxpayer noncompliance, such as filing status, β_N is a vector of coefficients, and ϵ_N is a random disturbance, discussed further below

The actual level of noncompliance N is determined as:

$$N = \begin{cases} N^* & N^* > 0; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

All returns must go through the classification screening process for the income component. We model this process as follows:

$$C^* = {\beta_C}' x_C + \epsilon_C, \tag{3}$$

In this expression C^* is a latent variable describing the propensity of the classifier to assign the income component to be examined, x_C is a set of variables associated with the classification decision, β_C is a vector of coefficients, and ϵ_C is a random disturbance. We observe the classification outcome C, where

$$C = \begin{cases} 1 & C^* > 0; \\ 0 & C^* \leq 0. \end{cases} \tag{4}$$

The outcome $C = 1$ means the income component has been classified for examination. In our empirical specification we include a dummy variable for each classifier who has worked at least 15 returns. We are therefore, able to control for differences in classification styles across classifiers.

We assume that the examiner always audits the income component if it has been classified for examination. On the other hand, if the component has not been classified, it is possible that the examiner will still elect to audit it. Typically, this will happen when the examiner learns some information over the course of his investigation that leads him to suspect non-compliance with respect to the component. We are able to model whether an unclassified income component is ultimately examined, because the NRP data sample records when a component has been examined, even when no adjustment has been made for the component. In our econometric model, we specify the probability that an unclassified income component will be examined as:

$$\frac{exp\{\alpha_0 + \alpha_1 N\}}{1 + exp\{\alpha_0 + \alpha_1 N\}}. \tag{5}$$

In other words, we allow the examination probability to depend on the magnitude of actual noncompliance. footnote[7]We note that there is a potential identification issue here; in subsequent work, we plan to incorporate the calibration sample to aid in identifying the model. The parameter α_0 in this expression determines the probability that an unclassified income component will be examined when it has been properly reported (i.e., when $N = 0$). The parameter α_1 determines the degree to which the probability of an examination changes with the magnitude of noncompliance. If α_1 is positive (the anticipated result), the probability of examination increases with the magnitude of noncompliance. In cases where the income component is neither classified nor examined on a return, the audit adjustment for the component (A) will be equal to zero. If the income component has been properly reported, this adjustment will accurately reflect that the report is fully compliant ($A = N = 0$). On the other hand, if the income component has been underreported on the return, the full amount of the understatement (N) will go undetected.

Now consider the cases where the income component either has been classified for examination, or has not been classified but has been examined anyway. In such cases, all, some, or none of the noncompliance that is present may be detected. We model the detection process using the fractional detection specification of Feinstein (1991):

$$D^* = {\beta_D}' x_D + \epsilon_D. \tag{6}$$

In this specification D^* is a latent variable describing the extent to which noncompliance is detected during the examination. The detection rate (D) represents the actual fraction of noncompliance on the income component that has been detected. It is defined as follows:

$$D = \begin{cases} 1 & D^* \geq 1 \text{ (complete detection);} \\ D^* & 0 < D^* < 1 \text{ (partial detection);} \\ 0 & D^* \leq 0 \text{ (nondetection).} \end{cases} \tag{7}$$

The term x_D in Equation (6) represents a set of explanatory variables associated with the detection, while β_D is a vector of coefficients, and ϵ_D is a random disturbance. Among the set of explanatory variables are dummy variables for individual examiners that have audited a sufficient number of returns (typically, 15 or more). This allows us to compare the relative performances of different examiners and to predict the extent to which they have been successful in uncovering any noncompliance that is present with respect to an income component.

Notice that if the income component has been reported properly ($N = 0$), the audit adjustment A will be equal to zero ($A = 0$) regardless of the effectiveness of the examination.

In this case, the recorded adjustment will properly reflect the fact that the return is fully compliant with respect to the item. On the other hand, if the income component has not been reported accurately ($N > 0$), the audit adjustment may fully reflect the magnitude of noncompliance ($A = N$; $N > 0$), only partially reflect the magnitude ($0 < A < N$), or not reflect the magnitude at all ($A = 0$; $N > 0$).

We assume that the error terms, ϵ_N and ϵ_C, in our model are bivariate normally distributed, with zero means, standard deviations of σ_N and 1, respectively, and correlation coefficient ρ. For simplicity, we assume that the error term ϵ_D is independent of ϵ_N and ϵ_C, and that it is normally distributed with mean zero and standard deviation σ_D. The assumption regarding ϵ_N implies that the propensity to commit noncompliance follows the displaced lognormal distribution. Such a distribution has a long and thin tail, consistent with empirical evidence that tax noncompliance tends to be highly skewed.

Likelihood Function for Model 1

As discussed in the previous section, there is an important problem that arises in analyzing the NRP and indeed nearly any audit data; namely, not all noncompliance is detected, and we have no direct information about noncompliance that the examiner failed to detect. Our likelihood function thus centers around not the true level of noncompliance N, but rather the audit adjustment A, which is equal to the true level of noncompliance times the detection rate:

$$A = N * D.$$

Thus the likelihood function involves the joint distribution function of the two observable variables A and C, worked out in terms of the underlying model processes and the underlying variables N, D, and C. In carrying out the transformation from the joint distribution of N, D, and C into the joint distribution of A and C, we account for the Jacobian of the transformation $J = 1/D$. footnote[8]In the course of our work on this project we discovered that Feinstein does not account for this Jacobian term when deriving the likelihood function in his 1991 paper, a lacuna in his analysis.

The likelihood function involves 5 distinct cases: (1) the income component is classified for examination, and there is no detected noncompliance: $C = 1,\ A = 0$; (2) the income component is classified for examination, and some noncompliance is detected: $C = 1,\ A > 0$; (3) the income component is not classified for examination, but the component is examined anyway with no detected noncompliance: $C = 0$, exam, $A = 0$; (4) the income component is not classified for examination, the component is examined anyway, and some noncompliance is detected: $C = 0$, exam, $A > 0$; and (5) the income component is not classified for examination, and the component is not examined: $C = 0$, no exam. We now present the likelihood function for each of the five cases.

Case 1: $C = 1,\ A = 0$

In this case, a return is classified for an examination of the income component, but the examiner does not discover any noncompliance with respect to the component. The likelihood function for this case can be computed as the difference between the marginal probability that $C^* > 0$ (income component classified) and the joint probability that $C^* > 0$, $N^* > 0$, and $D^* > 0$ (income component classified and at least some positive noncompliance detected):

$$L = \Phi\left(\beta_C{}'x_C\right) - BN\left(\frac{\beta_N{}'x_N - ln(h)}{\sigma_N}, \beta_C{}'x_C, \rho\right)\Phi\left(\frac{\beta_D{}'x_D}{\sigma_D}\right),$$

where $\Phi(z)$ represents the standard normal cumulative distribution function (c.d.f.) evaluated at z and $BN(z_1, z_2, \rho)$ represents the standard bivariate normal c.d.f. evaluated at z_1 and z_2 for correlation coefficient ρ.

Case 2: $C = 1,\ A > 0$

In this case, a return is classified for an examination of the income component, and some positive amount of noncompliance is detected with respect to the component. The likelihood function for this case accounts for the possibilities that noncompliance is either fully or partially detected:

$$
\begin{aligned}
L = & \frac{1}{\sigma_N(A+h)}\phi\left(\frac{ln(A+h)-\beta_N{}'x_N}{\sigma_N}\right)\Phi\left(\frac{\beta_C{}'x_C+\rho\left(\frac{ln(A+h)-\beta_N{}'x_N}{\sigma_N}\right)}{\sqrt{1-\rho^2}}\right) \\
& \Phi\left(\frac{\beta_D{}'x_D-1}{\sigma_D}\right)+\int_0^1\left[\frac{1}{\sigma_N\sigma_D(A+hD)}\phi\left(\frac{ln(A/D+h)-\beta_N{}'x_N}{\sigma_N}\right)\right. \\
& \left.\Phi\left(\frac{\beta_C{}'x_C+\rho\left(\frac{ln(A/D+h)-\beta_N{}'x_N}{\sigma_N}\right)}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{D-\beta_D{}'x_D}{\sigma_D}\right)\right]dD,
\end{aligned}
$$

where $\phi(z)$ represents the standard normal p.d.f. evaluated at z. The next three cases involve returns for which the income component has not been classified for examination. As discussed previously, when an income component is not classified for examination, we assume that the examiner audits the income component anyway with a probability that depends on the true level of noncompliance N:

$$
\frac{exp\{\alpha_0+\alpha_1 N\}}{1+exp\{\alpha_0+\alpha_1 N\}}.
$$

The parameter α_0 in the above expression determines the probability that an unclassified return will be examined when the income component is properly reported ($N=0$). The parameter α_1 determines the degree to which the probability of an examination changes with the magnitude of noncompliance. If α_1 is positive (the anticipated result), the probability of examination increases with the magnitude of noncompliance.

Case 3: $C = 0$, exam, $A = 0$

In this case, the income component has not been classified, but the examiner elects to audit it and does not detect any noncompliance. The likelihood function for this case is computed as the sum of two joint probabilities. The first is that the income component is not classified, that it is examined, and that it is noncompliant, but that the noncompliance has gone completely undetected. The second joint probability is that the income component is not classified, that it is examined, and that it is perfectly compliant. Specifically,

$$
\begin{aligned}
L = \Phi\left(\frac{-\beta_D' x_D}{\sigma_D}\right) \int_0^\infty \Bigg[& \frac{1}{\sigma_N (N+h)} \phi\left(\frac{ln(N+h) - \beta_N' x_N}{\sigma_N}\right) \\
& \Phi\left(\frac{-\beta_C' x_C - \rho\left(\frac{ln(N+h)-\beta_N' x_N}{\sigma_N}\right)}{\sqrt{1-\rho^2}}\right) \frac{exp\{\alpha_0 + \alpha_1 N\}}{1 + exp\{\alpha_0 + \alpha_1 N\}} \Bigg] dN \\
& + BN\left(\frac{ln(h) - \beta_N' x_N}{\sigma_N}, -\beta_C' x_C, \rho\right)\left(\frac{exp\{\alpha_0\}}{1 + exp\{\alpha_0\}}\right).
\end{aligned}
$$

Case 4: $C = 0$, exam, $A > 0$

As with Case 3, we observe this case only when the income component is not assigned by the classifier, but the examiner elects to review the component anyway. In this case, however, the review of the income component uncovers some noncompliance. The likelihood function for this case is somewhat similar to that given earlier for Case 2, which involves detected noncompliance for a classified return:

$$
L = \frac{1}{\sigma_N(A+h)}\phi\left(\frac{ln(A+h)-\beta_N{}'x_N}{\sigma_N}\right)\Phi\left(\frac{-\beta_C{}'x_C-\rho\left(\frac{ln(A+h)-\beta_N{}'x_N}{\sigma_N}\right)}{\sqrt{1-\rho^2}}\right)
$$
$$
\Phi\left(\frac{\beta_D{}'x_D-1}{\sigma_D}\right)\left(\frac{exp\{\alpha_0+\alpha_1 A\}}{1+exp\{\alpha_0+\alpha_1 A\}}\right)
$$
$$
+\int_0^1\left[\frac{1}{\sigma_N\sigma_D(A+hD)}\phi\left(\frac{ln(A/D+h)-\beta_N{}'x_N}{\sigma_N}\right)\right.
$$
$$
\Phi\left(\frac{-\beta_C{}'x_C-\rho\left(\frac{ln(A/D+h)-\beta_N{}'x_N}{\sigma_N}\right)}{\sqrt{1-\rho^2}}\right)\phi\left(\frac{D-\beta_D{}'x_D}{\sigma_D}\right)
$$
$$
\left.\left(\frac{exp\{\alpha_0+\alpha_1(A/D)\}}{1+exp\{\alpha_0+\alpha_1(A/D)\}}\right)\right]dD.
$$

<u>Case 5: $C = 0$, no exam</u>

In this case, the income component is not classified, and the examiner elects not to examine the return. The likelihood function for this case takes the form:

$$
L = \int_0^\infty\left[\frac{1}{\sigma_N(N+h)}\phi\left(\frac{ln(N+h)-\beta_N'x_N}{\sigma_N}\right)\Phi\left(\frac{-\beta_C'x_C-\rho\left(\frac{ln(N+h)-\beta_N'x_N}{\sigma_N}\right)}{\sqrt{1-\rho^2}}\right)\right.
$$
$$
\left.\left(\frac{1}{1+exp\{\alpha_0+\alpha_1 N\}}\right)\right]dN + BN\left(\frac{ln(d)-\beta_N'x_N}{\sigma_N}, -\beta_C'x_C, \rho\right)\left(\frac{1}{1+exp\{\alpha_0\}}\right).
$$

Model 2: Noncompliance and Detection

For many of the income components in our analysis, an examination of the component usually takes place whenever a return has reported a nonzero amount for the component. For such income components, we have worked with a simpler model that ignores the decision whether to classify the component for examination. In this model, we focus exclusively on whether noncompliance is present and the extent to which the examiner has been successful in detecting it.

There is a further issue for these types of income components. In particular, while most returns that report a nonzero amount of the component are subject to detailed examination of that component, in most cases for returns that report a zero amount for the component the component is not examined, at least not with the same intensity. Our model therefore employs separate specifications for returns that do and do not report a nonzero amount for each income component.

For returns that report a nonzero amount for an income component, the specification has two equations:

$$ln(N^* + h) = {\beta_N}' x_N + \epsilon_N \tag{8}$$

$$D^* = {\beta_D}' x_D + \epsilon_D, \tag{9}$$

where the observed level of noncompliance N is related to the latent variable N^* as follows:

$$N = \begin{cases} N^* & N^* > 0; \\ 0 & \text{otherwise.} \end{cases} \tag{10}$$

Similarly, the observed detection rate D is related to the latent variable D^* according to:

$$D = \begin{cases} 1 & D^* \geq 1 \text{ (complete detection);} \\ D^* & 0 < D^* < 1 \text{ (partial detection);} \\ 0 & D^* \leq 0 \text{ (nondetection).} \end{cases} \tag{11}$$

The two parts of this specification are identical with the corresponding parts for model 1; in particular the specification for noncompliance N^* and N are identical to equations (1) and (2), and the specification of the detection process is identical to equations (6) and (7). We maintain the assumptions that ϵ_N is normally distributed with mean zero and standard deviation σ_N; ϵ_D is normally distributed with mean zero and standard deviation σ_D; and that ϵ_N and ϵ_D are independently distributed.

As before, we work out the likelihood function in terms of assessed noncompliance, $A = N * D$. The likelihood function has two separate cases: $A = 0$ and $A > 0$. We consider each of these cases in turn.

<u>Case 1: $A = 0$</u>

In this case, either the taxpayer is compliant or is noncompliant but no noncompliance is detected. The likelihood function may be computed as one minus the probability that noncompliance is present and is at least partially detected:

$$L = 1 - \Phi\left(\frac{{\beta_N}' x_N - ln(h)}{\sigma_N}\right) \Phi\left(\frac{{\beta_D}' x_D}{\sigma_D}\right).$$

Case 2: $A > 0$

In this case, the taxpayer is noncompliant and the noncompliance is either fully or partially detected. Therefore, the likelihood function allows for detection rates ranging from zero to one:

$$L = \frac{1}{\sigma_N(A+h)}\,\phi\left(\frac{ln(A+h) - {\beta_N}' x_N}{\sigma_N}\right)\Phi\left(\frac{{\beta_D}' x_D - 1}{\sigma_D}\right) + \int_0^1 \frac{1}{\sigma_N \sigma_D (A+hD)}\,\phi\left(\frac{D - {\beta_D}' x_D}{\sigma_D}\right)\phi\left(\frac{ln(A/D+h) - {\beta_N}' x_N}{\sigma_N}\right) dD.$$

For returns that report a zero amount for the relevant income component, our specification separately addresses the likelihood that the income component is in fact present and the magnitude of that component conditional on it being present. Our specification allows for detection errors both with respect to identifying whether the income component is present and with respect to assessing its magnitude when present. We develop a specification for the joint likelihood that the income component is present and the chance that it will be detected if present:

$$P^* = {\beta_P}' x_P + \epsilon_P \tag{12}$$

$$D_P^* = {\beta_{DP}}' x_{DP} + \epsilon_{DP}, \tag{13}$$

where P^* is a latent variable describing the likelihood that some of the income component is present and D_P^* is a latent variable describing the propensity of the examiner to detect its presence. Unreported income is present if and only if $P^* > 0$. Likewise, this income is detected if and only if $D_P^* > 0$. We assume that ϵ_P and ϵ_{DP} are each normally distributed with zero means and unit standard deviations. For convenience, we also assume that they are independently distributed. The likelihood function for this portion of our model depends on whether the examiner has assessed that some of the income component is present.

Case 1: Examiner assesses that income component is present

In order for the examiner to assess that at least some of the income component is in fact present, it must be the case that both $P^* > 0$ and $D_P^* > 0$. Therefore, the likelihood function for this case is specified as:

$$L = \Phi(\beta'_P x_P)\Phi(\beta'_{DP} x_{DP}).$$

Case 2: Examiner assesses that income component is not present

If the examiner has assessed that the income component is not present, either $P^* < 0$ (component really is not present) or $D_P* < 0$ (detection error). The likelihood of this can be expressed as one minus the probability that $P^* > 0$ and $D_P^* > 0$:[9]

$$L = 1 - \Phi(\beta'_P x_P)\Phi(\beta'_{DP} x_{DP}).$$

So far, our model accounts for whether the income component is assessed to be present, but it does not account for the magnitude of the adjustment when the component is deemed to

[9] To ensure identification of this portion of our model, it is desirable that x_P includes at least one continuous variable that is excluded from x_{DP}.

be present. For returns with a positive adjustment for the income component, we assume that the magnitude of the adjustment depends on both the actual amount of the income component that is present and the extent to which it has been detected. More specifically, our specification includes the following two equations:

$$ln(N) = {\beta_N}' x_N + \epsilon_N \tag{14}$$

$$D^* = {\beta_D}' x_D + \epsilon_D, \tag{15}$$

where N represents the true magnitude of noncompliance (i.e., the magnitude of the income component that is present but which has been reported as zero), and D^* represents a latent variable for the propensity for noncompliance to be detected. We assume that ϵ_N is normally distributed with mean zero and standard deviation σ_N. Likewise, we assume that ϵ_D is normally distributed with mean zero and standard deviation σ_D. Since this portion of model is estimated over returns that are assessed to have at least some of the income component, it must be the case that detection is either partial or complete. The distribution of D^* is therefore truncated to lie above zero. The detection rate D is defined as:

$$D = \begin{cases} 1 & D^* \geq 1 \text{ (complete detection);} \\ D^* & 0 < D^* < 1 \text{ (partial detection).} \end{cases} \tag{16}$$

Unfortunately, there are relatively few examiners who have audited a sufficiently large number of returns (15 or more) that reported a zero amount of an income component of interest and were found to have a nonzero value for the component. In our analysis, we therefore apply the estimated parameters of the detection equation from our analysis of returns that reported a positive amount for the component. In effect, we are assuming that, once an examiner finds that the income component is present on a return that reports none of the component, his ability to detect the magnitude of noncompliance on that return is comparable to his ability to uncover noncompliance on a return that reports a nonzero amount for the income component.

As in our previous models, the observed assessed level of noncompliance is $A = D * N$. The likelihood function is:

$$L = \frac{1}{\Phi(\beta_D x_D)} \left[\frac{1}{\sigma_N A} \phi \left(\frac{ln(A) - \beta_N{}' x_N}{\sigma_N} \right) \Phi \left(\frac{\beta_D{}' x_D - 1}{\sigma_D} \right) \right.$$
$$\left. + \int_0^1 \frac{1}{\sigma_N \sigma_D A} \phi \left(\frac{D - \beta_D{}' x_D}{\sigma_D} \right) \phi \left(\frac{ln(A/D) - \beta_N{}' x_N}{\sigma_N} \right) dD \right].$$

V. EMPIRICAL RESULTS

In this section we present some data statistics and some of the results from our analysis. As noted in the introduction to this paper, due to issues of confidentiality and the need for the IRS to fully review and approve our results for release before making them publicly available, we are unable to report all of our results in this paper.

Tables 1 and 2 present statistics for income components estimated using model 1, which applies to components subject to substantial third party information reporting. We estimate model 1 for the following components:

(1) Wages.

(2) Taxable interest.

(3) Taxable state and local tax refunds.

(4) Dividends.

(5) Taxable pensions and IRA distributions.

(6) Gross social security benefits.

(7) Unemployment insurance.[10]

Table 1 presents information about the number of returns in our sample reporting nonzero amounts for each of these items, and the number reporting zero. The raw number of returns, rather than the population-weighted numbers are provided to give the reader a sense of the sample sizes that are available for estimation for each income component. As expected the majority of households report some wages; the majority also report some interest. Significant numbers report nonzero amounts for each of the other items. The table also presents the percentage of returns reporting nonzero amounts for which the item is classified for examination, and the percentage of returns reporting zero for which the item is classified for examination. These percentages have been weighted to give the reader a sense of the population characteristics. The majority of returns are not classified for examination

[10] There were too few cases involving alimony receipts to include in the analysis.

of a given income component. However, a modest percentage of returns reporting nonzero amounts for a given component are classified, the highest percentages being for interest, dividends, and social security benefits. For returns reporting zero amounts for a given income component, the percentage classified for exam in small; the one exception is interest, for which a nontrivial percentage of returns reporting zero interest income are classified for examination. The last two columns of the table show the number of returns examined for the specified item, both returns that were classified for examination and then examined and also returns that were not classified for examination but were examined anyway. While most examinations are for returns classified for exam, a nontrivial number of returns are examined for an item which was not classified.

Table 2 presents information about adjustments for noncompliance made during examinations for these same items. These numbers are weighted to reflect the filing population. The most striking feature of the table is the cells for which the percentage of cases having a positive adjustment is high. These include certain cells for items for which the household reported zero but the classifier assigned the item for examination – notably wages, interest, dividends, state and local tax refunds, and unemployment benefits. They also include certain cells for items that were not classified for examination but were examined anyway – notably interest, dividends, and state and local tax refunds. In the first group of cases the classifier presumably encountered a signal suggesting that there might be income present, and in the second group the examiner presumably encountered such a signal, even though the classifier had not. It is also interesting to note that for other items these signals are apparently less definitive – the adjustment rates are modest, for example, for social security benefits in cases in which the household reported zero and the classifier assigned the item for examination and cases in which this item was not classified for examination but the examiner audited it anyway. Assessment rates for items classified for examination are reasonable but not extraordinarily high, suggesting the classifiers use a fairly low threshold in triggering the decision to classify an item for examination.

Tables 3 and 4 present statistics for income components estimated using model 2. We estimate model 2 for the following issues/schedules:

(1) Net nonfarm sole proprietor income (Schedule C).

(2) Net farm sole proprietor income (Schedule F).

(3) Short-term capital gains.

(4) Long-term capital gains.

(5) Net rental and royalty income.

(6) Net partnership and S-corporation income.

(7) Other Schedule E income (such as estate and trust income).

(8) Supplemental gains reported on Form 4767.

(9) Other Form 1040 income.

Among the nine income components estimated based on this model, only two (net farm and nonfarm sole proprietor income) have a reasonable number of examiners who each audited that component on a sufficiently large number of returns (15 or more) to derive adequate estimates of the variation in detection rates across examiners. For the remaining seven components, we therefore estimated our model for each component jointly, restricting the parameters of the detection equation (with the exception of the constant term) to be common across all components.

Table 3 presents results for items (3)-(9) of the list, which are generally subject to partial third party information reporting. Columns 1 and 3 show the number of returns that report each of these components, and the number that do not. We again note that these numbers are raw numbers from the NRP and are not weighted to reflect the U.S. filer population. As such, they provide the reader a sense of the sample sizes we have to work with for the various income components. The table also shows for each item the percentage of returns among those that report the item for which there is a positive adjustment during the NRP examination process, and the percentage of returns among those that do not report the item for which there is a positive adjustment during the NRP examination process. These percentages have been weighted to reflect the U.S. filer population. As expected, a higher percentage of returns reporting the item have a positive adjustment than of returns not reporting the item. The highest adjustment rate is for households reporting rents and royalties income.

Table 4 presents similar information for Schedule C and Schedule F. Here we find that the percentage of returns reporting these items for which there is a positive adjustment during the NRP examination process is rather high, while the percentage of returns not reporting the items for which there is a positive adjustment is quite small.

We note that these tables do not present any information about the magnitude of noncompliance, only the rate. We cannot at present provide information about magnitudes, but hopefully some of this information will be presented in subsequent work.

Figures 1-3 present histograms based on our preliminary estimates of detection rates across examiners for three representative income items: wages, net rents and royalties, and Schedule C net income. For wages, we report results for all returns for which wages were examined, regardless of whether the taxpayer reported a nonzero or zero amount of wages on his return. In the cases of the latter two income components, we report results only for taxpayers who reported a nonzero amount of the component, for which examinations tended to be more thorough.

The histograms illustrate the distribution of detection rates among examiners who have audited the relevant income component on at least 15 different income tax returns. For each of the three components, there is a subset of examiners who are estimated as "near-perfect" detectors – meaning the econometric model has assigned them a detection rate of close to 100%. Essentially, these examiners serve as the benchmark against which the other examiners' detection rates are calibrated. The heterogeneity among detection rates is lowest for wages, with a very large subsample of examiners with estimated detection rates exceeding 90%. For both rents and royalties and Schedule C there is substantial heterogeneity, with the majority of examiners estimated as having detection rates below 50%, and a substantial number below 30%. These results are similar to the findings of our earlier analysis at a more aggregate level, but slightly more extreme. It would be important to explore whether the examiners estimated as perfect are more experienced, and also to check on the allocation of cases across examiners, as this might partly explain the substantial differences in detection rates.

In contrast to our results for examiner detection rates, we have found much less variation in the rate at which a given line item is classified for examination. To some extent, this may reflect common guidelines followed by classifiers for some classification decisions. However, the results might also reflect the fact that the classifiers were generally quite experienced, and therefore may have had similar work patterns.

VI. CONCLUSION

In this paper we have developed an econometric methodology to control for errors in assessments that result from multi-stage audit processes. We have applied our methodology to examine classification, examination, and noncompliance detection on a sample of federal individual income tax returns subjected to audits under the IRS National Research Program. Our models focus on noncompliance with respect to individual income components, the first time detection controlled estimation has been applied at this level of disaggregation.

Overall, we find that we are able to estimate the models successfully, and our preliminary results are broadly comparable to those obtained in previous analyses that have typically relied on more aggregated data. In particular, our initial estimates of noncompliance are broadly consistent with those found in many earlier studies. Our estimates of detection rates indicate substantial heterogeneity across examiners in detection rates for certain income items. We view the results shown here as preliminary and requiring further exploration.

References

Alexander, C. and Feinstein, J.S. (1987) 'A Microeconometric Analysis of Income Tax Evasion and Its Detection', mimeo, Massachusetts Institute of Technology.

Alm, J., Erard, B. and Feinstein, J.S. (1996) 'The Relationship Between State and Federal Tax Audits', in M. Feldstein and J.M. Poterba (eds.) *Empirical Foundations of Household Taxation*, Chicago: University of Chicago Press.

B. Erard & Associates (2005) 'IRS Tax Gap Estimation: Preliminary Results of Detection Controlled Analysis', PowerPoint presentation to Internal Revenue Service Office of Research, November 1.

______ (2006) 'Preliminary Econometric Results', results summary report submitted to Internal Revenue Service Office of Research, January 27.

______ (2007). 'Detection Controlled Estimation of Tax Noncompliance', PowerPoint presentation to Internal Revenue Service Office of Research and U.S. Treasury Office of Tax Analysis, February 5.

Brown, R.E. and Mazur, M.J. (2003) 'IRS's Comprehensive Approach to Compliance Measurement', *National Tax Journal*, 41(3):689–700.

Erard, B. (1993) 'Taxation with Representation : An Analysis of the Role of Tax Practitioners in Tax Compliance', *Journal of Public Economics*, 52(2): 163–197.

______ (1997) 'Self-selection with Measurement Errors: A Micro-econometric Analysis of the Decision to Seek Tax Assistance and its Implications for Tax Compliance', *Journal of Econometrics*, 81(2): 357–393.

______ (1999) 'Estate Tax Underreporting Gap Study', report prepared for the Internal Revenue Service Research Division, Order No. TIRNO-98-P-00406, March 4.

Feinstein, J.S. (1990) 'Detection Controlled Estimation', *Journal of Law & Economics*, Volume 33(1), pp. 233-276.

______ (1991) 'An Econometric Analysis of Income Tax Evasion and its Detection', *RAND Journal of Economics*,22: 14–35.

______ (1989) 'The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections, and Abnormal Occurrences', *Journal of Political Economy*, 109: 360–369.

Internal Revenue Service (2001) *National Research Program: Program Prospectus.*

______ (2003) *National Research Program: Individual Selection Summary Report*, February.

Note: All tables refer exclusively to returns subject to face-to-face examinations

Table 1: Weighted Classification Rates by Whether Nonzero Amount Reported, Group 1 Income Components

Income Component	**Returns with a Nonzero Report for Income Component**		**Returns with a Zero Report for Income Component**		**Raw # Total Examined Returns**	
	Raw # returns	**Weighted % that Were Classified**	**Raw # returns**	**Weighted % that Were Classified**	**Classified**	**Not Classified, but Examined**
Wages	26,418	12.1	11,452	1.8	2,507	934
Interest	27,937	27.6	9,933	16.3	11,123	688
Dividends	16,692	33.2	21,178	4.5	6,875	397
State & local tax refunds	10,190	11.8	27,680	5.4	2,738	198
Pensions and IRAs	8,076	21.1	29,794	5.4	2,850	217
Gross social sec benefits	3,989	36.5	33,881	1.3	1,952	181
Unemployment benefits	1,692	9.6	36,178	0.97	272	30

Table 2: Weighted Percentage of Examined Returns with a Positive Adjustment by Classification and Reporting Status, Group 1 Income Components

Income Component	**Income Component Classified**		**Income Component Not Classified, but Examined Anyhow**
	Nonzero report for income component	**Zero report for income component**	
Wages*	33.3	73.7	24.5
Interest	26.8	92.9	68.0
Dividends	27.3	82.5	67.9
State & local tax refunds	25.0	66.8	65.9
Pensions and IRAs	19.5	48.3	33.8
Gross social sec benefits	13.5	14.3	16.9
Unemployment benefits	17.9	82.6	44.9

*Wages variable excludes tip income.

Table 3: Reporting and Examiner Adjustment Statistics, Group 2 Income Components

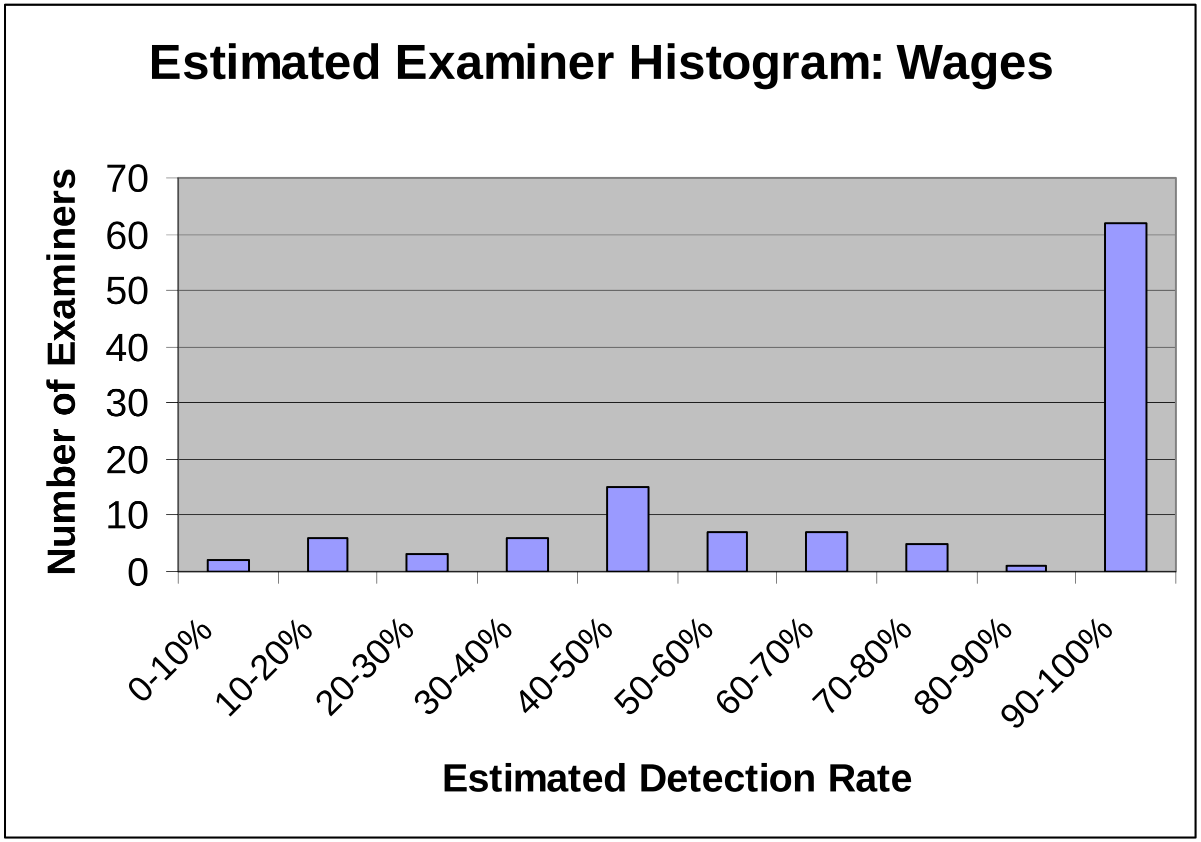
Income Component	**Income Component Reported**		**Income Component Not Reported**	
	Raw # returns	**Weighted % of returns with a positive adjustment**	**Raw # returns**	**Weighted % of returns with a positive adjustment**
Sched. D short term gains	7,981	11.1	29,889	1.3
Sched. D long term gains	13,571	14.7	24,299	2.8
Net rents and royalties	7,400	43.1	30,470	0.54
Net income from partnerships and S-corps	6,339	13.2	31,531	0.11
Other Sched. E income	1,004	14.6	36,866	0.03
Form 4797 gains	2,945	16.8	34,925	0.25
Other income	4,848	10.1	33,022	3.3

Table 4: Reporting and Examiner Adjustment Statistics, Group 3 Income Components*

Type of Sole Proprietor Schedule	Schedule Filed		No Schedule Filed	
	Raw # schedules	**Weighted % of schedules with a positive adjustment**	**Raw # returns**	**Weighted % of returns with a positive adjustment**
Non-farm (Schedule C)	23,943	55.4	17,557	1.7
Farm (Schedule F)	4,830	56.5	33,204	0.01

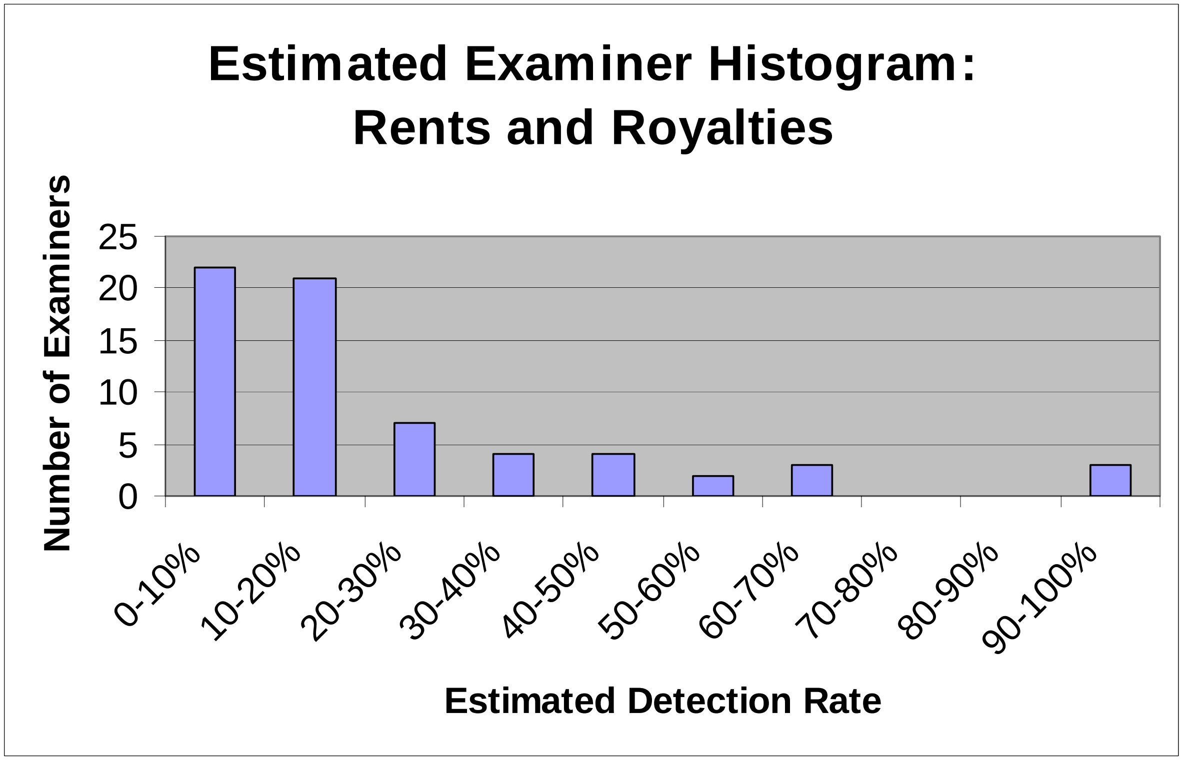
* Some taxpayers file multiple Schedule C or Schedule F returns; each schedule is counted separately in our statistics and our econometric analysis.

Figure 1



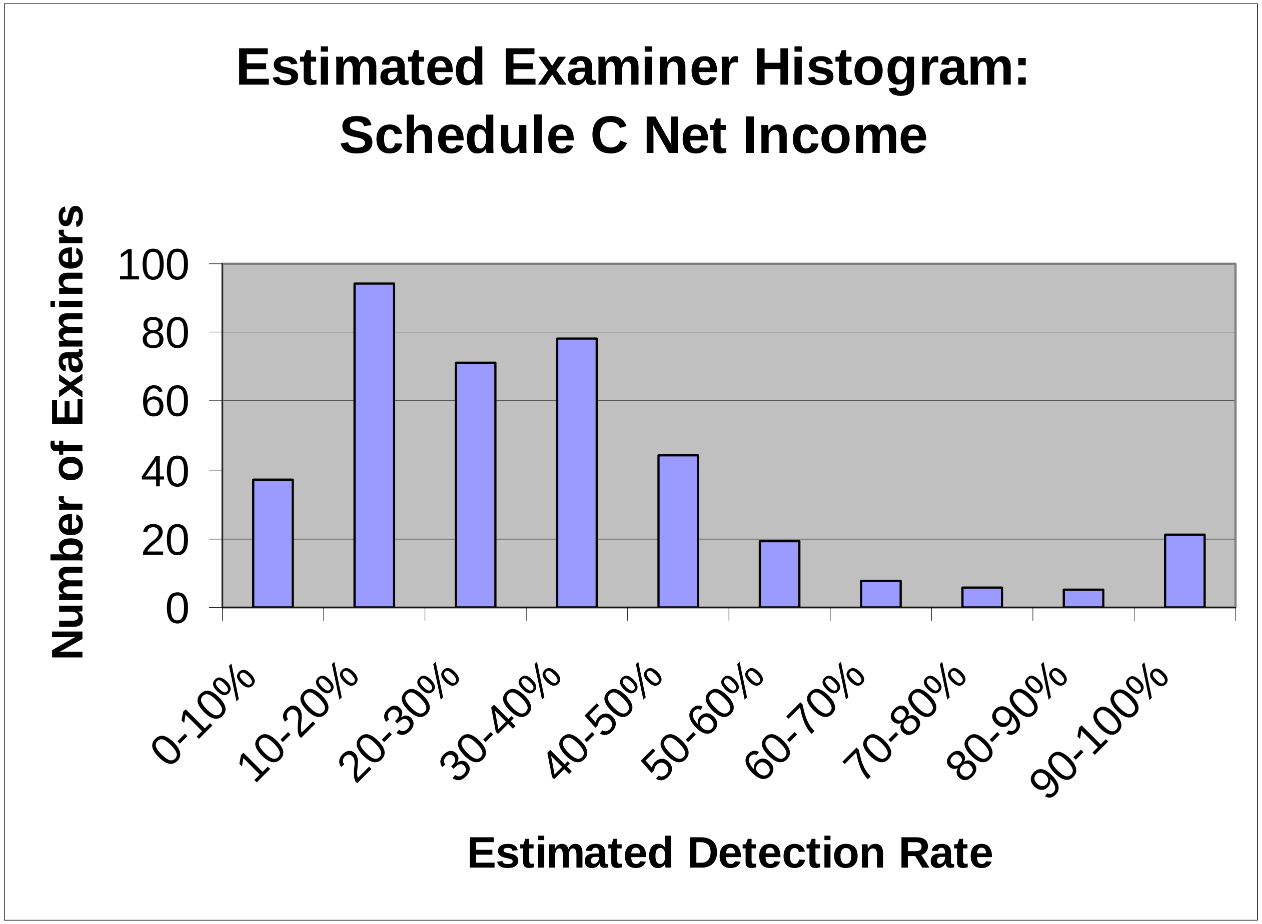
Average Detection Rate for All Examiners: 88%

Figure 2



Average Detection Rate for All Examiners: 43%

Figure 3



Average Detection Rate for All Examiners: 32%