The Individual Income Reporting Gap: What We See and What We Don't

Brian Erard, B. Erard & Associates and Jonathan Feinstein, Yale University

1. Introduction

Tax agencies are continually making decisions about the allocation of their resources across activities to promote tax compliance and combat evasion. The quality of these decisions is limited by their capacity to measure the overall level of compliance with taxpayer filing, reporting, and payment obligations, the frequency with which various types of transactions are misreported, and the characteristics of those who are responsible. In this paper, we provide an overview of how the IRS attempts to measure the degree to which filers of federal individual income tax returns properly report their incomes from various sources using data from the National Research Program (NRP).

The 2001 NRP provides a direct and nationally representative assessment of how much noncompliance IRS auditors are able to identify on individual income tax returns. However, willful tax evaders often undertake considerable efforts to conceal their misreporting, and NRP examiners are not always successful in uncovering this activity. The IRS therefore attempts to estimate not only the portion of the tax gap that we see from the NRP audit results, but also the portion that we don't.

Detection Controlled Estimation (DCE) is a statistical methodology that was initially developed by Feinstein (1990, 1991) to account for imperfections in examination processes (such as audits) to fully uncover violations (such as tax noncompliance). Under this methodology, one jointly models the detection process along with the underlying violation of interest. Under contract with the IRS, we have refined and generalized this methodology for application with National Research Program data to develop estimates of detected and undetected income underreporting for use in tax gap estimation. A key feature of the approach is that it accounts for differences among examiners in their ability to uncover noncompliance on tax returns. Intuitively, the methodology permits one to scale up the audit findings of less successful examiners to represent something closer to what the most successful examiners would have uncovered had they audited the returns.

A previous version of our methodology was employed in the development of the IRS estimates of the Tax Year 2001 individual income tax underreporting gap. Under that version, separate multiplier estimates were produced for "low visibility" and "high visibility" sources of income for each of two return categories ("business" and "nonbusiness"). Each multiplier represented an estimate of the ratio of the actual amount of underreporting present within that income source and return category to the amount that was detected during the NRP examinations. More recently, we have extended the methodology to produce more disaggregated estimates of detected and undetected underreporting by income line item in support of future tax gap estimates.

The DCE approach represents a significant departure from the earlier methodology employed by the IRS in developing its estimates of the tax gap. Under this earlier approach, an *ad hoc* adjustment was made to the portion of noncompliance on TCMP examinations that was identified by examiners without the aid of third-party information documents. While a common adjustment factor was applied to a wide range of income items, the value of this factor (3.28—meaning that there was an estimated $228 in undetected unreported income for every $100 of underreporting that was detected by examiners without the aid of third-party documents) had been derived based on findings from retrospective analysis of the random audits conducted under the Taxpayer Compliance Measurement Program (TCMP) for Tax Year 1976, the last year in which auditors in that program did not have the taxpayer's information documents available during the audit.

2. What We See

The NRP results provide details of what the taxpayer reported on each line item of the tax return as well as the NRP examiner's conclusion as to how much should have been reported for each line item. Since the returns were randomly selected, the results provide an indication of how much additional income (and tax) would have been detected if all federal individual income tax returns in the 2001 tax year population had been examined under the NRP process. They also provide an indication of what specific income and deduction items are commonly associated with compliance problems.
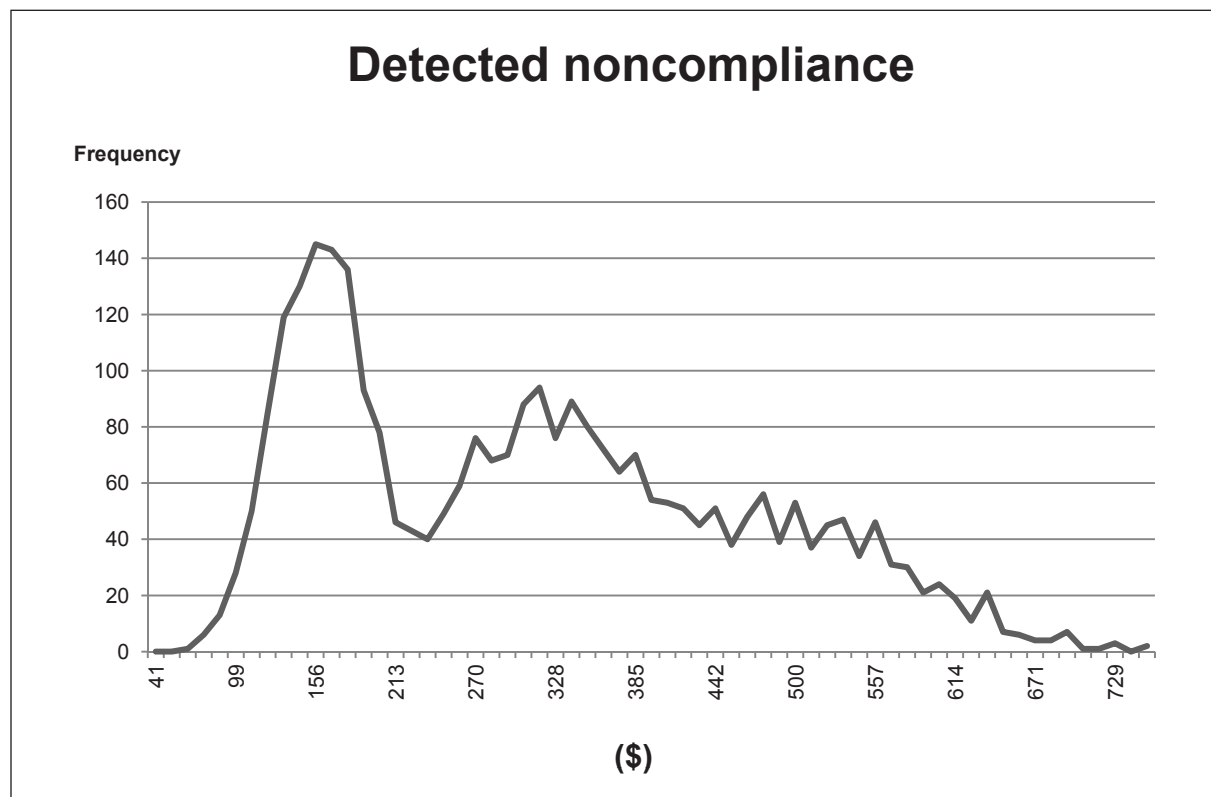
3. What We Don't See

In many instances taxpayers undertake considerable effort to conceal their tax transgressions from the tax authority. In such cases, it can be difficult for examiners to fully uncover all misreporting that is present. In general, one would expect that audit adjustments would allow us to observe many of the unintentional errors that taxpayers make in reporting their taxes, but only a portion of the deliberate cheating. Therefore, the raw NRP examination results are likely to provide an incomplete picture of the compliance landscape.

3.1 How to Measure What We Can't See

Intuitively, examiners will tend to vary in their experience and their skill at uncovering noncompliance. Some examiners may be globally superior at uncovering noncompliance with respect to all return issues; others may have a comparative advantage at uncovering noncompliance on particular issues. If we knew the relative abilities of different examiners to uncover noncompliance with respect to a particular tax issue or line item, we could "scale up" what was detected by a given examiner to approximate what the best examiner would have found in the audit.

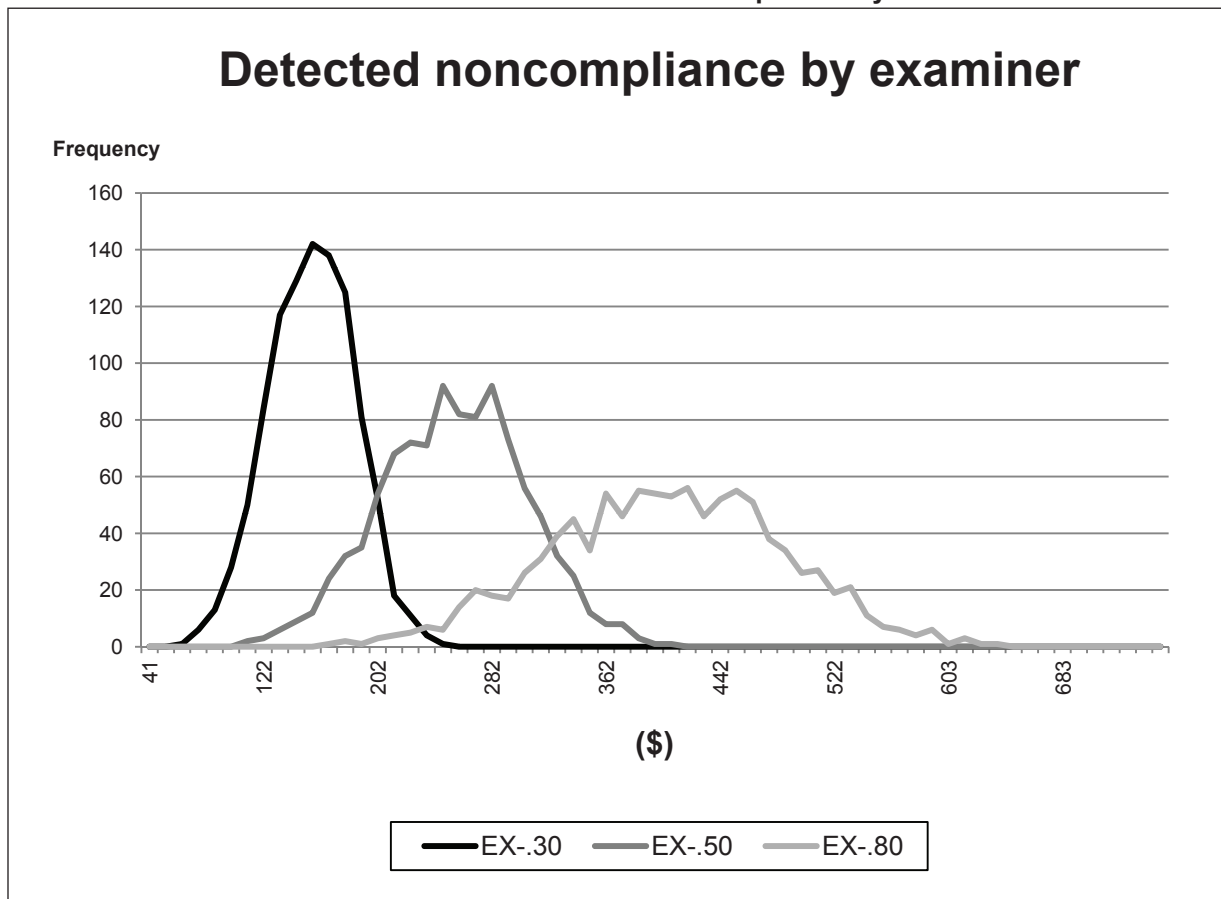**FIGURE 1. Illustration of Distribution of Detected Noncompliance**

To gain a sense of how the DCE approach works, it is helpful to consider the following scenario. Suppose that we are shown the following plot of the distribution of detected underreporting with respect to a given income source based on a random audit study:

Based on our discussion so far, we recognize that the actual distribution of noncompliance may differ from the above detected distribution, but how can we account for noncompliance that has gone undetected during the study?

Imagine that you were told that three different examiners had been randomly assigned to audit a share of the returns included in this study and that it was possible to identify the detected amounts of noncompliance that were attributable to each examiner. Suppose that a more detailed plot that illustrates the distribution of noncompliance detected by each examiner looks as follows:
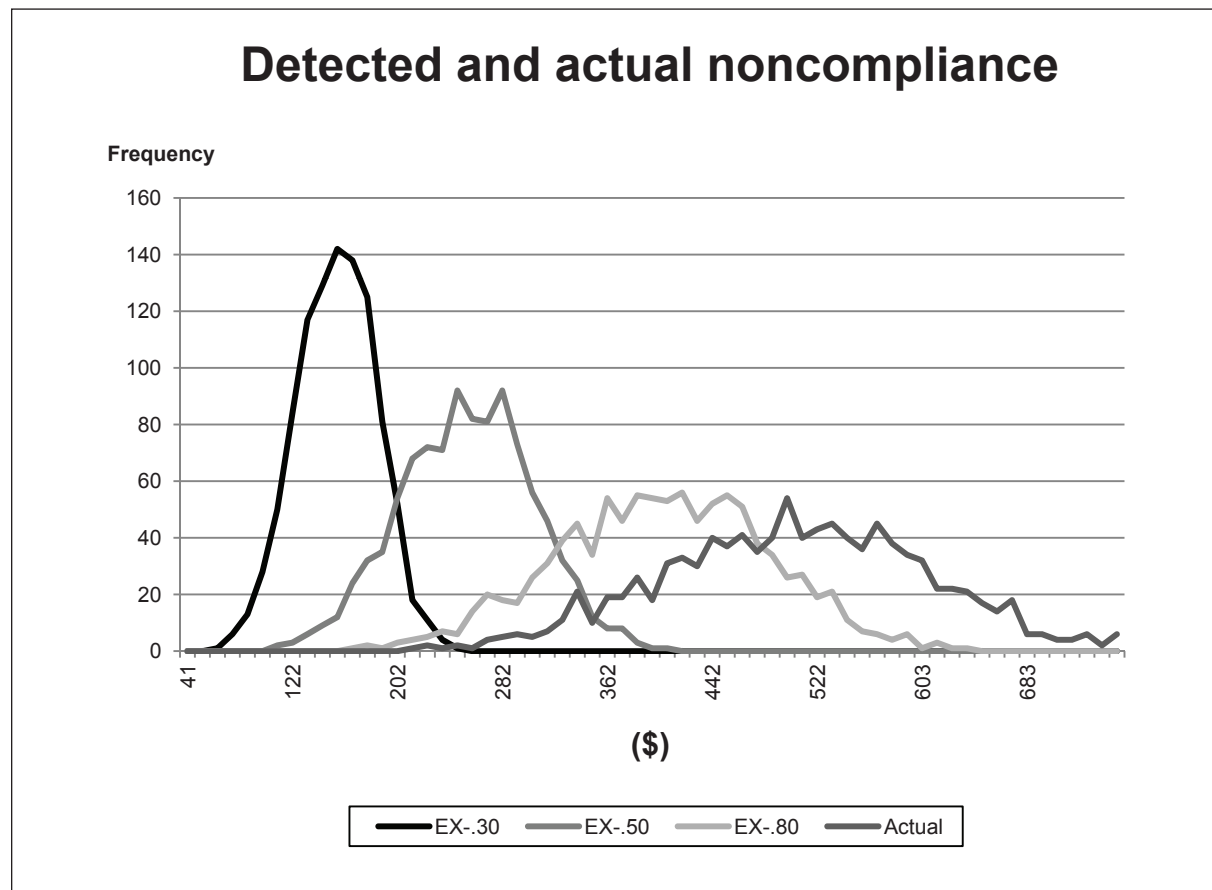
FIGURE 2. Illustration of Distribution of Detected Noncompliance by Examiner

We now recognize that each of the examiners had a fairly unique detection pattern. In particular, the examiner associated with the darkest line tended to detect relatively modest levels of noncompliance, while the examiners associated with the lighter gray lines each tended to uncover progressively larger amounts of noncompliance.

Using the observed relative detection rates for the three examiners, one can anticipate how the results for the examiners with the lower detection rates might be scaled up to approximate what the examiner with the highest detection rate would have uncovered had he been assigned to perform the audits in their place.

The actual distribution of noncompliance (including both the detected and undetected amounts in the population) is superimposed in Figure 3:

FIGURE 3. Illustration of Distributions of Actual and Detected Noncompliance

These results are based on a simulation in which the three examiners were able to detect, on average, about 30, 50, and 80 percent of noncompliance on a given return, respectively. After scaling up the results for the two examiners with the lower detection rates, one might have a predicted distribution of overall noncompliance in the population similar to the lightest gray line in the above figure. Certainly, this is much closer to the actual distribution of noncompliance than was the detected distribution presented in Figure 1. However, a comparison with the line representing the actual distribution of noncompliance indicates that it still somewhat under represents the true mean and variance of noncompliance in the population. If one knew something about the shape of the actual distribution of noncompliance, one might be able to make a more refined estimate of the overall distribution of noncompliance that improves on the lightest line.

Essentially, this is what the DCE approach does. It compares the relative detection performances of different examiners and combines this information with what is assumed about the distribution of noncompliance (e.g., that it has a skewed shape similar to the lognormal distribution) to scale up the examination results for a given line item to better represent the true level of noncompliance on a given return. Under such an approach, results for examiners with a relatively low detection rate on a given line item (when compared against examiners with similar levels of experience) receive a more substantial adjustment than those with a relatively high detection rate. Typically, examiners with the very highest detection rates receive only a very modest adjustment, suggesting that they were able to fully uncover nearly all noncompliance that was present for the line item on the returns that they examined.

4. NRP Data

We have adapted and refined the DCE methodology to estimate noncompliance on all key income sources on individual income tax returns using NRP data. This database contains the results of examinations of a stratified

random sample of approximately 45,000 tax returns. About 10 percent of these returns were either accepted as filed or subjected only to a correspondence examination targeting a small number of issues on the return. In past analysis, we have found that such returns have a rather limited potential for significant amounts of undetected noncompliance. Consequently, the focus in our research has been on those returns that were subjected to a more intensive face-to-face examination. Our estimation sample includes approximately 38,000 returns.

A key feature of the NRP face-to-face examinations is that not all line items on a given return were examined. Prior to the audit, an experienced IRS examiner known as a "classifier" reviewed the return as well as other available information known to the IRS (such as third-party information returns and prior tax return filings) and made some decisions regarding what line items should be examined. Some income sources were routinely examined (such as Schedule C and Schedule F when such schedules were filed with the return). Other line items were subject to the classifier's discretion and were "classified" for examination or not on the basis of his experience and judgment in light of the information available to him. So, for instance, if the amount of wages and salaries reported on a return was consistent with the amount shown on the W-2 forms available to the classifier, this line item might not be classified for examination. When a line item was classified for examination, the NRP examiner was instructed to conduct an audit of that item. In cases where a line item was not classified, the NRP examiner in most cases did not audit the item. However, the NRP examiner did have the discretion to audit an unclassified item if noncompliance was suspected. For instance, if an initial probe uncovered potentially unreported income on a given line item, the examiner was free to pursue this issue.

The TY2001 NRP data also include a "calibration sample" of approximately 1,200 returns that were subject to more thorough examination—something closer to the detailed line-by-line audit process employed under the predecessor Taxpayer Compliance Measurement Program that was in place through tax year 1988. We incorporate returns from the calibration sample in our analysis of selected income items; see section 5.2 below for details on how these calibration sample returns have been employed.

5. DCE Models

We employ DCE models to develop estimates of income underreporting by line item for most income sources reported on individual income tax returns.[1] These income sources fall broadly into one of two categories:

1. Income items that are fairly routinely classified for an NRP examination (at least when a nonzero amount is reported for the item on the return). This category consists of income items that are not subject to a high degree of third-party information, such as self-employment income and rental income.

2. Income items that are not routinely classified for examination. This category includes income items that are subject to substantial third-party reporting, such as wages and salaries and interest income.

For the first category of income items, we have developed a DCE model that incorporates equations describing the likelihood and magnitude of noncompliance as well as the propensity for noncompliance to be discovered during an examination. For the second category, we have extended this model to account for the classification process and for discretionary examinations of unclassified income items.

5.1 Model for Line Items with Routine Classification

Line items for income sources that are fairly routinely classified for NRP examination include:

1. Schedule C net nonfarm self-employment income;

2. Schedule F net farm self-employment income;

3. Schedule D net long-term capital gains;

4. Schedule D net short-term capital gains;

5. Schedule E net rents and royalties;

6. Schedule E other net income (partnerships, s-corporations, estates, trusts, etc.);

7. Form 4797 net supplemental gains; and

8. Form 1040 other income.

When a nonzero amount is reported for one of the above income sources, the line item or schedule is generally classified for examination. In our discussion, we focus on the portion of our model that addresses this situation. We also have developed and estimated a DCE specification for the case in which no income was reported for these income sources. However, we omit discussion of this case for the sake of brevity.

For the above line items, we specify a DCE model with three building-block equations. In this model, we distinguish between the actual level of noncompliance associated with an income source (N) and the detected level of noncompliance for the income source (as measured by the NRP examiner's adjustment A). If detection were perfect, the actual level of noncompliance for the income source would be equal to the detected amount (i.e., $N=A$); however, our model accounts for the possibility that detection is imperfect, in which case the adjustment (A) will understate the true level of underreporting on the line item (N) by some unobserved amount.

In addition to accounting for undetected noncompliance, our model allows for the fact that many taxpayers make fully compliant reports with respect to any given income item on a return (i.e., $N=0$). We do this by modeling the true level of noncompliance using a two-part specification:

$$P^* = \beta_P'x + \varepsilon_P$$
$$\ln N = \beta_N'x + \varepsilon_N.$$

This two-part specification accounts for two of the three building block equations in our DCE model. In this specification, P^* represents is a latent variable describing the propensity for noncompliance with respect to the income source being modeled. The propensity for noncompliance is assumed to depend on a set of taxpayer and tax return characteristics (x) as well as a random disturbance term (ε_P). The term β_P represents a set of coefficients of the explanatory variables that we estimate. If P^* is less than zero (implying a relatively low propensity for noncompliance), then the income source is fully reported on the return and noncompliance (N) is equal to zero. On the other hand, if P^* is greater than zero (implying a relatively high propensity for noncompliance), then the income source is underreported on the return to some extent, meaning than N is greater than zero. In that case, the magnitude of noncompliance is determined by the second equation of the model, which relates the natural log of N to our set of explanatory variables (x) and an error term (ε_N). The term β_N represents a second vector of coefficients that we estimate.

We employ the following rather standard two-part modeling assumptions:

1. ε_P and ε_N are independently distributed;

2. ε_P follows the standard normal distribution (mean zero and standard deviation one);

3. ε_N is normally distributed with mean zero and standard deviation σ_N.

Under these assumptions, the conditional magnitude of noncompliance (when it is present) follows the lognormal distribution. This modeling structure is consistent with experience, which suggests that many taxpayers make fully compliant reports and, among those who do understate their income, many do so by relatively modest amounts, while a small minority underreport by very large amounts. We note that we have experimented with alternative distributional assumptions (such as the generalized gamma); however we have found that the lognormal distribution performs reasonably well and makes estimation somewhat more straightforward.

In a standard two-part model, one observes the values of the dependent variable. If we wanted to assume that detection was perfect, we could in fact estimate our above specification by setting true noncompliance (N) equal to the examiner adjustment A. Although we still would not observe the latent noncompliance propensity P^*, we would in this case observe the noncompliance indicator P, defined by the expression:

$$P = \begin{cases} 1 & P* > 0 & (A > 0) \\ 0 & P* \leq 0 & (A = 0) \end{cases}$$

Having observable measures of the dependent variables P and N (as well as the set of explanatory variables x) would make it feasible to estimate our two-part specification, which would then permit us to examine how taxpayer compliance behavior on a given income source is associated with various taxpayer and tax return characteristics x. However, if our objective was simply to measure the detected level of underreporting within the tax return population with respect to the income source, we would not even need to estimate a model. Rather, we could just aggregate the individual NRP examiner adjustments A on each return using the NRP sample weights.

Thus, the fundamental reason for the complexity in our approach is that we want to account for the fact that NRP examiners are not always successful at uncovering noncompliance, meaning that actual noncompliance N is sometimes greater than the adjustment A. Our model accounts for imperfections in the NRP detection process via the third building block equation of our model:

$$D* = \beta_D'x_D + \varepsilon_D,$$

where $D*$ represents the propensity of the examiner to uncover noncompliance when it is present, x_D is a set of explanatory variables (including dummy variables for different NRP examiners, an indicator for whether the examination was conducted in the field rather than in the office, and the GS grade of the examiner), β_D is a vector of coefficients that we estimate, and ε_D is an error term assumed to follow the normal distribution with mean zero and standard deviation σ_D. We assume that this error term is independent of error terms in the first two equations. Let the detection rate (the fraction of noncompliance that the examiner is able to uncover be represented by D. Then the detection rate has the following relationship to the detection propensity $D*$:

$$D = \begin{cases} 1 & (Perfect\ Detection) & D* \geq 1 \\ D* & (Partial\ Detection) & 0 < D* < 1 \\ 0 & (Non-Detection) & D* \leq 0. \end{cases}$$

Our DCE model then consists of three equations that respectively describe the likelihood of noncompliance with respect to an income source on the tax return, the magnitude of noncompliance if it is positive, and the extent to which any noncompliance has been detected. We jointly estimate the parameters from all three equations of our model jointly using the method of maximum likelihood.[2] Although incorporating imperfect detection into our two-part specification of noncompliance significantly complicates the likelihood function, it adds an important sense of realism to the specification while still keeping it tractable to estimate.

The likelihood function for a model is defined in terms of the conditional probability distribution of the observed dependent variables given the control variables (x). So although the three equations of our model are defined in terms of the unobservable variables $P*$, N, and $D*$, the likelihood function must be defined in terms of the observed dependent variable A. In other words, we must derive the conditional distribution of A from the specified joint conditional distribution of these three unobserved response variables. Observe that the adjustment A is related to the actual level of noncompliance N and the detection rate D according to the following expression:

$$A = N * D.$$

Therefore, we can assess the conditional probability of an adjustment in the amount A by combining together the conditional probabilities associated with the various combinations of variables N and D that produce that value for A. To better understand how this process is carried out, it is useful to consider separately the cases where A is zero and where A is positive.[3]

When the adjustment A is equal to zero, there are two possibilities to consider:

1. The taxpayer was fully compliant in reporting the income item (i.e., $N=0$); or

2. The taxpayer understated the line item, but the examiner did not detect any of the noncompliance that was present (i.e., $N>0$ and $D=0$).

Observe that each of these cases will yield:

$$A = N*D = 0.$$

The likelihood associated with the first case is defined by the probability that $P* \leq 0$ (zero noncompliance). The likelihood associated with the second case is defined by the joint probability that $P*>0$ (some noncompliance is present) and $D* \leq 0$ (none of it was detected). The overall likelihood expression when the adjustment A is zero is computed as the sum of these two probabilities. Equivalently, it can be expressed as one minus the joint probability that $P*>0$ and $D* > 0$.

When the adjustment A is greater than zero, there are also two possibilities to consider:

1. All noncompliance was detected (i.e., $A=N$); or

2. Noncompliance was only partially detected (i.e., $A<N$).

The likelihood associated with the first of these cases is defined by the joint probability that $P*>0$ (some noncompliance is present) and $D* \geq 1$ (detection is perfect) multiplied by the probability density function for N, evaluated at $N=A$. Observe that for this case, we have:

$$A = N * D = N * 1 = N.$$

To determine the likelihood associated with the second case, one has to account for the fact that the detection rate D can take any value between 0 and 1. Therefore, it is necessary to integrate the joint probability density function for the adjustment A and the detection rate D over this range of values for D. This result is then multiplied by the probability that $P*>0$. Observe that for this case we have:

$$A = N * D < N; \ \ 0 < D < 1.$$

The overall likelihood expression when the adjustment A is positive is computed as the sum of the likelihood values associated with these two cases.

Estimation of the model yields estimates of the coefficients β_P, β_N, and β_D as well as the standard deviation terms σ_N and σ_D. Using these parameters, we are able to predict the conditional probability and magnitude of undetected noncompliance for a line item on a return given the NRP examiner's adjustment (detected noncompliance) A.

5.2 Model for Line Items with Non-Routine Classification

Line items for income sources that are not routinely classified for NRP examination include:

1. Wages and salaries;

2. Taxable interest;

3. Dividends;

4. State and local tax refunds;

5. Taxable pensions and IRAs;

6. Gross social security benefits;[4] and

7. Unemployment compensation.

For these income items, we extend our previous DCE model to incorporate a classification equation:

$$C^* = \beta_C' x_C + \varepsilon_C,$$

where C^* represents the propensity of a classifier to assign the line item to be examined on a return. In this probit specification, x_C represents a set of explanatory variables (including a set of classifier dummy variables as well as some variables measuring the discrepancy between the amount reported for a line item and the information shown on third-party reports) , β_C is a vector of coefficients to be estimated, and ε_C is a disturbance term that is assumed to follow the standard normal distribution. The return is classified for examination if and only if C^* is greater than zero. An important feature of our model is that we allow for the possibility that the classifier may observe some relevant information about the taxpayer (such as details from prior year tax returns) that is unavailable to us. We do so by allowing nonzero correlations between the classification equation error term ε_C and the noncompliance equation error terms ε_P and ε_N. These correlations account for factors not observed by us that may make it possible for the classifier to more effectively select which returns should be examined for a given line item.

When a line item is classified for examination in our model, an examination takes place and the three equations from our prior model continue to describe the probability and magnitude of noncompliance and the extent of detection. When a line item is not classified, we account for the possibility that the examiner uses his discretion and elects to audit the item. We do this by specifying the probability of an unclassified line item being examined as:

$$\Pr(Audit \mid Not\ \ Classified) = \frac{\exp(\alpha_0 + \alpha_1 N)}{1 + \exp(\alpha_0 + \alpha_1 N)}.$$

Under this logistic probability expression, the likelihood that an unclassified income item is examined is depends positively on the level of noncompliance with respect to the item (N). What we have in mind here is that examiners who decide to audit an unclassified line item probably have uncovered some signal that significant noncompliance is likely to be present. As a consequence, the unclassified returns they choose to audit will tend to be the ones with relatively large levels of noncompliance for the line item. The parameters α_0 and α_1 are coefficients that we estimate along with the other parameters of our model.[5]

The introduction of a classification equation and a logistic specification for the likelihood of an audit of an unclassified item complicate the likelihood function. To avoid an overly technical discussion, we will not provide a detailed explanation of the derivation of the likelihood function for this case. However, we do note that the likelihood function now involves a distinct expression for each of the following cases:

1. Classified, Positive Adjustment;

2. Classified, Zero Adjustment;

3. Not Classified, Examined, Positive Adjustment;

4. Not Classified, Examined, Zero Adjustment; and

5. Not Classified, Not Examined.

With the introduction of a classification equation, the model becomes more difficult to identify; specifically, it can be challenging to reliably estimate the correlation terms between the errors of the classification and noncompliance equations in a model of this sort. To improve identification, we have incorporated observations from the calibration sample into our analysis. For the calibration sample observations, we assume that each of the line items was examined on all returns. Since there is no classification issue for these observations, they provide an independent source of information about the noncompliance equation parameters, thereby making it easier to distinguish between the coefficients of these equations and the correlation terms of the model.

Estimation of the model yields estimates of the coefficients β_P, β_N, β_D. β_C, α_0, and α_1 as well as the standard deviation terms σ_N and σ_D. Using these parameters, we are able to predict the conditional probability and magnitude of undetected noncompliance for a line item on a return given the classification and examination outcomes that have been observed.

5.3 Need for Joint Estimation of Line items

A key feature of our methodology is that it exploits heterogeneity among examiners in their ability to detect noncompliance. To do this effectively, one needs to have a reasonable number of examiners who have each audited a given line item on a significant number of returns (say, 15 or more). While this condition is satisfied for Schedule C and Schedule F reports, it is not generally satisfied for the remaining income items that are the subject of our analysis.

We have therefore undertaken a joint estimation strategy for estimating groups of income items subject to a common detection equation. Essentially, our approach assumes that a given examiner has the same potential for detecting noncompliance (when it is present) on any of the line items included in the group. However, the specification continues to allow for differences in detection abilities across examiners and across groups of line items. It is important to note that our joint estimation strategy does not restrict either the level or the rate of undetected noncompliance to be the same for different members of a group of income items. The level and rate can vary across group members, both because neither the likelihood nor the magnitude of noncompliance have been constrained to be the same for different income sources and because the sets of examiners that have audited each source do not perfectly overlap.

We have two distinct groups of income items that are employed under our joint estimation strategy. The first is our set of seven income items that are subject to a high degree of third-party information reporting:

1. Wages and salaries;

2. Taxable interest;

3. Dividends;

4. State and local tax refunds;

5. Taxable pensions and IRAs;

6. Gross social security benefits; and

7. Unemployment compensation.

Recall that we employ our DCE model for return line items with non-routine classification as described in Section 5.2 for this group of income sources.

The second group includes the following six income items, which are subject to less substantial third-party information reporting:

1. Schedule D net long-term capital gains;

2. Schedule D net short-term capital gains;

3. Schedule E net rents and royalties;

4. Schedule E other net income

5. Form 4797 net supplemental gains; and

6. Form 1040 other income.

Recall that this group of income sources is estimated using our DCE model for return line items that are subject to routine classification as described in Section 5.1.

Since we have sufficient examiners who have each audited a significant number of Schedule C and Schedule F returns, we estimate our DCE model (for return line items subject to routine classification) separately for these income sources without grouping them with other line items.

6. Results

In general, the estimated detection rates for each of our models indicate significant heterogeneity across examiners, ranging from very low (sometimes single digits) to near-perfect rates.

Table 1 presents the average implicit DCE multiplier for several categories of income items. The implicit multiplier represents the conversion factor to produce an estimate of overall noncompliance (detected plus undetected) from an estimate of detected noncompliance.[6]

The high third-party information reporting group includes wages and salaries, taxable interest, dividends, state and local tax refunds, taxable pensions and IRAs, gross social security benefits, and unemployment compensation. For this group, the overall implicit DCE multiplier is 2.52, indicating that there is approximately $152 in undetected noncompliance on these line items for every $100 that is detected. Table 1 also breaks down the implicit multipliers for the cases where an income item was classified for examination and where the item was not classified. When items in the group were classified for examination, the implicit multiplier was only 1.46; however, it was much higher (5.37) when items were not classified. Recall that the examination of unclassified income items was at the discretion of the NRP examiner, and in the majority of cases no examination was conducted. The higher multiplier for unclassified income items accounts for undiscovered noncompliance on the significant portion of returns that went unexamined for the line items as well as undetected noncompliance on the smaller portion that were examined.

Table 1 also provides the implicit DCE multiplier for the group of six income items (net short-term and long-term capital gains, net rental and royalty income, other Schedule E income, Form 4797 net supplemental gains, and Form 1040 other income) that were routinely classified for examination (when reported on the return). Although our description of the DCE specification in Section 4.1 focused on the case where these income items were reported on the return, we mentioned that we also estimated an econometric specification for the case where they were not reported on the return. The overall DCE multiplier (3.26) for this group presented in the table accounts for both cases, indicating that approximately $226 in noncompliance went undiscovered for every $100 that was detected for these items. Intuitively, this is higher than the multiplier for the high third-party information return income category as it is more difficult to detect noncompliance in the absence of comprehensive third-party information reporting.

For this group of income sources, Table 1 breaks the DCE multiplier down for the cases where the income items were and were not reported on the tax return. Examination was fairly routine within this group when the income items were reported on the return. For this case, the multiplier was relatively low (2.86). In contrast, examinations were less common when the income items were not reported. Intuitively, the multiplier was much larger for this case (4.80).

TABLE 1. Implicit DCE Multipliers by Income Category

Income Category	Multiplier	Income Category	Multiplier
High 3rd Party Information Reporting*		**Schedule C**	
Classified	1.46	Schedule Reported	2.92
Not Classified	5.37	Scheduled Not Reported	16.4
Overall	2.52	Overall	3.4
Routinely Classified*		**Schedule F**	
Items Reported	2.86	Schedule Reported	3.18
Items Not Reported	4.80	Schedule Not Reported	20.0
Overall	3.26	Overall	3.41

*These implicit multipliers are averaged over several line items, which were estimated separately.

Table 1 also presents implicit DCE multiplier estimates for Schedule C (nonfarm) and Schedule F (farm) net self-employment income. These estimates reflect the cases where these schedules were and were not filed with the return. As we would expect, the implicit multipliers for Schedule C (3.47) and Schedule F (3.41) are high relative to the other income categories as these income sources are subject to a very low degree of third-party information reporting.

Table 1 breaks down the overall multipliers for Schedule C and Schedule F to compare the cases where the schedules were and were not filed with the income tax return. In the former case, examination was routine and the multipliers were relatively low (2.92 and 3.18, respectively). In the latter case, examination was at the discretion of the examiner and often was not undertaken. The implicit DCE multipliers for this case were much higher (16.4 and 20.0, respectively), suggesting that a significant number of taxpayers failed to report their self-employment earnings and escaped detection during the NRP examination process.

Table 2 presents the estimated Net Misreported Amounts (NMAs) and Net Misreporting Percentages (NMPs) associated with the same income categories covered in Table 1. The estimated NMA represents the difference between the amount of income that federal individual income tax filers are required to report on their tax returns and what they actually report. Our estimated NMA accounts not only for the noncompliance on returns subject to face-to-face audits in our study, but also returns that were accepted as filed or subject to a correspondence examination. For the face-to-face audit cases, the results are based on our DCE estimates. For the other cases, our estimates rely on the additional tax recommended by the NRP examiners without any adjustment for undetected noncompliance. Based on prior analysis, we do not believe that the magnitude of undetected noncompliance for such cases is likely to be very substantial. Across all of the income sources we have analyzed in our study, the overall NMA is estimated to be $805 billion.[7]

TABLE 2. Net Misreported Amount and Net Misreporting Percentage by Income Source

Income Category	Net Misreported Amount ($B)	Net Misreporting Percentage
High 3rd Party Information Reporting	88	1.6%
Routinely Classified	359	28.5%
Schedule C	330	54.8%
Schedule F	27	51.4%
All Income Categories Combined	805	10.9%

The Net Misreporting Percentage (NMP) is also a measure of aggregate reporting noncompliance for a given income source, but is expressed as a rate. It is computed as the ratio of the NMA for the income source to the sum of the absolute values of the amounts that should have been reported across returns.[8] Consistent with prior IRS research, the estimated NMP for income items in the high third-party information reporting category of returns is very low (less than 2 percent). In contrast, the NMP for the group of six income items subject to routine classification (and less substantial information reporting) is much higher (28.5 percent). Again, this is an indication that income sources that are subject to less comprehensive third-party information reporting

tend to have significantly greater potential for noncompliance. For Schedule C and Schedule F, which are subject to very little third-party reporting, the estimated NMPs are even higher (54.8 and 51.4 percent, respectively). Across all of the income categories in our analysis, the overall NMP is just under 11 percent, suggesting that, as a group, U.S. federal individual income tax filers do report a very substantial portion of the income they are required to report on their returns.

Table 3 breaks down the shares of the overall estimated NMA attributable to different income sources. Overall, a very large share of the gap (41 percent) is attributable to underreporting of Schedule C net income. When Schedule F underreporting is included, the share of the overall NMA attributable to understated net self-employment earnings amounts to 44.4 percent. A similar share of the overall gap (44.6 percent) is attributable to the six income items in our category of income sources that are generally subject to only a modest degree of third-party information reporting. In contrast, only about 11 percent of the overall NMA is attributable to income items that are subject to reasonably comprehensive third-party information reporting.

TABLE 3. Share of Overall Income NMA by Income Category

Income Category	Net Misreported Amount ($B)	Share of Overall NMA
High 3rd-Party Information Reporting	88	11.0%
Routinely Classified	359	44.6%
Schedule C	330	41.0%
Schedule F	27	3.4%
All Income Categories Combined	805	100.0%

7. Directions for Future Research

There are several directions for future research. First, the IRS is working to incorporate our line item DCE estimates of income underreporting into its model for developing updated estimates of the individual income tax gap—the difference between individual income taxes that should have been paid and the amount actually paid on time without enforcement action. Second, it would be desirable to incorporate the DCE results into a micro-simulation model that would permit the IRS to analyze a wide variety of "what-if" questions regarding changes in the composition of the taxpayer population or the tax treatment of various line items on the return. Such a model would be particularly useful if it also incorporated other taxpayer information, such as taxpayer burden estimates. A third avenue for future research concerns how best to adapt the DCE methodology to account for the new NRP sampling design, which involves annual audits of stratified random samples of approximately 13,000 individual income tax returns per year. A fourth direction for further research relates to the application of the DCE methodology to other NRP data sources, such as the recent NRP studies of S-corporation and employment taxes.

References

Feinstein, Jonathan S. (1990) "Detection Controlled Estimation," *Journal of Law and Economics,* Vol. 33, No. 1, April, pages 233-76.

Feinstein, Jonathan S. (1991) "An Econometric Analysis of Income Tax Evasion and Its Detection," *RAND Journal of Economics*, Vol. 22, No. 1, Spring, pages 14-35.

Endnotes

[1] We employ our DCE analysis to control for undetected misreporting with respect to income sources. Although various offsets (such as itemized deductions and credits) are also subject to misreporting on tax returns, the burden of proof for such items is on the taxpayer to justify the amounts claimed. Consequently, our working assumption is that examiner adjustments for offset items are a reasonably accurate reflection of noncompliance with respect to these items.

[2] In some cases, the number of parameters to be estimated precludes simultaneous estimation of all parameters owing to computer memory limitations. In such cases, we divide the parameters into groups and employ an iterative stepwise maximization procedure that converges to the global maximum over all parameters.

[3] In our analysis, we treat negative audit adjustments (i.e., cases where the examiner has determined that income has been over reported) as an assessment of zero.

[4] We model underreporting with respect to gross rather than taxable social security benefits to focus on cases of direct misreporting. In other words, we wanted to focus on cases where taxable social security benefits were understated as a result of understating gross benefits. The degree to which gross social security benefits are taxable depends on the amount of income one has from other income sources. We wanted to exclude from our analysis cases of indirect misreporting of taxable social security benefits that were solely attributable to understatements of other forms of income.

[5] Since this probability expression depends on the unobserved true level of noncompliance N, it is necessary to integrate over the possible values that N may take given the observed examiner adjustment A when estimating our specification.

[6] Note that these implicit multipliers are not comparable to the aforementioned 3.28 multiplier derived from the 1976 TCMP study. That multiplier (because of the way it was derived) was applied only to amounts detected without the aid of information documents; it was not applied to all adjustments that examiners made as these implicit multipliers could be. The overall implicit multiplier corresponding to the 3.28 figure, which would be comparable to these figures, was therefore significantly less than 3.28. Of course, it is also important to keep in mind that the predecessor TCMP program involved intensive line-by-line audits of the entire tax return. As discussed, NRP audits are more selective and therefore may have greater potential for undetected noncompliance.

[7] Note that this is not a tax gap estimate but rather an estimate of aggregate income underreporting. It would be necessary to apply a tax calculator to assess the degree to which this income underreporting translates into an understatement of tax liability. In a full tax gap analysis, one would also account for the tax implications of misstatements of adjustments, itemized deductions, and credits.

[8] The denominator of the NMP measure was computed by adding the aggregate estimated level of underreporting with respect to a given income source to the sum of the absolute values of the amounts actually reported for this source across returns. This approach is somewhat different than the official IRS approach of first combining the estimates of reported and unreported income before taking absolute values. Consequently, the NMP reported in this study will tend to differ from the official IRS measure.