

# Estimates of the Tax Year 2006 Individual Income Tax Underreporting Gap

*Kim Bloomquist, Ed Emblom, Drew Johns, and Patrick Langetieg, IRS Office of Research*

---

IRS recently estimated the Tax Year (TY) 2006 gross tax gap (the difference between the total tax imposed by law and the amount of tax paid timely) to be \$450 billion. Individual income taxes were underreported by \$235 billion, representing 19 percent of the amount of individual income taxes that should have been reported on returns that were filed on time.<sup>1</sup> Self-employment taxes were underreported by \$57 billion, 59 percent of the amount that should have been reported on individual income tax returns that were filed on time. The individual income tax and self-employment tax underreporting gaps combined accounted for 65 percent of the total gross tax gap in TY 2006. The tax gap estimates show that as information reporting increases, underreporting of that income tends to decrease. Only 8 percent of income subject to substantial information reporting (but not withholding) was underreported, while 56 percent of income subject to little or no information reporting was underreported.<sup>2</sup>

Prior to the release of the TY 2006 tax gap estimates in January 2012, the last tax gap estimates were for TY 2001, which were released in February 2006. The TY 2006 individual underreporting gap reflects new data and significant advancements in methodology over the TY 2001 estimates. The ideal data for estimating reporting compliance would both represent the population of tax returns and be complete in the detection of underreported tax. The individual underreporting gap estimate was not an actual tallying of observed misreported taxes across all individual accounts. Instead, misreported individual income taxes for the population were estimated by auditing a stratified random sample of individual income tax returns. Since the taxpayers were randomly selected, the sample was representative of the population of individual income tax returns. However, misreported income detected during the course of a given audit most likely did not account for all of the income that should have been reported on the return.

In order to account for all underreporting, the amount of income that was not detected during the audit was estimated using an econometric technique called detection controlled estimation (DCE). The DCE technique was applied for both TYs 2001 and 2006 tax gap estimates. However, for TY 2006 the DCE methodology was expanded to produce estimates of undetected income for individual line items—unlike for TY 2001 where estimates were made for aggregated line items. The TY 2006 estimate also reflects an improvement in the calculation of the tax liability based on the “true” income as expanded by the DCE estimates of undetected income. For TY 2006, IRS used a tax calculator to estimate the marginal increase in tax related to the unreported income on each tax return—an improvement over the prior use of aggregate average marginal tax curves. The remainder of this paper describes in greater detail the data and methodology used to estimate the TY 2006 individual income tax underreporting gap.

## National Research Program

The IRS National Research Program (NRP) designs and administers reporting compliance studies for the IRS.<sup>3</sup> The first NRP study of individual reporting compliance consisted of a stratified random sample of about 45,000 TY 2001 individual income tax returns filed during calendar year 2002.<sup>4</sup> That study served as the basis for the TY 2001 individual underreporting gap estimates. NRP used a process called classification to determine the type of audit for each return selected and the mandatory issues to be examined.<sup>5</sup> In the classification process, examiners compared information return documents (W-2s, 1099s, etc.) with the actual tax return in order to identify discrepancies and also identify items that appeared large, unusual, or questionable. Some line items on the return, typically those that could not be verified through information returns, were always classified as mandatory to audit. In the case of simple returns where information could be easily reconciled with

the information returns, taxpayers were not audited and actually not even contacted. Returns that had only a small number of simple issues identified in classification were routed to correspondence exams where the exam could be handled through telephone calls, faxes, and traditional mail. More complicated returns were assigned to one of two types of audits that involved face-to-face interaction with an examiner: either an office audit handled by a Tax Compliance Officer (TCO) or a field audit handled by a Revenue Agent (RA) who may actually visit the taxpayer's place of business.

The purpose of the classification process was to limit the burden on taxpayers by selecting an appropriate audit technique and set of issues. The number of mandatory issues on an NRP audit still typically exceeded the number of issues that would have been examined had the return been selected through one of the typical IRS compliance risk-based return selection processes. In that sense they were more complete audits, which was beneficial from a research perspective. Examiners also had the discretion to expand the audit to include non-classified issues, typically whenever additional information was uncovered during the course of the audit that caused the examiner to question that issue.

Although research from the study provided long-term benefits,<sup>6</sup> the random selection of 45,000 tax returns for a single tax year was a disruption to the normal exam workload. Instead of randomly selecting a large number of returns for a compliance study every few years, IRS decided to begin selecting a smaller number of returns on an annual basis. The smaller number of returns would provide more timely information and could be combined over several years to provide compliance estimates at a similar level of reliability as a single-year larger study. The smaller annual studies were designed to randomly select about 13,000 returns each year, beginning with TY 2006. Returns would go through a similar classification process as the TY 2001 study; there would just be fewer returns for a given tax year. Since the TY 2006 underreporting gap was estimated before enough of the annual studies were complete, both the TY 2001 and TY 2006 NRP studies were used to estimate the TY 2006 individual underreporting gap, as described below.

## Detection Controlled Estimation (DCE)

Not all underreported income is detected by every audit, even ones of the scope and quality of NRP audits. This was confirmed by the 1976 IRS Taxpayer Compliance Measurement Program (TCMP) individual reporting compliance study, which was the last IRS reporting compliance study to audit taxpayers without the auditors having the use of third-party information return documents. IRS later compared the information return documents to the audit findings and found that for every \$1.00 of detected unreported income that was reported on information documents, an additional \$2.28 went undetected.<sup>7</sup> As a result of that study, IRS began multiplying the portion of income detected without the use of information documents by a multiplier, typically 3.28, in order to estimate the individual underreporting gap.

In the late 1980s, Jonathan Feinstein developed an econometric technique for estimating undetected income that he termed detection controlled estimation.<sup>8</sup> The intuition behind the methodology was that examiners have varying abilities for detecting income that can be observed through patterns in the data collected from taxpayer audits. Feinstein explained that the observed audit adjustment actually reflects the product of the true (unobserved) unreported income and the propensity of the examiner to detect unreported income. Feinstein's application of the methodology to TCMP data resulted in comparable estimates for the amount of undetected income as the IRS was assuming based on the 1976 TCMP study.

The original DCE methodology focused on estimating overall noncompliance for a given return, while IRS was also interested in the sources of noncompliance. The IRS Office of Research contracted with Dr. Brian Erard (B. Erard and Associates) and Professor Feinstein (Yale School of Management) to extend and refine Professor Feinstein's original DCE methodology.<sup>9</sup> The first extension of the DCE methodology for the IRS estimated noncompliance separately for two groups of returns and two types of income items using the face-to-face audit results from the TY 2001 NRP study as part of the TY 2001 underreporting gap estimation process. Returns without reported Schedule C or Schedule F income and with reported total positive income (TPI) less than \$100,000 were estimated separately from all other returns. Similarly, income lines subject to less information reporting were estimated separately from all other income lines.

This extension provided IRS with four new multipliers that were used to estimate the TY 2001 individual underreporting gap. While an improvement over the old TCMP multiplier method, that approach was still primarily an aggregate approach to DCE estimation. Only the aggregate undetected income for income lines expected to have similar compliance characteristics and detection rates was estimated. In order to estimate the tax gap for specific income items, the TY 2001 DCE estimates required the assumption that average detection rates for income items grouped together for estimation were equal. Additionally, the use of a multiplier meant that undetected income was allocated only to returns where some unreported income was initially detected by the auditor. This approach also precluded the use of a micro tax calculator to estimate the tax value of the underreporting, so tax was estimated using average marginal tax rates for the TY 2001 tax gap estimates.

These assumptions meant that some line items received more undetected income than they should have while other income items received less, and that some returns were allocated more undetected income than they should have been while other returns were allocated less. For example, net capital gains or losses were grouped with income items covered by significant amounts of information reporting (wages and salaries, interest income, dividend income, social security income, etc.). Given the complex nature of capital gains transactions and the fact that the basis was not reported on information documents, unreported capital gains may actually have been more difficult to detect than other income items in its TY 2001 DCE estimation group.

### ***Estimation at the Income Item Level***

For estimating the TY 2006 individual underreporting gap, a second extension of the DCE methodology provided separate estimates of undetected income for income lines on the Form 1040. This extension allowed for greater variability in the average detection rates across line items.<sup>10</sup> DCE estimation requires explicit modeling of a detection equation whose arguments include the type of examiner (TCO or RA), the experience of the examiner, and binary variables that take the value of 0 or 1 to indicate which examiner conducted the exam. In order to differentiate the detection capabilities of different examiners, the examiners included in the detection equation must have audited a sufficient number of returns with the income item being modeled. Typically, this requirement would be 15 or more returns. Since the TY 2006 NRP study was less than a third of the size of the TY 2001 study, there were not enough observations to estimate the DCE equations. It may take three or more annual NRP studies before IRS has enough observations to estimate the DCE equations on the more recent data. Therefore, the TY 2001 NRP data was used to estimate the equation parameters for the second DCE extension. An imputation methodology, discussed later, was developed to allocate the DCE estimates derived from TY 2001 NRP data to the new TY 2006 NRP returns.

In addition to the detection equation, the second extension of the DCE methodology included a two-part specification for modeling the noncompliance of a line item. The first noncompliance equation modeled the likelihood of noncompliance while the second equation modeled the magnitude of noncompliance conditional on the presence of noncompliance. Since some income items with significant information reporting were not routinely classified, the extension also included an additional modeling of the likelihood of the item being classified based on its return characteristics and actual mismatches with information documents for these items.

The data requirements for DCE meant that some income items still needed to be grouped together for purposes of estimating the detection equation, even when using data from the larger TY 2001 NRP study. Table 1 shows the specific groupings of income items used for estimation. Income items that were routinely classified (typically because of the general lack of complete information reporting) were modeled separately from items subject to significant information reporting (wages, interest income, etc.). Schedules C and F income were primarily estimated independently of each other and of other routinely classified income items. Other routinely classified income items (capital gains, rental and royalty income, partnership and S corporation income, etc.) were estimated jointly with a common detection equation. Similarly, items subject to significant information reporting were also estimated jointly with a common detection equation.

**TABLE 1. Grouping of Income Items for Joint Estimation**

Items Subject to Significant Information Reporting	Items Routinely Classified	
Estimated Jointly	Estimated Jointly	Estimated Separately
Wages and Salaries	Short-term Capital Gains	Schedule C
Interest	Long-term Capital Gains	Schedule F
Dividends	Rents and Royalties	
State and Local Tax Refunds	Partnership, S Corporation, Estate, Other Income	
Pensions and IRAs	Form 4797 Net Gains	
Gross Social Security	Other Income	
Unemployment		

The joint estimation of some line items with a common detection equation meant that the expanded methodology assumed that a given examiner had similar detection capabilities across all of the income items within the group. Unlike the first extension of DCE, noncompliance of each income item was modeled using separate equations and parameters even though detection was modeled using a common equation. In other words, the equations and parameters that modeled the likelihood and magnitude of noncompliance were not constrained to be identical across line items within a group while the detection equation and parameters were constrained. The second extension explicitly provided separate estimates of undetected income for each income item, a marked improvement over the first extension. Additionally, because different examiners may have examined different income items, the overall average detection rates for a given line item could still vary within the group. Although separate detection equations would be preferred to the use of a common detection equation, there were simply not enough audits in the sample to support that level of detail.

The use of multipliers to expand from detected underreported income to total underreported income in the original TY 2001 DCE implementation meant that income could be allocated only to returns where unreported income was detected by the examiner. The second extension of the DCE methodology provided return level predictions of undetected income based on the probability and magnitude of undetected income conditional on whether or not unreported income was initially detected by the examiner. One significant implication of this enhancement was a more realistic distribution of undetected income. The more realistic distribution of undetected income enabled IRS to use a tax calculator, discussed later, to estimate the tax associated with all unreported income on each return in the sample separately.

## **Two-Stage Imputation of Undetected Income from TY 2001 to TY 2006**

The smaller size of the TY 2006 NRP sample prevented IRS from using that data to estimate the DCE equation parameters. The primary purpose of the DCE estimation is to estimate how much underreported income was not detected on the NRP audits. Under the assumption that the average propensity of examiners to detect underreported income remained stable between the TY 2001 and TY 2006 NRP studies, the detected underreported income from the TY 2006 NRP data could still provide much of the information needed for estimating the TY 2006 individual underreporting gap. The imputation of undetected income from TY 2001 NRP data to TY2006 NRP data took place over two stages. The first stage generated 10 simulated TY 2001 NRP data sets with return level predictions of undetected income. During the second stage, those 10 simulated TY 2001 data sets were used to generate 10 simulated TY 2006 NRP data sets with return level predictions of undetected income.

### ***Stage 1: TY 2001 DCE Simulations***

The DCE formula underlying the return level predictions predicts a positive probability of undetected income for most returns (though this is typically very small for returns where no unreported income was detected). Simply multiplying the predicted probability of undetected income by the predicted magnitude of undetected

income would have resulted in nearly every return receiving some positive amount of undetected income for each income item—probably an unreasonable outcome. A small probability of undetected income for an income item actually means that undetected income would be present on a relatively small number of returns for that item. Since one goal was to have a more realistic allocation of undetected income, a simulation approach was developed in order to apply the DCE prediction formulas. The simulation process essentially randomly allocates undetected income for a given income item based on the probability of undetected income for that item on each return.

The specific steps of the TY 2001 simulation are described below. For each return:

- Step 1: Calculate the predicted probability of the presence of undetected income conditional on whether unreported income was detected by the examiner.
- Step 2: Calculate the predicted magnitude of total unreported income conditional the presence of undetected income.
- Step 3: Draw a random number between 0 and 1.
- Step 4: If the random number is less than or equal to the predicted probability from Step 1, allocate the predicted total (detected + undetected) amount of unreported income from Step 2. Otherwise, allocate only the detected amount of unreported income (if any).
- Step 5: Calculate the weighted sum of predicted total unreported income from Step 4 across all returns to estimate unreported income for the population.

Steps 3 to 5 were repeated 10 times for each income item to create 10 sets of TY 2001 data with simulated undetected income.

### ***Stage 2: TY 2006 DCE Imputations***

In the past, IRS has relied upon “implicit” multipliers defined as the total estimated underreported income divided by detected underreported income. Implicit multipliers cannot accommodate the allocation of undetected income to returns on which no income was detected. Since the new DCE methodology explicitly provides estimates for these returns, IRS desired a new approach to applying the estimates of the average propensity to detect underreported income across NRP studies. The 10 simulations created during Stage 1 meant that undetected income could now be allocated directly to each TY 2006 NRP return (by income item), conditional on certain assumptions. First, detection of unreported income for a given line item was assumed to be different for returns that reported that line item versus returns where the taxpayer did not report any income for that line item. Second, if income was reported for the line item, detection was assumed to vary with the amount of the line item that was reported. Third, if income was not reported for the line item, detection was assumed to vary with the amount of adjusted gross income (AGI) that was reported.

Undetected income was imputed separately for each income item. Each TY 2001 NRP simulated data set was divided into returns that reported the income item and returns that did not report the income item. For returns that reported the income item, weighted deciles of the reported amount were calculated. For returns that did not report the income item, weighted deciles of AGI were calculated. If there was not sufficient data to calculate deciles, an alternative percentile was calculated.<sup>11</sup> For each income item, the calculation of deciles provided up to 20 groups of returns, or bins, (10 income item deciles and 10 AGI deciles). The TY 2006 NRP data were similarly apportioned according to income item and AGI deciles.

Using the first TY 2001 simulated data set, for each bin:

- Step 1: Calculate weighted probability of the presence of undetected income, defined as the weighted number of returns with undetected income divided by the total weighted number of returns.
- Step 2: Calculate the weighted mean amount of undetected income conditional on the presence of undetected income, defined as the weighted sum of undetected income divided by the weighted number of returns with undetected income.

For each TY 2006 return:

- Step 3: Draw a random number between 0 and 1.
- Step 4: If the random number is less than or equal to the weighted probability from Step 1, allocate the mean amount of undetected income from Step 2. Otherwise, do not allocate undetected income.
- Step 5: Multiply the mean amount of undetected income by the ratio of the weighted detected underreported amount from TY 2006 to the weighted mean detected underreported amount from TY 2001.

Steps 1 to 5 were repeated for each income item for each of the 10 TY 2001 simulated data sets to create 10 TY 2006 simulated data sets with return level predictions of undetected income.

## **Additional Income Adjustments**

### ***Tip Income***

For some line items, DCE is unlikely to fully account for all undetected income. Since tip income is relatively concentrated in a few industries and occupations, tip income represents a relatively small amount of overall wages, salaries, and tips. However, since a significant portion of tip income is paid in cash by customers, tip income is subject to less information reporting than most wages and salaries. The lack of complete information reporting and the cash nature of tips suggest that tip income had a lower compliance rate than other wages and salaries and was harder to detect during an audit. Given the concentration of tip income and the nature of the NRP samples, the design of the TY 2001 and TY 2006 NRP studies did not support estimates of unreported tip income. Furthermore, tip income earners who filed a Form 1040-EZ and who reported all of the income reported to them on their W-2 would be less likely to be subject to a face-to-face NRP audit compared to other cash intensive sources of income, like Schedule C income. Therefore, a separate estimate of unreported tip income based on prior IRS studies of tip income compliance supplemented the DCE estimate of undetected wages and salaries.

### ***S Corporations, Partnerships, Estates and Trusts***

With flowthrough income, there are two potential sources of misreporting: misreporting by the individual shareholder/partner and misreporting by the S corporation/partnership on the K-1 sent to both the shareholder/partner and IRS. On most NRP audits, examiners reconciled the income reported on the K-1 with the income reported by the individual taxpayer on Schedule E. Examiners looked at other individual level issues such as various limitations on the deduction of losses related to basis or passive-activity rules. It was very uncommon for examiners to examine the related closely held S corporation and partnership entities. Therefore, very little entity level misreporting was detected on the TY 2001 and TY 2006 NRP audits.

IRS previously conducted a separate reporting compliance study of S corporations spanning Tax Years 2003 and 2004. Preliminary results of the TYs 2003/2004 NRP S corporation study were reported at the 2009 IRS Research Conference and suggested the net misreporting percentage (NMP) of income by S corporations was 15 percent.<sup>12</sup> However, the percentage of flowthrough income misreported by individuals based on the TY 2001 and TY 2006 NRP individual studies, after attempting to account for undetected income using DCE, was below the estimated percentage of misreported income by S corporations from the TY 2003/2004 NRP S corporation study. Since this seemed unlikely, the final tax gap estimate adds a small amount of underreported income such that the overall percentage of misreported income for the S corporations, partnerships, estates and trusts rises to 15 percent, the level observed in the TY 2003/2004 NRP study of S corporations.

## **Tax Calculator**

The imputation of return-level predictions of undetected income from the TY 2001 simulations to the TY 2006 NRP data provided estimates of total underreported income, but not underreported tax. To estimate

underreported taxes resulting from the underreported income, a tax calculator was applied to individual observations (i.e. tax returns) from the ten simulated TY 2006 data sets. The use of a tax calculator was an improvement over prior methodologies of estimating average marginal tax rates, particularly since the tax calculator could account for different statutory tax rates between long-term capital gains and dividends versus other sources of income.<sup>13</sup> This process provided ten underreporting gap estimates for each line item which were then averaged to produce the final underreporting gap estimate. The final line item underreporting gap estimates were summed to estimate the overall individual income tax underreporting gap. The specific process for estimating the underreporting gap for each income item using the tax calculator is described below. Essentially, the additional income for each income item was added (or subtracted) to the reported amount of income and tentative tax calculated. Then that additional income was dropped and the process repeated for the next income item.

Using the first simulated TY 2006 NRP data set from Stage 2 of the imputation:

#### **Income and Deductions**

Step 1: Calculate tentative tax based on reported income and deductions.

Step 2: Add net misreported wages, salaries, and tips and recalculate tentative tax.

Step 3: Subtract tentative tax calculated in Step 1 from tentative tax calculated in Step 2. This was the estimate of the underreporting gap for wages, salaries, and tips from the first simulated data set.

Step 4: Remove the unreported wages, salaries, and tips added during Step 2.

Step 5: Repeat Steps 2 to 4 for the remaining income items and deductions separately.

#### **Credits<sup>14</sup>**

Step 1: Calculate total credits based on reported income and deductions.

Step 2: Add all net misreported income and deductions to all line items and calculate total credits.

Step 3: Subtract total credits calculated in Step 2 from total credits calculated in Step 1. This was the estimate of the underreporting gap for total credits from the first simulated data set.

The separate steps for income and deductions and credits were repeated for each of the 10 simulated TY 2006 data sets and then the results were averaged. Had net misreported income been added to all line items simultaneously, the resulting calculation of tentative tax would have been larger than the estimate obtained by summing the marginal increases to tentative tax. Increased total income would have increased marginal tax rates due to the progressivity of income taxes, increasing the estimate of the total underreporting gap. However, the DCE estimation and imputation methodology included other assumptions, such as the imputation of the mean undetected income by decile. Those assumptions also have potential impacts on the distribution of income and therefore marginal tax rates. Taken as a whole, it's not clear that the complete DCE estimation, imputation and tax calculation methodology resulted in an estimate that could be interpreted as a lower bound.

### ***Self-Employment Taxes***

Self-employment taxes are required to be reported by individuals with self-employment income on individual income tax returns. The underreporting of self-employment income (primarily income reported on Schedules C and F) results in underreported self-employment taxes. Therefore, the TY 2001 and TY 2006 NRP studies and the tax calculator were used to estimate the self-employment tax underreporting gap. Self-employment taxes were calculated before and after including unreported self-employment income for each of the ten simulated TY 2006 data sets. Each spouse on a joint return has a separate earned income threshold above which the combined wages and self-employment income are subject to Medicare taxes but not Social Security taxes. Since the undetected income is allocated to the return as a whole and not to individual spouses on the return, self-employment tax was calculated under two different assumptions concerning the threshold. Under the first assumption, the unreported self-employment income was allocated to a single spouse. Under the second

assumption, the unreported self-employment income was allocated evenly to both spouses. The final self-employment tax underreporting gap estimate was the average of the estimates resulting from the two different assumptions. For TY2001 the average of the two estimates (\$35.3 billion and \$42.7 billion) resulted in the same estimate under the new DCE methodology as the estimate under the original DCE methodology. For TY2006, the average of the two estimates (\$52.4 billion and \$61.1 billion) was \$57 billion.

## TY 2006 Estimates

Table 2 shows the estimated tax gaps by individual income tax component for TY 2006 and TY 2001. Business income reported on Schedules C, E, and F accounted for just over half of the total individual income underreporting gap in both TY 2001 and TY 2006. Overall, the estimated NMP for TY 2006 was 19 percent, not significantly different from the estimated NMP in TY 2001, shown in Table 3. The underreporting gap associated with credits increased from \$17 billion in TY 2001 to \$28 billion in TY 2006. The increase in the underreporting gap for credits, shown in Table 3, was consistent with the growth in refundable credits between TY 2001 and TY2006. Notably, the reported amount of the earned income credit grew 30 percent while reported amount of the additional child tax credit tripled.

**TABLE 2. Individual Income Tax and Self-Employment Tax Underreporting Gaps, Tax Years 2001 and 2006**

(In billions of dollars)

Tax Gap Component	TY 2006	TY 2001
<b>Individual Income Tax</b>	<b>235</b>	<b>197</b>
Nonbusiness Income	68	56
Business Income	122	109
Adjustments, Deductions, Exemptions	17	15
Credits	28	17
<b>Self-Employment Tax</b>	<b>57</b>	<b>39</b>

**TABLE 3. Individual Income Tax Underreporting Gap Estimates by Visibility Category, Tax Years 2001 and 2006**

Tax Return Line Items Grouped by Visibility Category	Underreporting Gap (In billions of dollars)		Net Misreporting Percentage*	
	TY 2006	TY 2001	TY 2006	TY 2001
<b>Total Individual Income Tax</b>	<b>235</b>	<b>197</b>	<b>19</b>	<b>18</b>
Substantial Information Reporting and Withholding <sup>1</sup>	11	11	1	1
Substantial Information Reporting <sup>2</sup>	12	9	8	5
Some Information Reporting <sup>3</sup>	64	51	11	9
Little or No Information Reporting <sup>4</sup>	120	110	56	54
Tax Credits	28	17	37	26

\* Net misreporting percentage is the net misreported amount divided by the sum of the absolute values of the amounts that should have been reported.

<sup>1</sup> Wages and salaries

<sup>2</sup> Pensions & annuities, unemployment compensation, dividend income, interest income, Social Security Benefits

<sup>3</sup> Deductions, exemptions, partnership/S corporation income, capital gains, alimony income

<sup>4</sup> Nonfarm proprietor income, other income, rents and royalties, farm income, Form 4797 income, adjustments

NOTE: Components might not sum to totals because of rounding.

Table 3 also shows that the main finding from TY 2001 has not changed. As the level of information reporting increases, the percentage of income that is misreported (as reflected by the NMP) decreases. When there was little or no information reporting, 56 percent of income was underreported in TY 2006. That contrasts with income subject to substantial information reporting, where only 8 percent of income was underreported. In the case of wages and salaries, where there was both substantial information reporting and withholding, only 1 percent of income was underreported.

## Conclusion

The TY 2006 individual income tax underreporting gap estimate represents a significant methodological advance compared to prior tax gap estimates. Specifically, the latest estimate incorporates line item level estimates of undetected underreported income (instead of aggregate multipliers) and a tax calculator to more accurately compute marginal tax rates (in place of average marginal tax rates used in previous years). The latest estimate also makes extensive use of newly available TY 2006 NRP data. Future work will focus on modifying the DCE technique to use pooled NRP sample data for 3 or more years and to apply the modified DCE methodology to produce updated estimates of undetected underreported income.

## References

- B. Erard & Associates, "Development of a Strategy for Estimating the Federal Income Tax Reporting Gap Among Small Corporations," Draft Report for IRS Contract TIRNO-03-P-00429, July 16, 2004.
- B. Erard & Associates, "IRS Tax Gap Estimation: Preliminary Results of Detection Controlled Analysis," PowerPoint presentation to Internal Revenue Service Office of Research, November 1, 2005.
- B. Erard & Associates, "Preliminary Econometric Results," Results summary report submitted to Internal Revenue Service Office of Research, January 27, 2006.
- B. Erard & Associates, "Adjustment of Income Tax Underreporting Using Detection Controlled Estimation," Final Report for IRS Contract Number TIRNO-05-D-00050 0001, November 15, 2007.
- B. Erard & Associates, "Estimates of Income Underreporting Based on Generalized Detection Controlled Estimation Methodology," Task 1 Final Report for IRS Contract TIRNO-10-D-00021 0001, July 22, 2011.
- Brian Erard and Jonathan S. Feinstein, "The Individual Income Reporting Gap: What We See and What We Don't," *IRS Research Bulletin*, Publication 1500 (Rev. 2012).
- Internal Revenue Service, *Income Tax Compliance Research: Estimates for 1973–1981, Appendix B*. Department of Treasury, 1983.
- Internal Revenue Service, *Income Tax Compliance Research: Gross Tax Gap Estimates and Projections for 1973–1992*, IRS Publication 7285 (3–88), Washington, D.C., 1988.
- Jonathan S. Feinstein, "Detection Controlled Estimation," *Journal of Law and Economics*, 33(1):233–276, 1990.
- Jonathan S. Feinstein, "An Econometric Analysis of Income Tax Evasion and its Detection," *Rand Journal of Economics*, 22(1):14–35, 1991.

## Endnotes

- <sup>1</sup> Estimates of the amount of underreported income and the percentage of underreported income reflect underreported income net of overreported income. Individual income taxes reported (and underreported) on returns filed after the applicable filing deadline, including any valid extensions, are part of the individual income tax nonfiling gap.
- <sup>2</sup> Income subject to substantial information reporting (but not withholding) includes dividend income, interest income, pensions and annuities, social security benefits, unemployment insurance, and state income tax refunds. Income subject to little or no information reporting includes nonfarm sole proprietor income reported on Schedule C, farm income reported on Schedule F, rental and royalty income reported on Schedule E, Form 4797 income, and income reported on the “other income” Form 1040 line.
- <sup>3</sup> NRP conducts more than just individual reporting compliance studies. It should be assumed for the remainder of this paper that references to an NRP study refers to an individual reporting compliance study unless explicitly stated otherwise.
- <sup>4</sup> The TY 2001 individual reporting compliance study consisted of returns with tax periods ending between July 2001 and June 2002, the overwhelming majority of which ended on December 31, 2001, and were filed in early 2002.
- <sup>5</sup> Examples of issues include line items on the return, filing status, number of dependents, whether an activity is engaged in for profit or as a hobby.
- <sup>6</sup> Research from the TY 2001 NRP study improved the targeting of audits towards taxpayers most likely to have compliance issues. It also inspired legislation that increased information reporting requirements, such as the reporting of basis for stocks and the reporting of merchant card and third-party payments.
- <sup>7</sup> See Internal Revenue Service (1983) and Internal Revenue Service (1988) for a discussion of the 1976 Information Return Program document matching study and the derivation of the multipliers applied to TCMP audit results.
- <sup>8</sup> See Feinstein (1990, 1991).
- <sup>9</sup> See B. Erard & Associates (2005, 2006, 2007, and 2011) and Erard and Feinstein (2012).
- <sup>10</sup> Detection rate here is defined as the amount of unreported income detected as a percentage of the total unreported income. The smaller the detection rate, the larger the amount of total underreporting is relative to detected underreporting.
- <sup>11</sup> The only income item for which there was insufficient data to use deciles was alimony income. In this case, a single p-tile was used for returns with reported alimony income and a separate p-tile for returns with no reported alimony income.
- <sup>12</sup> See <http://www.irs.gov/pub/irs-soi/09resconawardscorp.pdf>. Net Misreporting Percentage (NMP) for income is defined as the sum of the net misreported amount of income divided by the sum of the absolute values of the amounts of income that should have been reported.
- <sup>13</sup> Average marginal tax curves were used for the TY 2001 individual underreporting gap estimates released in June 2006. Because the prior DCE methodology included capital gains and dividends with other line items and relied on multipliers, the average marginal tax curves were an appropriate approach. However, they likely overstated the marginal tax rate on capital gains.
- <sup>14</sup> In addition to misreported eligibility criteria, the misreporting of income and tax often results in the misreporting of credits. Therefore the total misreporting of credits was calculated with undetected income included in the calculation of total income.