Preliminary Update of the Tax Year 2001 Individual Income Tax Underreporting Gap Estimates

Alan Plumley, Internal Revenue Service

Background

At a July 2004 hearing before the Senate Finance Committee, Internal Revenue Service (IRS) Commissioner Mark Everson committed to releasing by March 31, 2005, an updated estimate of the individual income tax underreporting gap based on the random audit data being compiled by IRS's National Research Program (NRP). At that time, the latest estimates of the tax gap were projections to Tax Year 2001 from compliance data related to Tax Year 1988 and earlier. After 13 years without any new comprehensive tax compliance data, much anticipation had built concerning what the Tax Year 2001 NRP data (which were to be ready for analysis in January 2005) would tell us about any possible changes in taxpayers' compliance behavior, and the estimates scheduled for release by March 31, 2005, would be our first look at the answer to that question.

The gross tax gap is the difference between the aggregate tax liability imposed by law for a given tax year and the amount of tax that taxpayers pay voluntarily and timely for that year. It can be divided into three major components: (1) the nonfiling gap, which is the amount of tax liability that is not paid on time by those who do not file a required return on time; (2) the underreporting gap, which is the amount of tax understated on timely filed returns (net of overstatements); and (3) the underpayment gap, which is the amount of tax that is reported on timely returns but is not paid on time. The net tax gap is the amount of the gross tax gap that remains unpaid after all enforced and other late payments are made for the tax year in question.

The Tax Year 2001 NRP study focused only on the individual income tax underreporting gap. It consisted of a formal review of a stratified random sample of about 46,000 individual income tax returns. All of the returns were given a manual review by experienced auditors, who classified the returns into three general categories as to how to proceed with them: (1) accept the return as filed, given completely corroborating information reported by third parties, or gathered from other sources; (2) contact the taxpayer by correspondence

to investigate up to three items on the return that could not be corroborated; or (3) conduct an in-person audit of the return.

The primary compliance measure derived from the raw results of these contacts is the Voluntary Reporting Rate (VRR), which is the aggregate amount of tax originally reported on the returns in the sample (projected to the entire population) divided by the corresponding amount that should have been reported according to the NRP classifiers and auditors. As summarized in Table 1, the VRR based on the Tax Year 2001 NRP study was 92.4 percent—slightly lower than for the most recent prior study (1988), but well within the historical range. However, that traditional definition of overall tax reporting compliance fails to reflect any misreporting associated with refundable credits (the Earned Income Tax Credit and the Additional Child Tax Credit, which are treated as payments on the tax return and are therefore subtracted from the total tax figure used in the VRR calculation. Refundable credits were a more significant factor in 2001 than they were in 1988; the Additional Child Tax Credit did not exist in 1988, and the EITC has expanded greatly. It is possible to modify the VRR to reflect refundable credits, but that requires a detailed case-by-case analysis of both the 2001 and the 1988 data, rather than the aggregate analysis performed at this stage; so, those estimates will be released later.

Table 1.--Raw Voluntary Reporting Rate for Selected Years

Source and Tax Year	Voluntary Reporting Rate (VRR)
TCMP 1979	91.6%
TCMP 1982	92.7%
TCMP 1985	92.0%
TCMP 1988	93.2%
NRP 2001 *	92.4%

* = Preliminary.
TCMP = Taxpayer Compliance Measurement Program.
NRP = National Research Program.

Like the Taxpayer Compliance Measurement Program (TCMP) studies of the past, however, the NRP audits likely did not account well for three types of noncompliance: (1) underreported informal supplier income;[1] (2) underreported tip income; and (3) underreported income in general that is not subject to third-party information reporting. Prior tax gap studies supplemented the random audit data to account for these deficiencies. The extent of the informal sector of the economy was estimated from customized diary-based surveys of consumers, and separate tip studies provided the basis for estimating tips that were not detected in the compliance audits. To estimate

the extent to which the compliance audits did not fully detect underreporting of other types of income, multipliers were applied to most audit adjustments that had been detected without the aid of third-party information documents. These multipliers were based on a study of Tax Year 1976 TCMP audits that had been conducted without the benefit of such information documents. These three supplements to the raw compliance data were combined with the data in a simulation/calculation model, with which the final tax gap estimates were produced for the year of the audits and for several years beyond that. These estimates were typically published around 2 years after the audit data became available for analysis.

Similar analyses are being conducted to supplement the raw NRP data, but the methodologies being employed have changed. The current plan is to estimate the extent of underreported informal supplier and tip income using existing Census and consumer expenditure data, and the extent to which auditors and classifiers generally do not detect all underreported income is being estimated by a statistical methodology called "Detection-Controlled Estimation," which takes into account the variation in auditors' and classifiers' ability to detect underreported income of each type separately. These analyses will be combined with the raw NRP data in a microsimulation model to produce final, detailed tax gap estimates by the end of 2005.

Since this detailed analysis would clearly not be completed by the March 31 deadline, IRS had to develop a way to produce preliminary estimates based on the new NRP data. This paper outlines that methodology and explains the preliminary estimates that it produced.

Methodology for Developing the Preliminary Estimates

The preliminary approach had to incorporate a quick way to supplement the raw NRP data to account for underreported informal supplier and tip income, as well as for the expected incomplete detection of other types of underreported income in the NRP study. Fortunately, previous tax gap studies (based on TCMP audits for Tax Years 1982, 1985, and 1988) produced aggregate estimates of two key amounts for each major source of income and each category of offset (e.g., adjustments, deductions, exemptions, and credits). These two amounts were: (1) the raw TCMP net misreported amount (e.g., understated dividend income net of overstated dividend income) weighted to the population; and (2) the corresponding expanded net misreported amount that reflected all of the supplemental analyses to account for what was not detected by the TCMP audits. The basic approach of the preliminary NRP-based estimation was to apply the ratio of (2)/(1) to the raw NRP misreported

amounts for each of the same income and offset categories. This applied to the new data in the aggregate an average effective multiplier based on the earlier studies in lieu of the more detailed case-by-case expansion that had been conducted in the past, and will be conducted for the NRP data in due course.

Following that general approach, we derived average effective multipliers for each major tax return line item for each of the three most recent TCMP tax years (1982, 1985, and 1988). These multipliers varied somewhat over this period, and the variation was not the same for all line items. Given the uncertainty about which (if any) of these multipliers might be most applicable to Tax Year 2001, we used both the lowest and the highest ratio for each line item to derive two sets of tax gap estimates: a low estimate and a high estimate. The low estimate applied the lowest effective multiplier for each line item (regardless of which year had the lowest) to the raw aggregate net misreported amount from the NRP data, and the high estimate applied the highest effective multiplier for each line item. This produced low and high expanded NRP estimates of the net misreported amount for each line item (e.g., dividend income or itemized deductions). We estimated the tax impact of each of these aggregate misreported amounts by multiplying them by an average marginal tax rate specific to each line item. We estimated these from the 2001 Statistics of Income sample for individual income tax returns (assuming that everyone who reports an income or an offset item underreports it by the same percentage as everyone else). The total preliminary (low and high) estimates of the underreporting gap, then, are the sum of the corresponding estimates of the tax impact across all line items (income and offsets). Table 2 illustrates how this methodology was applied to the Schedule C net income line item.

As we pursued this approach, however, it became clear that the raw NRP data (much like the raw TCMP data in the past) contained a variety of "anomalies" that, if not accounted for, would misrepresent the magnitude of the tax gap.[2] Some of these anomalies would have a significant impact on our preliminary methodology in particular, even if they had little or no impact on the detailed analysis that would be conducted later. That is because the preliminary analysis was an aggregate analysis conducted on each major return line item separately. One of the most common anomalies (including in TCMP studies of the past) occurred when a taxpayer reported an income or offset item on the wrong line of his or her tax return. This "misclassification" might have no tax impact (e.g., cash wages reported as "other income" rather than as wage income), but if the auditor corrected the return, the NRP results could contain offsetting adjustments as the amount was shifted from one line (which would show a decrease in the amount that should have been reported) to another line (which would show an increase in the amount that should have

Table 2.--Example of Preliminary Tax Year 2001 Underreporting Gap Estimation Based on the NRP Individual Income Tax Reporting Compliance Study: Non-Farm Proprietor Income
(Money amounts are in billions of dollars)

		Calculation	Tax Year and Data Source				
			1982 TCMP	1985 TCMP	1988 TCMP	2001 Preliminary NRP	
						Low [3] Estimate	High [3] Estimate
Raw TCMP / NRP Results							
A	Amount Reported		50.0	79.7	118.5	216.3	216.3
B	Amount Should Have Reported		74.0	113.1	157.6	311.6	311.6
C	Raw Net Misreported Amount	B-A	23.9	33.4	39.0	95.3	95.3
D	Correction for Anomalies					2.1	2.1
E	Corrected Net Misreported Amount	C-D				93.2	93.2
Final (Detailed) Tax Gap Estimates [1]							
F	Adjusted Net Misreported Amount [Accounts for informal suppliers and nondetection]		71.6	90.6	109.9		
G	Tax Gap Associated With Line Item		17.9	24.0	25.2		
Preliminary Tax Gap Estimation							
H	Effective Net Misreported Amount Multiplier	F/C	2.99	2.71	2.81	2.7	2.9
I	Adjusted Net Misreported Amount [Accounts for informal suppliers and nondetection]	E*H				252.5	278.8
J	Assumed Average Marginal Tax Rate [2]					23.3%	23.3%
Preliminary Line Item Tax Gap Estimate		**I*J**				**59**	**65**

[1] Final (detailed) tax gap estimates go beyond the raw compliance data to reflect both informal supplier and tip income, as well as other unreported income that was not detected by the compliance audits.

[2] Average marginal tax rate as reported on returns (SOI). Using this for additional amounts that should have been reported implies that all taxpayers understated their incomes by the same percentage (i.e., a uniform Net Misreporting Percentage for each income type).

[3] The Low Estimate uses the lowest effective multiplier from prior years, while the High Estimate uses the highest effective multiplier from prior years. Both methods use the average marginal tax rate from SOI for 2001. Note that the Low and High Estimates are not necessarily lower and upper bounds on the true gap.

been reported). Likewise, the misclassification could have some tax impact, but not as much as the separate adjustments would imply. For example, if a taxpayer reported business income as wages, the NRP business income (Schedule C) line item would suggest that he or she underreported business income and therefore contributed to the tax gap. In reality, this understatement would generally be offset by the overstatement in wages. However, the business income may have legitimate business expenses that had not been claimed but would partially offset the income, and the net business income would likely be subject to self-employment tax. Although compliance with the self-employment tax may be treated separately from the individual income tax, half of any increase in the amount of self-employment tax would be deductible as a statu-

tory adjustment on the individual income tax return, thus reducing the overall income tax. The point is that aggregating all NRP changes to a given line item (as was done for the preliminary analysis outlined above) could misrepresent the impact (if any) of such misclassifications on the tax gap.

An anomaly that was unique to NRP was the change of a taxpayer's filing status to Married-Joint. In such cases, there may have been no misreporting on either the NRP return or the spouse's return, but the differences between the "corrected" combined return and the original NRP return could mistakenly be treated as underreporting and included in the sum of all NRP adjustments for a given return line item. For example, if a return selected for the study was filed as a Married-Separate return, but in the course of the NRP audit it became clear that it would have been to the taxpayer's advantage to file a joint return, the NRP "per return" data would include just the income on the return of the one spouse who was selected for the sample, but the "per exam" data would include the income of both spouses. The difference between these two income amounts would normally be interpreted as underreporting, but that would not necessarily be true in this case. Moreover, if the selected spouse's return reported very modest income (e.g., mostly wages) it would typically have a very high sample weight (i.e., it would represent thousands of such returns in the general population). Then, if the spouse had a very large amount of such income, the "change" in income, when multiplied by the large sample weight, would appear to represent billions of dollars in underreported income, when in fact the difference might not be attributable to noncompliance at all. These types of cases were excluded from TCMP studies of the past, but they were deliberately not excluded from NRP so that we could build a more accurate picture of the entire taxpayer population. However, that commitment requires additional work to ensure that such cases do not cause us to make incorrect inferences from the data.

One final example will illustrate the kinds of "anomalies" that could easily distort our preliminary estimates of the tax gap based on the aggregate amount misreported on any given return line item taken in isolation. Let's say a taxpayer did not report a special source of income (e.g., a cash prize) because he donated the entire amount to a charity (which he also did not report, thinking it completely offset the income that he did not report). Even if the full deduction for the donation could be claimed so that there was no tax impact, the preliminary (i.e., aggregate) tax gap estimating methodology would have assumed that all of the underreported income increased the tax gap and that all of the unreported deduction decreased the tax gap, giving a misleading view of noncompliance for each of those two line items.

To correct our aggregations of the raw NRP data for the impact of these and other types of anomalies, we reviewed the cases that had the largest

weighted adjustments on any given return line item to determine how much of the adjustment would not affect the tax gap. We then corrected (usually lowered) the raw NRP weighted net adjustments accordingly prior to applying the preliminary tax gap estimation methodology outlined above. These corrections reduced the tax gap estimates by 7.7 billion dollars when projected to the entire population.

The Preliminary Estimates

Table 2 illustrates the method used to estimate the underreporting gap associated with a given return line item (in this case, non-farm proprietor [Schedule C] net income). We followed this approach for each category of income and offsets separately. The detailed results are summarized in Table 3. The preliminary estimates of the overall individual income tax underreporting gap range from $150 billion to $187 billion, which is somewhat more than we had projected from the 1988 estimates. As with prior estimates of the individual income tax underreporting gap, over half is attributable to the underreporting of business income. This reflects a greater opportunity to underreport business income due to the lack of third-party information reporting or withholding. In general, although these preliminary estimates are likely to change as a result of the detailed tax gap analysis, the results are quite consistent with prior estimates.

The preliminary estimates of the individual income tax underreporting gap suggest that it may be somewhat larger than we had projected from earlier compliance data (see Table 4). However, a closer look at the estimates reveals a possible explanation. As shown in Table 4, the total amount of income reported on individual income tax returns roughly doubled between 1988 and 2001. Wage and salary income almost doubled as well. However, business income almost tripled—a growth that was driven almost entirely by the growth in flowthrough (partnership and S corporation) income, which more than quadrupled. This is consistent with the growth in that portion of the underreporting gap beyond what we had projected for business income overall based on earlier data. Thus, it appears that any decline in reporting compliance since 1988 is due to the different "mix" of income sources, even if each type of income is reported with the same degree of accuracy as in the past. As more and more taxpayers receive a growing share of their income from sources not subject to information reporting or withholding, we should expect the overall underreporting gap to increase faster than the growth in receipts overall.

These preliminary estimates of the individual income tax underreporting gap are combined with projections of other components of the Tax Year 2001

Table 3.--Preliminary[1] Estimates of the Individual Income Tax Underreporting Gap, Tax Year 2001
(Money amounts are in billions of dollars)

Line Item	Preliminary Low [2] Estimates	Preliminary High [2] Estimates
Total Individual Income Tax Underreporting Gap	**150**	**187**
Non-Business Income	**42**	**57**
Wages, salaries, tips, etc.	13	15
Interest income	1	1
Taxable dividends	2	4
State income tax refunds	*	*
Alimony income	*	*
Taxable pension, annuities, IRA distributions	4	8
Unemployment compensation	*	1
Taxable Social Security benefits	1	1
Schedule D - net capital gains & distributions	3	6
Form 4797 - other gains net income	2	3
Other income	14	18
Business Income	**83**	**99**
Sch C - Sole Proprietor net income	59	65
Sch E - Rent & Royalty net income	7	8
Sch E - Partnership, S-Corp, Estate & Trust net inc.	16	24
Schedule F - Farm net income	2	3
Offsets to Income	**14**	**17**
Adjustments: 1/2 of Self-Employment Tax [3]	-6	-7
Adjustments: All Other	*	1
Deductions	15	18
Exemptions	5	5
TOTAL Credits	**11**	**14**
Net Math Error (other than EITC)	*****	*****

Note: Amounts may not add to the totals because of rounding.

* Less than $0.5 billion

[1] These line-item estimates are particularly sensitive to data anomalies and to the uncertainties inherent in the preliminary methodology. The final estimates may be much different. The totals and subtotals (in bold) are likely to be stronger estimates than the detailed estimates.

[2] These are not necessarily lower and upper bounds on the true tax gap.

[3] These estimates are negative because when business income is understated, the corresponding self-employment tax is also understated—half of which could have been claimed as a statutory adjustment, which would have reduced income tax liability.

Table 4.--Growth in the Individual Income Tax Gap Arising From Underreported Business Income, Tax Years 1988 to 2001
(Money amounts are in billions of dollars)

Line Item	TY2001 Tax Gap Estimates			Net Income Reported		
	Preliminary Low Estimates	Preliminary High Estimates	Previous[1] Projections	TY1988	TY2001	Ratio of TY2001 to TY1988
Total Individual Income	**150**	**187**	**149**	**3,230**	**6,416**	**1.99**
Wage, salary, tip income				**2,238**	**4,407**	**1.97**
Business Net Income	**83**	**99**	**81**	**184**	**534**	**2.90**
Nonfarm Proprietor net income	59	65	59	119	216	1.82
Rent and Royalty net income	7	8	8	-3	28	
Partnership, S corporation, Estate, and Trust net income	16	24	7	72	299	4.15
Schedule F--Farm net income	2	3	7	-3	-9	

[1] Projected to 2001 from estimates made for Tax Years 1988 and 1992, assuming constant compliance rates for each major component of the gap (e.g., business income overall), and that the growth in the gap was due entirely to the growth in actual receipts (which was driven by the growth in incomes and, therefore, in aggregate tax liability).

gross tax gap in Table 5. (These other projections are based on estimates of the gap made for prior tax years, which were expanded in proportion to the growth in actual tax receipts in each major category.) Table 5 highlights the fact that we do not yet have estimates for several components of the tax gap, and that our estimates for some other key components (e.g., individual income tax nonfiling and corporation underreporting) are based on very old data.

Conclusions

The preliminary estimates of the individual income tax underreporting gap suggest that it may be somewhat larger than we had projected from earlier compliance data. However, the aggregate analysis conducted so far suggests that this may be driven by a change in the mix of income sources, rather than a decline in the propensity of taxpayers to report a given type of income. Income from flowthrough businesses such as partnerships and S corporations (which has historically been reported less accurately than wages) has grown much faster than other sources of income, presumably because more

Table 5.--Preliminary Estimates[1] of the Tax Year 2001 Gross Tax Gap by Type of Tax and by Type of Noncompliance
(Money amounts are in billions of dollars)

Type of Tax	Nonfiling Gap	Under-reporting Gap	Under-payment Gap [2]	Total Tax Gap	Non-compliance Rate (percent) [3]
Individual Income Tax	28	**150 - 187**	19	**198 - 234**	**17.5-20.1**
Corporation Income Tax	*	30	2	32	18.4
Employment Tax [4]	*	**66 - 71**	7	**73 - 78**	**9.9-10.4**
Estate and Gift Tax	2	4	2	8	23.2
Excise Tax	*	*	1	1	*
TOTAL	30	**251 - 291**	32	**312 - 353**	**15.0-16.6**

Note: Amounts may not add to the totals because of rounding.

* No estimates available.

[1] Amounts in **bold** are preliminary updates based on the 2001 NRP individual income tax underreporting study. The ranges arise from different sets of assumptions, and are not necessarily upper and lower bounds. Unbolded portions of the underreporting and nonfiling gaps were projected to 2001 from compliance data related to Tax Year 1988 and earlier.

[2] Actual amounts tabulated from IRS records; no estimates necessary.

[3] Gross tax gap expressed as a percentage of total true tax liability for a given tax year.

[4] The underreporting gap estimates include preliminary estimates associated with Self-Employment Tax, which were derived from the preliminary estimates of the individual income tax underreporting gap arising from understated self-employment income.

people have invested more money in these opportunities than past trends would have led us to expect. Fortunately, the National Research Program's next reporting compliance study will focus on flowthrough entities and is set to begin later this year.

Overall, the preliminary analysis of the new NRP data on reporting compliance suggests that voluntary compliance among individuals remains fairly strong. We estimate that individuals pay at least 80 percent of their tax liability voluntarily and on time, much the same as we had estimated from previous compliance studies. Individual income tax still seems to be the largest component of the total tax gap (around two-thirds), but this is primarily because

individuals bear the largest share of the overall Federal tax burden, and because much individual income is not subject to third-party information reporting or other public incentives for accurate tax reporting. These preliminary estimates will be updated at the end of this year when detailed estimates are completed. The detailed estimates will supplement the raw NRP results to account for underreported informal supplier income, underreported tip income, and the extent to which IRS auditors do not fully detect other types of unreported income.

Endnotes

[1] Informal suppliers are sole proprietors who operate in an informal business style, often on a cash basis, with poor or nonexistent books and records—including moonlighters, street vendors, roadside stands, door-to-door salesmen, etc.

[2] Some of these data issues are really not "anomalies" in the sense of being unexpected departures from the norm, but they all have in common the fact that they introduce into the raw audit results changes to what was originally reported that represent much less noncompliance than the raw results would otherwise suggest, and often have no impact on the tax gap at all.