Complex Survey Sample Design in IRS' Multi-objective Taxpayer Compliance Burden Studies

John Guyton
Wei Liu
Michael Sebastiani
Internal Revenue Service, Office of Research, Analysis & Statistics
1111 Constitution Ave. NW, Washington, DC 20224[1]

Abstract

The Internal Revenue Service periodically conducts complex surveys to measure the pre-filing and filing burden of individual taxpayers in response to the requirements of the U.S. federal tax system. The sample design for the survey needs to balance three major objectives. The first is to ensure a sufficient number of respondents within and across strata to meet the needs of the modeling of compliance burden. The second is that it must be efficient so that the estimates are reliable. The third is to facilitate the comparisons between the current year study and the previous study. An iterative procedure for a stratified random sample design is proposed to search for the optimal sample allocation. The proposed procedure utilizes the optimality in the Neyman allocation method, and incorporates the minimum sample size requirements for modeling and different nonresponse rates across strata. Our adjustment on the Neyman allocation causes loss of efficiency for descriptive analysis, but such loss of efficiency is minimized so that the estimates still meet the precision requirements. Furthermore, such loss is well compensated by the gains in modeling and analytical capabilities.

Key words: Complex survey sample design, Neyman allocation, Nonresponse rate, Sample size, Optimality, Sensitivity analysis

1. Introduction

The Internal Revenue Service periodically conducts complex surveys to measure the pre-filing and filing burden of individual taxpayers in response to the requirements of the U.S. federal tax system. One of the challenges of this type of research is incorporating what one learns from one study in the design of the subsequent study while maintaining comparability between the studies. Our sample design specifications are developed to balance three issues. The first and most important is to ensure that there are a sufficient number of cases to meet the needs of the modeling tool to identify the determinants of burden, both within and across stratum. The second is that it must be efficient in the way the sample is distributed so that estimates from the sample are reliable (i.e., meet confidence interval range requirements). The third is that the design should facilitate the comparisons between the studies. This paper discusses our approach to the sample design for Individual Taxpayer Burden (ITB) TY2010[2] survey embedding the previous TY2007 survey sample design for comparability.

---

[1] The views expressed are those of the authors and not the official positions of the Internal Revenue Service.
[2] TY2010 refers to tax year 2010. We will use this notation throughout the paper.

An iterative procedure for a stratified random sample design is proposed to search for the optimal sample allocation. The proposed procedure utilizes the optimality in the Neyman allocation method, and incorporates the minimum sample size requirements for modeling and different nonresponse rates across strata. Our adjustment on the Neyman allocation causes loss of efficiency for descriptive analysis, but such loss of efficiency is minimized so that the estimates still meet the precision requirements. Furthermore, such loss is well compensated by the gains in modeling and analytical capabilities.

To make the study of ITB TY2010 survey comparable with the one of TY2007 survey, we continue to use the same design variable, total monetized burden, the same stratified random sampling approach, and the same stratification variables that were used in the TY 2007 study.[3] The Neyman allocation method was used to determine the optimal sample size for each stratum subject to the total sample size of 15,000 in the sample design for the ITB TY2007 survey. It aimed to minimize the variance of the estimated mean burden, but it left several strata with too few observations to model. A common rule of thumb is that a sample must include at least 10 or 15 observations per independent variable in a regression model (Stevens, 2002; Bartlett et al., 2001). We choose 15 to be conservative. The expected number of independent variables is 15, so the minimum expected number of complete responses for modeling is 225 for each stratum.

Our objective is to minimize the variance of the estimated mean burden constrained on this minimum expected number of complete responses for modeling, with response rate incorporated. The total sample size increases to 20,000 in the ITB TY2010 survey. We start with the same total sample size of 15,000 as in TY2007, considering this as our base sample. The remaining 5,000 is the reserved sample to make any adjustments for the purposes of modeling and increasing the precision of the estimate. An iterative procedure is used to search for our optimal sample allocation. The final allocation with key inputs is shown in Table 1. This design results in an overall CV of 1.01%, less than 2%, the general requirement on precision.

[3] This approach is discussed in further detail in Brick, *et al*, 2009 and Contos, *et al*, 2010.

Table 1 Sample allocation for ITB TY2010 survey

Monetized Burden Strata	Projected Pop Count	Est. Mean	Est. Std. Dev.	Est. Response Rate	Final Allo-cation
11 paid, low	9,822,075	190.46	241.53	0.2558	**880**
12 paid, low-medium	26,114,402	295.10	370.49	0.3213	**1,644**
13 paid, medium	15,940,360	619.92	980.87	0.3916	**2,656**
14 aid, medium-high	15,732,824	946.43	1,157.12	0.3970	**3,092**
15 paid, high	10,685,596	1,837.13	2,524.26	0.3894	**4,582**
21 self, low	3,503,015	85.97	115.25	0.3594	**626**
22 self, low-medium	2,707,918	157.75	225.08	0.3436	**655**
23 self, medium	1,695,808	499.83	709.51	0.4355	**517**
24 self, medium-high	770,422	715.88	876.97	0.4046	**556**
25 self, high	288,597	923.48	881.83	0.4119	**546**
31 soft, low	10,478,344	116.18	159.24	0.3058	**736**
32 soft, low-medium	15,971,640	185.25	228.28	0.3678	**619**
33 soft, medium	10,942,941	518.45	713.67	0.4620	**1,327**
34 soft, medium-high	6,336,666	769.97	1,015.50	0.4396	**1,093**
35 soft, high	1,639,707	1,278.71	1,615.97	0.4772	**472**
Total	132,630,316	551.90			**20,000**
Overall CV					**1.01%**

2. Iterative Procedure for Sample Allocation

TY2010 population counts within each stratum are projected based on TY2009 population counts and the projected growth rates. TY2007 survey data are used to estimate the mean, standard deviation, and item response rate of monetized burden for each stratum. With these inputs, we are able to determine the sample allocation as follows.

2.1 Step 1

The adjustment procedure in Step 1 is shown in Table 2. The Neyman allocation method is first applied to the base sample of 15,000, the sample size of the TY2007 ITB survey sample. The expected numbers of respondents are obtained for each stratum according to the estimated item response rates[4]. We compare these numbers with 225, the minimum sample required to model each stratum. Nine strata are identified that do not contain enough respondents. The sample sizes for these strata are adjusted so that the expected numbers of respondents that are highlighted in Table 2 equal 225, while the sample sizes for the remaining strata maintain the Neyman allocation. This procedure results in a total sample size of 18,540, indicating that we have more sample to be allocated to the base sample to increase the precision of the estimate.

[4] The estimated item response rates are the lower bounds of the 99% Wilson Score confidence intervals in TY2007, adjusted for the overall expected lower bound of the response rate at 40% in TY2010.

Table 2: Adjusted Allocation Step 1: Adjustment of Neyman Allocation for Modeling

Monetized Burden Number of Strata	Neyman Allocation	Est. Item Response Rate	Expected Number of Respondents	Minimum Sample Size for Modeling	Adjusted Allocation: Step 1
11 paid, low	**362**	**0.2558**	**93**	**225**	**880**
12 paid, low-medium	1,478	0.3213	475	225	1,478
13 paid, medium	2,388	0.3916	935	225	2,388
14 paid, medium-high	2,780	0.3970	1,104	225	2,780
15 paid, high	4,119	0.3894	1,604	225	4,119
21 self, low	**62**	**0.3594**	**22**	**225**	**626**
22 self, low-medium	**93**	**0.3436**	**32**	**225**	**655**
23 self, medium	**184**	**0.4355**	**80**	**225**	**517**
24 self, medium-high	**103**	**0.4046**	**42**	**225**	**556**
25 self, high	**39**	**0.4119**	**16**	**225**	**546**
31 soft, low	**255**	**0.3058**	**78**	**225**	**736**
32 soft, low-medium	**557**	**0.3678**	**205**	**225**	**612**
33 soft, medium	1,193	0.4620	551	225	1,193
34 soft, medium-high	983	0.4396	432	225	983
35 soft, high	**405**	**0.4772**	**193**	**225**	**472**
Total	15,000	0.4003	5,861	3,375	18,540

2.2 Step 2

In Step 2, we determine the maximum sample size for base sample for the Neyman allocation, given the nine strata identified for further adjustment in the Step 1. The following inequality (1) is used to find the maximum sample size for base sample.

$$n_b \frac{\sum_{h \in A} N_h S_h}{\sum_{h \in H} N_h S_h} + (20{,}000 - n_b) \geq 5{,}600, \qquad (1)$$

Where n_b is the total base sample size, N_h and S_h are the population count and standard deviation for stratum h, $A = strata\{11,21,22,23,24,25,31,32,35\}$, H is the entire set of all strata. The minimum total sample size is 5,600 for all nine strata identified in Step 1 (the sum of the highlighted numbers in the last column of Table 2).

The first component on the left-hand side of inequality (1), $n_b \frac{\sum_{h\in A} N_h S_h}{\sum_{h\in H} N_h S_h}$, represents the sum of the sample sizes for all nine strata in A from the base sample, due to the Neyman allocation. The second component, $20{,}000 - n_b$, represents the sum of the sample sizes for all nine strata from the reserved sample to adjust for modeling. We solve Inequality (1) for n_b and $n_b \le 16{,}693$.

The adjustment procedure in Step 2 is shown in Table 3. The Neyman allocation for the base sample of size 16,693 is calculated, and then the adjustment procedure from Step 1 is repeated. Step 2 results in a total sample size of 20,008. This is because the number of strata to be adjusted is reduced from nine to eight since the base sample size increases in Step 2. Specifically, the sample size in stratum 32 that results from Neyman allocation becomes sufficient for modeling and does not need to be adjusted. The value for $n_b = 16{,}693$ is obtained assuming that nine strata need to be adjusted; however the result shows only eight strata need to be adjusted, indicating that this “non-matching” or “non-steady” state requires another iteration.

Table 3: Adjusted Allocation Step 2

Monetized Burden Number of Strata	Neyman Allocation	Estimated Response Rate	Expected Number of Respondents	Minimum Sample Size for Modeling	Adjusted Allocation: Step 2
11 paid, low	**403**	**0.2558**	**103**	**225**	**880**
12 paid, low-medium	1,644	0.3213	528	225	1,644
13 paid, medium	2,657	0.3916	1,041	225	2,657
14 paid, medium-high	3,094	0.3970	1,228	225	3,094
15 paid, high	4,584	0.3894	1,785	225	4,584
21 self, low	**69**	**0.3594**	**25**	**225**	**626**
22 self, low-medium	**104**	**0.3436**	**36**	**225**	**655**
23 self, medium	**204**	**0.4355**	**89**	**225**	**517**
24 self, medium-high	**115**	**0.4046**	**46**	**225**	**556**
25 self, high	**43**	**0.4119**	**18**	**225**	**546**
31 soft, low	**284**	**0.3058**	**87**	**225**	**736**
32 soft, low-medium	*620*	*0.3678*	*228*	*225*	*620*
33 soft, medium	1,327	0.4620	613	225	1,327
34 soft, medium-high	1,094	0.4396	481	225	1,094
35 soft, high	**450**	**0.4772**	**215**	**225**	**472**
Total	16,693	0.4003	6,523	3375	20,008

2.3 Step 3

In Step 3, we adjust the sample size for base sample so that the final total sample size reaches exactly 20,000, given the eight strata identified for adjustment in Step 2. Inequality (2) is used to find the sample size for base sample.

$$n_b \frac{\sum_{h\in B} N_h S_h}{\sum_{h\in H} N_h S_h} + (20{,}000 - n_b) \geq 4{,}988\,, \qquad (2)$$

Where, $\mathrm{B} = strata\{11,21,22,23,24,25,31,32\}$. 4,988 is the minimum total sample size for all eight strata identified in Step 2--the sum of the highlighted numbers in the last column of Table 3. We solve Inequality (2) for n_b and $n_b \leq 16{,}684$. We start with the base sample of 16,684 for the Neyman Allocation, and repeat the same adjustment procedure as in Step 2. The number of strata requiring adjustment remains the same in this round, implying that we have reached a “steady state” and the resulting total sample size reaches the exact 20,000 as required. The adjustment procedure is shown in Table 4.

Table 4: Adjusted Allocation Step 3: Final Allocation.

Monetized Burden Number of Strata	Neyman Allocation	Estimated Response Rate	Expected Number of Respondents	Minimum Sample Size for Modeling	Adjusted Allocation: Step 2
11 paid, low	**403**	**0.2558**	**103**	**225**	**880**
12 paid, low-medium	1,644	0.3213	528	225	1,644
13 paid, medium	2,656	0.3916	1,040	225	2,656
14 paid, medium-high	3,092	0.3970	1,228	225	3,092
15 paid, high	4,582	0.3894	1,784	225	4,582
21 self, low	**69**	**0.3594**	**25**	**225**	**626**
22 self, low-medium	**104**	**0.3436**	**36**	**225**	**655**
23 self, medium	**204**	**0.4355**	**89**	**225**	**517**
24 self, medium-high	**115**	**0.4046**	**46**	**225**	**556**
25 self, high	**43**	**0.4119**	**18**	**225**	**546**
31 soft, low	**283**	**0.3058**	**87**	**225**	**736**
32 soft, low-medium	619	0.3678	228	225	619
33 soft, medium	1,327	0.4620	613	225	1,327
34 soft, medium-high	1,093	0.4396	480	225	1,093
35 soft, high	**450**	**0.4772**	**215**	**225**	**472**
Total	16,684	0.4003	6,519	3375	**20000**

In summary, the allocation adjustment procedure described above maximizes the optimality in the Neyman allocation method subject to the minimum sample size necessary for modeling within each stratum, while incorporating response rate. As we can also see, the adjustment for modeling in the above iterative procedure is minimized subject to the total sample size of 20,000. Even though we expect a sacrifice in precision due to adjustment of the Neyman allocation, this sacrifice has been minimized.

3. Evaluation of the Proposed Sample Design

Table 5 presents a comparison between the final design (Design I) and the Neyman allocation of 20,000 without any adjustment (Design II). The expected number of respondents from our design satisfies the minimum sample size for modeling within each stratum, while the expected numbers of respondents for seven strata from the Neyman allocation without any adjustment are far less than the minimum sample size.

The overall CV from the Neyman allocation without any adjustment is 0.95%, while the overall CV from our design is 1.01%. The precision in our design decreases, as expected, which is a trade-off for incorporating modeling and response rate. However, our overall CV is still comparable with the overall CV from the Neyman allocation. Such sacrifice in precision is minor and is well compensated by the gains in modeling. The overall CV based on the expected number of respondents is 1.62% from our design, and again it is comparable with the corresponding CV of 1.53% from the Neyman allocation.

Table 5: Comparisons between Two Designs:
Design I: the proposed design with adjustment for modeling and response rate
Design II: Neyman allocation without any adjustment

Monetized Burden Strata	Estimated Response Rate	Design I	Expected Number of Respondents	Design II	Expected Number of Respondents
11 paid, low	0.2558	880	225	483	**124**
12 paid, low-medium	0.3213	1644	528	1970	633
13 paid, medium	0.3916	2656	1040	3184	1247
14 paid, medium-high	0.3970	3092	1228	3707	1472
15 paid, high	0.3894	4582	1784	5493	2139
21 self, low	0.3594	626	225	82	**30**
22 self, low-medium	0.3436	655	225	124	**43**
23 self, medium	0.4355	517	225	245	**107**
24 self, medium-high	0.4046	556	225	138	**56**
25 self, high	0.4119	546	225	52	**21**
31 soft, low	0.3058	736	225	340	**104**
32 soft, low-medium	0.3678	619	228	742	273
33 soft, medium	0.4620	1327	613	1590	735
34 soft, medium-high	0.4396	1093	480	1310	576
35 soft, high	0.4772	472	225	540	257
Total	0.4003	20000	7701	20000	7815
Overall CV		**1.01%**	**1.62%**	**0.95%**	**1.53%**

The Neyman allocation method assumes population counts and standard deviations are both known for all strata, however in practice they are often estimates. We evaluate the projection method of the TY2010 population by comparing the actual count in TY2009 and the projected count in TY2009 using the same projection method. The average relative error rate among all the strata is 0.0037, supporting our projected counts for TY2010. However, there is not a straightforward measure to see how reliable our estimated standard deviations are. It is of interest to investigate how robust our design is to different values of population standard deviations. We obtain different sets of standard deviation estimates, and assume they represent the true population parameter values in TY2010 and the overall CVs can be calculated based on these estimates and our sample design. Table 6 shows that our proposed sample design can still satisfy the precision requirement even when the population standard deviations are mis-specified. Therefore, our design is robust to different estimates on population standard deviations.

Table 6: Robustness of the proposed sample design to different values of standard deviations

Std. Dev. I: estimates used in our final design
Std. Dev. II: estimates using multiple imputations
Std. Dev. III: estimates including outliers
Std. Dev. IV: estimates excluding the observations above the 99^{th} percentile
Std. Dev. V: estimates from ITB TY99/00 survey

Monetized Burden Strata	Std. Dev. I	Std. Dev. II	Std. Dev. III	Std. Dev. IV	Std. Dev. V
11 paid, low	241.53	240.83	708.43	240.74	384.28
12 paid, low-medium	370.49	369.22	403.19	281.10	569.25
13 paid, medium	980.87	953.64	998.00	615.62	722.5
14 paid, medium-high	1,157.12	1,309.89	1,210.67	930.71	974.14
15 paid, high	2,524.26	5,041.15	4,063.85	2,051.27	2,143.75
21 self, low	115.25	112.86	306.64	208.21	517.76
22 self, low-medium	225.08	225.99	285.97	224.35	507.05
23 self, medium	709.51	680.96	709.51	709.51	337.65
24 self, medium-high	876.97	849.45	868.18	868.18	631.62
25 self, high	881.83	839.23	881.83	881.83	695.25
31 soft, low	159.24	158.26	294.59	206.34	425.51
32 soft, low-medium	228.28	232.43	279.12	211.39	464.06
33 soft, medium	713.67	709.62	713.21	532.83	442.47
34 soft, medium-high	1,015.50	1,001.92	1,014.18	708.57	755.34
35 soft, high	1,615.97	1,842.07	1,611.69	1,198.00	1,015.03
Overall CV	**1.01%**	**1.40%**	**1.29%**	**0.79%**	**1.02%**

Finally, we consider the implications of our proposed design on the other three relevant variables, total money, total monetized time, and total time with respect to both precision of the estimates and modeling. Since the estimated item response rates for these three variables are almost identical to the ones for total monetized burden, immediately it follows that the expected number of respondents from our design will satisfy the minimum sample size requirement for modeling these three variables. Table 7 shows the overall CV for each of the three variables based on our design, compared with the CV from the Neyman allocation for each variable separately.

As shown in Tables 7-1, 7-2, and 7-3, although our design based on the total monetized burden still can satisfy the modeling of all three variables, the sacrifice in precision is greater comparing the Neyman allocation with each variable as the design variable. This indicates that a more advanced algorithm is needed if we want to simultaneously estimate and model total money, total monetized time, and total time. We expect that our proposed design will provide sufficient responses in the various strata to support further refinements in future designs. However, the Neyman allocation without any adjustment for any of the three variables results in modeling issues.

4. Conclusions

An iterative procedure is proposed to search for the optimal sample allocation to optimize the Neyman allocation subject to the minimum sample size for modeling purposes, with response rates incorporated.

On one hand, any adjustment on the Neyman allocation may cause loss of efficiency; on the other hand, the Neyman allocation without any adjustment will leave us with some strata with too few observations to model. Our proposed sample allocation procedure shows that such loss of efficiency can be minimized with appropriate adjustment; moreover, it can be well compensated by the gains from modeling and the ability to conduct valid predictive analysis.

Contemporary surveys often serve multiple purposes—both descriptive and predictive analyses. How to balance between the objectives to achieve a certain level of precision on the estimates and to conduct valid predictive analysis requires more theoretical development. Our future research will extend the approach described here. We are also interested in exploring the optimal allocation algorithm when the objective of the survey is to estimate and model multiple variables simultaneously.

Finally, Neyman allocation is often criticized because it ignores possible different response rates across strata. When response rates are quite different across different strata, the objective function is to minimize the variance of the mean estimate based on the respondents, and the sample size can be calculated from the number of respondents and response rate. We expect that a more rigorous optimization procedure reflecting these response rate differences will further improve the performance of the sample design.

Table 7-1: Implications of the proposed design on total money
Design I: our design with the total monetized burden as the design variable
Design II: Neyman allocation with total money as the design variable

Strata	Estimated Response Rate	Design I	Expected Number of Respondents	Design II	Expected Number of Respondents
11 paid, low	0.2558	880	225	646	**165**
12 paid, low-medium	0.3213	1,644	528	3,109	999
13 paid, medium	0.3916	2,656	1,040	1,655	648
14 paid, medium-high	0.3970	3,092	1,228	4,264	1,692
15 paid, high	0.3894	4,582	1,784	7,423	2,890
21 self, low	0.3594	626	225	59	**21**
22 self, low-medium	0.3436	655	225	44	**15**
23 self, medium	0.4355	517	225	240	**105**
24 self, medium-high	0.4046	556	225	69	**28**
25 self, high	0.4119	546	225	8	**3**
31 soft, low	0.3058	736	225	350	**107**
32 soft, low-medium	0.3678	619	228	469	**173**
33 soft, medium	0.4620	1,327	613	959	443
34 soft, medium-high	0.4396	1,093	480	442	**194**
35 soft, high	0.4772	472	225	263	**125**
Total		20,000	7,701	20,000	7,610
Overall CV for total money		**1.26%**	**2.06%**	**1.08%**	**1.77%**

Table 7-2 Implications of our proposed design on total monetized time
Design I: our design with the total monetized burden as the design variable
Design II: Neyman allocation with total monetized time as the design variable

Strata	Estimated Response Rate	Design I	Expected Number of Respondents	Design II	Expected Number of Respondents
11 paid, low	0.2558	880	225	426	**109**
12 paid, low-medium	0.3213	1,644	528	1,803	579
13 paid, medium	0.3916	2,656	1,040	3,550	1,390
14 paid, medium-high	0.3970	3,092	1,228	3,582	1,422
15 paid, high	0.3894	4,582	1,784	4,810	1,873
21 self, low	0.3594	626	225	90	**32**
22 self, low-medium	0.3436	655	225	139	**48**
23 self, medium	0.4355	517	225	269	**117**
24 self, medium-high	0.4046	556	225	149	**60**
25 self, high	0.4119	546	225	61	**25**
31 soft, low	0.3058	736	225	362	**111**
32 soft, low-medium	0.3678	619	228	825	304
33 soft, medium	0.4620	1327	613	1,811	837
34 soft, medium-high	0.4396	1,093	480	1,508	663
35 soft, high	0.4772	472	225	614	293
Total		20,000	7,701	20,000	7,863
Overall CV for total money		**1.26%**	**2.01%**	**1.18%**	**1.89%**

Table 7-3: Implications of our proposed design on total time
Design I: Our design with the total monetized burden as the design variable
Design II: Neyman allocation with the total time as the design variable

Strata	Estimated Response Rate	Design I	Expected Number of Respondents	Design II	Expected Number of Respondents
11 paid, low	0.2558	880	225	397	**102**
12 paid, low-medium	0.3213	1,644	528	1,736	558
13 paid, medium	0.3916	2,656	1,040	1,565	613
14 paid, medium-high	0.3970	3,092	1,228	8,269	3,283
15 paid, high	0.3894	4,582	1,784	2,464	960
21 self, low	0.3594	626	225	103	**37**
22 self, low-medium	0.3436	655	225	86	**29**
23 self, medium	0.4355	517	225	145	**63**
24 self, medium-high	0.4046	556	225	106	**43**
25 self, high	0.4119	546	225	27	**11**
31 soft, low	0.3058	736	225	526	**161**
32 soft, low-medium	0.3678	619	228	625	230
33 soft, medium	0.4620	1,327	613	890	411
34 soft, medium-high	0.4396	1,093	480	1,920	844
35 soft, high	0.4772	472	225	1,140	544
Total		20,000	7,701	20,000	7,888
Overall CV for total time		**2.43%**	**3.84%**	**1.86%**	**2.98%**

References

Andreoni, James, Brian Erard and Jonathan Feinstein. *Tax Compliance*. Journal of Economic Literature. Vol. XXXVI (June 1998) 818-869.

Brick, Michael, George Contos, Karen Masken, and Roy Nord. *Response Mode and Bias Analysis in the IRS' Individual Taxpayer Burden Survey*, 2009 Joint Statistical Meeting Proceedings, August 2009.

Cochran, William G. Sampling Techniques, *3rd edition.* John Wiley & Sons.

Contos, George, Guyton, John L., Langetieg, Patrick, and Vigil, Melissa. *Individual Taxpayer Compliance Burden: The Role of Assisted Methods in Taxpayer Response to Increasing Complexit*. 2010 IRS Research Conference.

Guyton, John L., John F. O'Hare, Michael P. Stavrianos, and Eric J. Toder. *Estimating the Compliance Cost of the U.S. Individual Income Tax*. National Tax Journal, September 2003, 673-688.