Technical Paper

October 2022

The Individual Income Tax and Self-Employment Tax Nonfiling Tax Gaps for Tax Years 2014-2016

Publication 5785 (10-2022) Catalog Number 93941Y Department of the Treasury **Internal Revenue Service** www.irs.gov

This page intentionally left blank.

The Individual Income Tax and Self-Employment Tax Nonfiling Tax Gaps for Tax Years 2014-2016

Suggested Citation:

Hertz, T.N., Langetieg, P.T., Payne, J.M., and Plumley, A.H.

The Individual Income Tax and Self-Employment Tax Nonfiling Tax Gaps for Tax Years 2014-2016
Research, Applied Analytics & Statistics Technical Paper

Internal Revenue Service
Publication 5785
Washington, DC
October 2022

.

This page intentionally left blank.

Table of Contents

Tables and Figures

This page intentionally left blank.

The Individual Income Tax and Self-Employment Tax Nonfiling Tax Gaps for Tax Years 2014-2016

Executive Summary

Taxpayers are required by the Internal Revenue Code to file income tax returns with the IRS by the established due date, on which they are to report all of their tax liability; it also requires them to pay that tax liability on time. However, not all taxpayers file required tax returns on time (or at all), and some of their tax liability is therefore not paid on time. The nonfiling tax gap[1] is the amount of true tax liability not paid on time by those who do not file on time. Since some nonfilers pay some or all of their true tax liability on time (e.g., through withholding), not all nonfilers actually contribute to the tax gap. Nonetheless, the nonfiling gap is comprised of two major components: the portion associated with those who file late ("late filers"), and the portion associated with those who never file at all ("not-filers"). Thus, from a tax gap perspective, nonfilers include both late filers and not-filers. With the passage of time, some not-filers file late returns, so the distinction between these two groups is merely a pragmatic one for estimating the gap. It is easier to estimate the contribution that late filers make to the nonfiling gap since we have their tax return; it is much harder to estimate the gap associated with those who have not filed any return by the time the estimate is made.

We estimate that the average annual individual income tax nonfiling gap over the TY 2014 through TY 2016 period was $32.6 billion, and the corresponding self-employment tax nonfiling gap was $6.5 billion. Since self-employment tax is to be reported on the same tax return as individual income tax, the methodologies described herein produce estimates of the nonfiling gap for each of these taxes. This paper provides details about these estimates and the methodologies used to produce them.

The paper is organized as follows: Section 1 explains the steps used for estimating the gap associated with not-filers, and how that methodology changed since the previous (TY 2011-2013) estimates; Section 2 explains the steps for estimating the gap associated with late filers; Section 3 summarizes the resulting nonfiling gap estimates; and Section 4 provides a summary of the methodological changes we implemented for these estimates compared with the method we used for our Tax Year 2011-2013 estimates.

[1] Hereinafter referred to simply as the nonfiling gap.

This page intentionally left blank.

Introduction

The nonfiling gap is the amount of true tax liability not paid on time by those who do not file on time. Since some nonfilers pay some or all of their true tax liability on time (e.g., through withholding), not all nonfilers actually contribute to the tax gap. Nonetheless, the nonfiling gap is comprised of two major components: the portion associated with those who file late ("late filers"), and the portion associated with those who never file at all ("not-filers").

We have estimated the individual income tax nonfiling gap and the self-employment tax nonfiling gap together (since self-employment tax is reported and reconciled on the Form 1040 individual income tax return), but we report them separately. For the first time, these estimates are based on an updated method in which annual Current Population Survey - Annual Social and Economic Supplement (CPS-ASEC) demographic surveys were linked to comprehensive IRS administrative data that the Census Bureau received from the IRS under section 6103(n) of the Internal Revenue Code. This approach is demonstrably superior to both the "Census Method" used for the TY 2008-2010 tax gap estimates, and the "Administrative Data Method" used for the TY 2011-2013 tax gap estimates.[2]

Section 1 below explains the new method used to estimate the gap associated with not-filers; Section 2 explains the steps for estimating the gap associated with late filers; and Section 3 summarizes the resulting nonfiling gap estimates.

1. Not-Filers

1.1 Overview of Previous Estimation Methods

Since not-filers do not declare their income or eligibility for deductions and credits on income tax returns, this is the most difficult portion of the nonfiling gap to estimate. Several methods have been employed in the past to estimate this portion of the tax gap. For example, in the early 1990's, the IRS estimated the nonfiling gap using a special study of Tax Year 1988 nonfilers under the Taxpayer Compliance Measurement Program (TCMP). This study selected a random sample of nonfilers, attempted to contact them and secure delinquent returns from them when possible; those secured returns were then subjected to line-by-line examinations to determine the true tax liability.[3] This approach is not only very costly, but it still requires estimating the gap associated with any not-filers for whom the IRS could not secure a delinquent return.

For the Tax Year 2001 tax gap estimates, the IRS turned to a different method: an "Exact Match" between Census and IRS data. This approach involved identifying respondents in the annual Current Population Survey who appeared not to have filed an income tax return, then estimating their income tax liability. This approach is much simpler, but the Census data do not capture all income, it is not always possible to determine whether a Census respondent filed on

[2] See Hertz *et al.* (2021).

[3] See Internal Revenue Service (1996) and Erard and Ho (2001).

time, and there was not a good method available to estimate the extent to which nonfilers paid at least some of their tax liability on time.

To estimate the Tax Year 2006 tax gap associated with not-filers, the IRS assembled a sample of individuals not appearing on filed tax returns, identified the income reported to the IRS for them by third parties, grouped them into family (tax) units (guided by Census data), imputed some additional income, deductions, and credits to them, then estimated their tax liability less credits and withholding. However, this approach (which we call the Administrative Sample Method) lacked information on income not reported to the IRS by third parties, and starting with a *sample* of individuals, created challenges for grouping people together into presumed tax units.[4]

For the Tax Year 2008-2010 tax gap estimates, the IRS used two methodologies: (1) an improved "Census Method" in which Census survey records were linked to limited tax administrative data; and (2) an improved "Administrative Data Method," which was based on population data rather than a sample. Because each of those methods had strengths and weaknesses, the separate estimates were averaged to arrive at the final estimates. The Census Method involved identifying respondents in the annual Current Population Survey who appeared not to have filed an income tax return, imputing income to them based on models trained on tax data for filers, placing them into tax units based primarily on Census data, then estimating their income tax liability. The income imputations made this method better than earlier "Exact Match" methods.[5]

To estimate the Tax Year 2011-2013 tax gap associated with not-filers, the IRS relied solely on a slightly improved Administrative Data Method because the Census Method became increasingly inaccurate while the Census survey data could be linked only to limited tax administrative data.[6]

The current (Tax Year 2014-2016) estimates are based on the best of both worlds—relying on detailed micro information on income from a greatly expanded set of tax administrative data that are linked at the person level with detailed demographic data from the Census. That is, it takes advantage of both the demographic information provided in the Census Method (to assign not-filers into tax units: filing status, dependents, etc.) and comprehensive tax information used in the Administrative Data Method (obviating the need to impute most types of income). Details are provided in Section 1.2.

We average our estimates over Tax Years 2014 through 2016 to correspond with the individual income tax underreporting gap estimates provided in the combined tax gap report.

1.2 New Census Method

For the current estimates, we retained many of the basic elements of the old Census Method, but we improved it in several important ways.

[4] See Internal Revenue Service (2012).

[5] See Langetieg *et al.* (2016).

[6] See Internal Revenue Service (2019) and Langetieg *et al.* (2017).

1.2.1 Elements Retained from the Prior Census Method

- Census has continued to improve their ability to assign an anonymous Protected Identification Key (PIK) to most respondents in the CPS-ASEC and to all the records Census receives from the IRS for the population from both income tax returns and from third-party information documents. This allows us to link Census survey records with tax administrative records to identify not-filers.[7]

- We used the third-party information about the income of the not-filers, together with demographic information about them contained in the CPS-ASEC to impute certain deduction and credit amounts.

- We estimated the tax liability of the not-filers using a detailed tax calculator.

- We supplement this estimate of the gap associated with not-filers with a separate estimate for late filers derived from IRS administrative data (see Section 3).

1.2.2 Differences from the Prior Census Method

- The most significant change in methodology arose from a new source of data at the Census Bureau: comprehensive tax administrative data from both tax returns and third-party information returns for the entire population. This was made possible through a special short-term IRS research project created under the authority of Internal Revenue Code section 6103(n). This obviated the need for most of the income imputations that were necessary for prior estimates.

- Another significant change in methodology related to our method for re-weighting the Census survey records that could be linked to the tax administrative data. This is necessary because there are some CPS-ASEC records that could not be matched to the IRS data because a PIK could not be assigned to them with adequate certainty. As with prior Census Method approaches, we therefore restricted our analysis to the records that *could* be matched and re-weighted them. However, unlike in the past (when we re-weighted the linked records to represent their share of the entire Census population) we have now created new weights that are designed such that the linked records represent the full population of potential nonfilers in the tax administrative data. This was important because nonfilers are relatively more likely not to have a good PIK assigned to them than tax return filers (who are more "visible").

- We also made better imputations of net self-employment income for these estimates. Instead of training our imputation model on the amount of self-employment income reported on filed returns, we trained it on data from the IRS National Research Program (NRP)—a stratified random sample of returns that were selected for a full audit. Instead of using the amount *reported* by taxpayers, the imputations are now based on the values of net self-employment income *as corrected* by the auditor. While the corrected self-employment amounts are closer

[7] See Jones and O'Hara (2014), and Wagner and Layne (2012).

to the "true" earnings than what is reported on returns, significant amounts of this income remain undetected by the auditors. But we are still assuming that the self-employment income of timely filers can be used as a basis for imputing such income to similarly situated nonfilers, as well.

- Finally, we were able to identify from third-party information documents (primarily Forms W-2 and 1099) the amount of income tax withheld from the income of the not-filers in the matched dataset. Although this allowed us to account for withholding quite accurately, the comprehensive tax data available at Census does not have information about other pre-payments of tax, such as estimated tax payments. We were nonetheless able to account for these miscellaneous timely payments by nonfilers using a small aggregate adjustment derived from IRS tax data.

1.2.3 The Expanded Tax Administrative Data

Section 6103 of the Internal Revenue Code protects the confidentiality of tax records. Sub-section (a) describes the General Rule: "Returns and return information shall be confidential, and except as authorized by this title [no one] shall disclose any return or return information obtained by him in any manner in connection with his service as such an officer or an employee or otherwise or under the provisions of this section."

Sub-section (j) provides for the statistical use of federal tax records by the Department of Commerce:

> Upon request in writing by the Secretary of Commerce, the Secretary [of the Treasury] shall furnish—
>
> (A) such returns, or return information reflected thereon, to officers and employees of the Bureau of the Census, and
>
> (B) such return information reflected on returns of corporations to officers and employees of the Bureau of Economic Analysis,
>
> as the Secretary may prescribe by regulation for the purpose of, but only to the extent necessary in, the structuring of censuses and national economic accounts and conducting related statistical activities authorized by law.

The IRS regularly provides limited data to the Census Bureau under this sub-section per detailed regulations.

Sub-section (n) further authorizes the release (e.g., to the Census Bureau) of protected tax information for the purposes of tax administration specifically:

> Pursuant to regulations prescribed by the Secretary [of the Treasury], returns and return information may be disclosed to any person, including any person described in section 7513(a), to the extent necessary in connection with the processing, storage, transmission, and reproduction of such returns and return information, the programming, maintenance, repair,

testing, and procurement of equipment, and the providing of other services, for purposes of tax administration.

Because estimating the extent and drivers of income tax nonfiling furthers tax administration, and because the data available under 6103(j) are inadequate for that purpose, the IRS entered into an agreement with the Census Bureau to transmit to Census a much more complete set of individual income tax records for a special short-term tax administration research project. This paper is one of the earliest studies based on these expanded (n) data.

Whereas the (j) data contained several indicators of the *presence* of certain types of income and very few income amounts, the (n) data include the *amounts* reported on income tax returns as well as the amounts reported on third-party information returns *for virtually every type of income* reported on such returns to the IRS. It is these latter amounts that allow us to estimate the filing obligations, tax liabilities, and balance due (or refund) for each person not represented on a timely filed tax return. Moreover, this level of income detail also allowed us to re-weight the linked records not connected with a filed return so that they represent not the original CPS-ASEC population, but rather the population of potential nonfilers among the tax records. The (n) data used for current estimates pertained to IRS Processing Years 2015-2019 (Tax Years 2014-2019), including both late filers and not-filers.

1.2.4 Imputing Self-Employment Income at the Individual Level

Given that only a small portion of self-employment income is reported on Form 1099-MISC (miscellaneous), we use regression models to impute this income to not-filers based on the net self-employment income amounts that should have been reported on filed returns as corrected by IRS National Research Program (NRP) audits.[8] Our imputations were restricted to the corrected amounts of net sole proprietorship income reported on Schedule Cs. We estimated the likelihood that a not-filer has self-employment earnings falling into one of the following three categories: (a) negative net self-employment earnings; (b) net self-employment earnings between $1 and $433; and (c) net self-employment earnings in excess of $433 (since taxpayers with more than $433 in net self-employment earnings are required to file a tax return and pay self-employment tax).

The econometric framework involved three separate models. The first was a probit specification for the likelihood that an individual has nonzero self-employment earnings:

$$SE^{*} = \gamma' x + \mu \qquad (1)$$

where SE^{*} is a latent variable describing the propensity for net self-employment earnings to be present, x is a vector of explanatory variables, and γ is a vector of coefficients to be estimated. The explanatory variables include five age categories, region, indicators for the presence of key income types as represented on third-party information documents (wages, interest, dividends, taxable state and local tax refunds, nonemployee compensation, gross amount of payment card and third party network transactions, capital gains, pensions, Schedule E, social security, unemployment compensation, and other income) and estimated payments, and the log of each of these income and payment amounts. The error term μ is assumed to follow the standard normal

[8] NRP audits are conducted on a stratified random sample representing all filed individual income tax returns.

distribution. Estimation of this model permits us to develop a prediction equation for the unconditional likelihood that an individual has a positive or negative net income from self-employment. Each individual was assigned a random number from a uniform distribution, and if the value of this number was below the predicted probability then the person was determined to have positive or negative net self-employment income.

Our second model was an ordered probit specification for the dollar amount category that net self-employment earnings fall into when they are present (negative, \$1 to \$433, or over \$433):

$$I^*_{SE} = \delta' x + \nu \qquad (2)$$

where I^*_{SE} is a latent variable for the propensity for net self-employment earnings to fall into one of these categories, x is the same set of explanatory variables used in the probit model, δ is a coefficient vector to be estimated, and v is a standard normal random disturbance. The model also includes a limit parameter l to be estimated.[9] The indicator I_{SE} for the net self-employment earnings category is assigned as follows:

$$I_{SE} = \begin{cases} 1 & net \quad earnings \quad < \quad \$0 \\ 2 & \$0 \quad < \quad net \quad earnings \quad \le \quad \$433 \\ 3 & net \quad earnings \quad > \quad \$433. \end{cases} \qquad (3)$$

Our third model is a regression specification for the magnitude of net self-employment earnings when they exceed \$433. Our specification is:

$$\ln(SE) = \beta' x + \varepsilon, \qquad (4)$$

where ln(SE) represents the natural log of net self-employment earnings, x is the same set of explanatory variables used in the preceding models, β' is a vector of coefficients to be estimated, and ε is assumed to be a normal random error term with mean zero and standard deviation σ. Under this specification, the distribution of self-employment earnings is assumed to be log normal. Each individual is assigned a random number from a normal distribution, which is multiplied by the root mean squared error and added to the predicted log amount. Furthermore, we have imposed the constraint that the imputed self-employment income amount cannot exceed the amount corresponding to the 99.99th percentile of self-employment income on filed returns.

1.2.5 Reweighting the CPS-ASEC Data

The administrative (n) data made it possible to re-weight the linked records so that the linked records represent our target population: the set of all individuals for whom the IRS received information documents[10] indicating economic activity for the tax year in question, but who do not appear on a timely filed Form 1040 for that year as either a primary filer or a spouse on a

[9] This parameter serves as a threshold for separating the various levels of the response variable.

[10] The information returns considered were: Forms W-2, W-2G, 1099-R, 1099-SSA, 1099-INT, 1099-DIV, 1099-MISC, 1099-LTC, 1099-PATR, 1099-Q, 1099-G, 1099-C, 1099-OID, 1099-S, 1041-K1, 1120S-K1, 1065-K1, 5498, 5498-SA, 1098, 1098-E, and 1098-T.

married-filing-jointly return. These are potential nonfilers; they are "potential" because they might not have a filing requirement (or a tax liability) due to having income that is less than the relevant filing thresholds. We limit these to records with a valid Taxpayer Identification Number (TIN—essentially, a valid social security number[11]) and a PIK. This is necessary because without the assurance that the TINs are valid, and belong to a unique individual, it is not possible to ascertain whether the information document in question can be linked to an individual in an unambiguous way. This does, however, imply that some people who are working under invalid social security numbers, including some foreign-born workers not lawfully present in the United States, will be dropped from our count of nonfilers. This imparts a downward (conservative) bias to our estimates.

We next require that each TIN is in the master list of individuals provided to the IRS by the Social Security Administration, and that the payee on the form has a birthdate that restricts them from being over 110 years old in the tax year in question. We also require that they were not deceased prior to that tax year. This left an average of 50.5 million potential nonfilers over the 2014-2016 period.

The weights were constructed using a probit model run on the full administrative population of potential nonfilers, with an outcome variable equal to 1 for records that were linked with a respondent in the CPS-ASEC, and 0 otherwise. The predictors in the model include age, gender, and the presence and decile locations (with respect to the full potential nonfiler population in the IRS administrative data) for each of the following types of taxable income: wages, interest, dividends, capital gains, farm, unemployment compensation, social security, pension, rents and royalties, tax refunds, the imputed amount of net self-employment income, and other.[12] Finally, each record in the linked data is assigned a weight equal to the inverse of the predicted probability of being linked. The overall average of that weight would be the total number of potential returns in the (n) data divided by the number of linked records from the CPS-ASEC data, but the weight for any given linked record would be somewhat higher or lower than that average based on the values of the predictor variables for that record. We have demonstrated that this new weighting methodology successfully allows the linked records to represent the full population of potential nonfilers.[13]

1.2.6 Potential Nonfilers

Having assigned a new weight to each record in the linked sample, we can easily tabulate a variety of income-related statistics for the population of individuals who are potential nonfilers. In Table 1 we show the counts of potential nonfilers with different types of income and the sum of the amounts of income for each income type, comparing estimates based on population data with weighted results using the CPS-ASEC records linked to the comprehensive (n) data. Table 2 shows the corresponding percentage of potential nonfilers with the different types of income and the mean amounts following the same column ordering as in Table 1.

[11] The anonymized data do not contain social security numbers (SSNs), but they do include indicators from the tax administrative data as to the type of TIN provided by the taxpayer, including whether it was a valid SSN.

[12] We found this to be the best specification for replicating the aggregate counts and amounts of income in the full population of potential nonfilers as found in the administrative data.

[13] See Hertz *et al.* (2021).

Column A in both Table 1 and Table 2 shows the counts and amounts for the full population of potential nonfilers using income found on third-party documents, except that net self-employment income is imputed based on formulas developed using amounts on filed returns as corrected by National Research Program audits.[14] In terms of income, this is the administrative data baseline that we would like to replicate using the linked sample (which is necessary in order to obtain greater micro-level accuracy in the creation of tax units and in the assignment of dependents).

Column B estimates the same values using the weighted records of the CPS-ASEC that were linked with the comprehensive (n) tax data. Like Column A, Column B includes the imputed amounts for net self-employment income. Notice that the estimates from the linked records in Column B match quite closely the corresponding estimates in Column A from the entire population.

[14] Remember that this population excludes those for whom a valid PIK could not be assigned to the taxpayer identified on the third-party documents.

Table 1. Estimates of the Number and Income of Potential Individual Income Tax Nonfilers (Person Level), Population (n) Data vs. Linked CPS-ASEC Sample, Tax Years 2014-2016

Tax Year:	2014		2015		2016	
Data source:	Administrative Population (n)	CPS-ASEC Linked to (n) data	Administrative Population (n)	CPS-ASEC Linked to (n) data	Administrative Population (n)	CPS-ASEC Linked to (n) data
Weights:	None (population)	IRS re-weights	None (population)	IRS re-weights	None (population)	IRS re-weights
Income source:	3rd-party information returns	Administrative (n) data with SE imputation	3rd-party information returns	Administrative (n) data with SE imputation	3rd-party information returns	Administrative (n) data with SE imputation
Type of income	**A**	**B**	**A**	**B**	**A**	**B**
Aggregate Counts (Millions)						
Total population count	49.68	49.63	50.16	50.34	51.63	51.65
Wages	15.08	15.09	15.77	15.98	17.04	17.27
Interest	6.96	7.01	6.72	6.74	6.79	6.85
Dividends	4.58	4.57	4.45	4.41	4.35	4.35
Capital gains	2.14	2.12	2.15	2.13	1.90	1.91
Pensions	6.75	6.81	6.87	7.01	7.06	7.12
Social security	23.79	23.96	24.02	24.92	24.34	24.53
Unemployment compensation	1.06	1.10	0.87	0.92	0.85	0.88
Net business and farm	4.99	5.06	5.14	5.28	5.43	5.46
Positive net business and farm	4.60	4.68	4.73	4.86	4.99	5.00
Negative net business and farm	0.41	0.40	0.42	0.42	0.44	0.46
Schedule E income	0.65	0.65	0.65	0.68	0.65	0.65
Other income	2.29	3.07	2.06	2.02	1.92	2.56
Aggregate Amounts (Billions US$)						
Total income	$657.0	$666.9	$701.1	$723.2	$738.8	$760.4
Wages	$215.3	$213.7	$242.5	$244.3	$269.8	$280.8
Interest	$1.5	$1.5	$1.5	$1.3	$1.5	$1.4
Dividends	$2.9	$2.9	$2.8	$2.6	$2.7	$2.6
Capital gains	$2.1	$1.8	$1.9	$1.6	$1.2	$0.9
Pensions	$51.2	$52.4	$54.6	$57.1	$56.9	$56.4
Social security	$266.6	$269.7	$275.3	$287.3	$282.0	$284.7
Unemployment compensation	$3.6	$3.8	$3.4	$3.6	$3.5	$3.7
Net business and farm	$87.5	$92.8	$93.2	$96.8	$98.8	$101.7
Positive net business and farm	$90.0	$95.2	$95.8	$98.8	$101.5	$104.0
Negative net business and farm	-$2.5	-$2.4	-$2.6	-$2.0	-$2.7	-$2.3
Schedule E income	$4.2	$5.3	$4.4	$7.6	$4.1	$6.3
Other income	$22.1	$23.0	$21.5	$21.0	$18.3	$21.9

Census Bureau Disclosure Review Board release authorizations CBDRB-FY2021-CES005-020 and CBDRB-FY22-P2599-R9418.

Table 2. Estimates of the Incidence and Average Amount of Income of Potential Individual Income Tax Nonfilers (Person Level), Population (n) Data vs. Linked CPS-ASEC Sample, Tax Years 2014-2016

Tax Year:	2014		2015		2016	
Data source:	Administrative Population	CPS-ASEC Linked to (n) data	Administrative Population	CPS-ASEC Linked to (n) data	Administrative Population	CPS-ASEC Linked to (n) data
Weights:	None (population)	IRS re-weights	None (population)	IRS re-weights	None (population)	IRS re-weights
Income source:	3rd-party information returns	Administrative (n) data with SE imputation	3rd-party information returns	Administrative (n) data with SE imputation	3rd-party information returns	Administrative (n) data with SE imputation
Type of income	**A**	**B**	**A**	**B**	**A**	**B**
			Incidence			
Wages	30.4%	30.4%	31.4%	31.7%	33.0%	33.4%
Interest	14.0%	14.1%	13.4%	13.4%	13.1%	13.2%
Dividends	9.2%	9.2%	8.9%	8.8%	8.4%	8.4%
Capital gains	4.3%	4.3%	4.3%	4.2%	3.7%	3.7%
Pensions	13.6%	13.7%	13.7%	13.9%	13.7%	13.7%
Social security	47.9%	48.3%	47.9%	49.5%	47.1%	47.5%
Unemployment compensation	2.1%	2.2%	1.7%	1.8%	1.6%	1.7%
Net business and farm	10.0%	10.2%	10.3%	10.5%	10.5%	10.6%
Positive net business and farm	9.3%	9.4%	9.4%	9.7%	9.6%	9.7%
Negative net business and farm	0.8%	0.8%	0.8%	0.8%	0.8%	0.9%
Schedule E income	1.3%	1.3%	1.3%	1.4%	1.3%	1.3%
Other income	4.6%	6.2%	4.1%	4.0%	3.7%	5.0%
			Average Amounts			
Wages	$4,333	$4,305	$4,834	$4,854	$5,225	$5,436
Interest	$31	$30	$29	$26	$29	$27
Dividends	$58	$58	$55	$51	$53	$50
Capital gains	$42	$37	$37	$32	$23	$17
Pensions	$1,031	$1,056	$1,089	$1,134	$1,103	$1,091
Social security	$5,367	$5,435	$5,489	$5,707	$5,463	$5,512
Unemployment compensation	$73	$76	$68	$72	$68	$71
Net business and farm	$1,761	$1,870	$1,859	$1,924	$1,914	$1,970
Positive net business and farm	$1,812	$1,918	$1,910	$1,963	$1,966	$2,014
Negative net business and farm	-$50	-$47	-$51	-$39	-$52	-$44
Schedule E income	$84	$107	$87	$151	$79	$123
Other income	$445	$463	$430	$418	$355	$425

Census Bureau Disclosure Review Board release authorizations CBDRB-FY2021-CES005-020 and CBDRB-FY22-P2599-R9418.

1.2.7 Forming Tax Units

Having identified the potential nonfilers in the linked data and having re-weighted those records to represent the full population of potential nonfilers, we were able to adapt the standard tax model we had developed for use with the (j) data to assign the potential nonfilers into tax units. But, instead of using just the CPS-ASEC records matched to the IRS third-party documents, as explained later, we also in some cases use CPS-ASEC records with PIKs that are not found on a filed return. To build the tax units, we: (1) combined the records of spouses;[15] (2) assigned them the Married filing jointly filing status; and (3) assigned all others to either Single or Head of Household filing status, depending on their demographics in the CPS-ASEC. Using family demographics like this at the micro level is a key benefit of linking the tax administrative data to the CPS-ASEC survey records.

After creating tax units in this way, we were then able to identify which of them appear to have had a filing requirement. We took into account two key filing thresholds. First, tax units with more than $433 of net self-employment income are required to file a return to report that income; even if they do not have an income tax liability, they may have a self-employment tax liability. This is one reason we impute self-employment income to the potential nonfilers.[16]

The more common filing threshold applies to everyone. For Tax Years 2014-2016, people were required to file a tax return if their gross income exceeded the sum of their standard deduction, any additional standard deduction on account of being 65 or older or blind, and the value of the personal exemption(s) of the primary taxpayer (and spouse, if any).[17] This means that the gross income filing threshold for singles who were neither 65 or older nor blind was $10,150 for Tax Year 2014, $10,300 for Tax Year 2015, and $10,350 for Tax Year 2016. Again, someone's gross income could give them a filing requirement even if they do not have any income tax liability but the offsetting costs, expenses, deductions, credits, etc. that would reduce their tax liability to zero would need to be reported on a filed tax return.

We also assigned children to these tax units based on the information in the CPS-ASEC for the linked record. These would reduce income tax liability through dependent exemptions, credits, etc.

1.2.8 Calculating Tax and Balance Due

As in our previous estimates derived from the (j) data linked to the CPS-ASEC,[18] the tax model computes self-employment tax, the adjustment for one-half of the self-employment tax, exemptions, the standard and itemized deductions, taxable income, tentative tax, nonrefundable credits, refundable credits, and tax balance due after credits. Included in these computations are imputations for deductions, nonrefundable credits other than the Child Tax Credit, and

[15] In the case when one spouse could be linked to the (n) data and the other could not, for the purposes of this paper we used the income reported in the CPS-ASEC as that spouse's income. This may partially overcome the fact that not all information documents in the (n) data had been assigned a reliable PIK (see Section 1.2.2). However, the IRS re-weights do not account for that missing spouse among the potential nonfilers.

[16] Note that we impute net self-employment earnings.

[17] This threshold does not take into account any dependents; those need to be claimed on a tax return.

[18] See Langetieg *et al.* (2016, p. 4-5).

adjustments other than the adjustment for one-half of self-employment tax.[19] Subtracting the amount of tax withheld (per Forms W-2 and other information documents when using the (n) data and imputed withholding amounts otherwise), we arrive at an estimate of the net balance due or refund. We aggregate the positive balance due amounts and other quantities on the mock tax return to the population of nonfilers using the IRS re-weights described earlier.

Table 3 presents the estimated number of tax units associated with each filing status, the counts of tax units having the given income types, and the corresponding dollar amounts for the income and tax categories. However, there are several differences between the person-level estimates in Table 1 and the tax unit estimates in Table 3—beyond the facts that the tax unit estimates combine spouses into tax units and all tax units are restricted to those that have a filing obligation. Unlike Table 1, Table 3 uses income reported on the CPS-ASEC when no information returns in the (n) data are matched to the individual (see footnotes 13, 14, and 15). The tax units in Table 3 potentially include everyone in the CPS-ASEC with a valid PIK who was not matched to a Form 1040 and associated with a tax return that was estimated to be required. On the other hand, the person-level estimates in Table 1 include only those for whom the CPS-ASEC record matches to at least one 3rd-party information document in the (n) data. The CPS-ASEC income is included only for spouses who are not linked to any IRS third-party data.

[19] For each of these imputations, a two-step model is applied. In the case of deductions, the first model estimates the likelihood that the taxpayer itemizes deduction rather than taking the standard deduction. For adjustments and credits, the first model estimates the likelihood that the taxpayer has a positive, non-zero amount. In all three cases, the second model estimates the log amount for each of the aggregate line items.

Table 3. Estimates of the Number, Income, and Tax Items of Individual Income Tax Not-filers (Return Level), CPS-ASEC Linked to Administrative (n) Data, Tax Years 2014-2016

Type of income	Tax Year			2014-2016 Average
	2014	2015	2016	
		Aggregate counts (Millions)		
Tax units	10.57	11.09	11.90	11.19
Single tax units	5.37	5.47	5.95	5.60
Married filing jointly tax units	1.99	2.31	2.29	2.20
Head of household tax units	3.22	3.31	3.66	3.40
Wages	6.25	6.85	7.59	6.90
Interest	1.58	1.59	1.65	1.61
Dividends	1.36	1.20	1.13	1.23
Capital gains	0.72	0.66	0.52	0.63
Pensions	1.96	2.11	2.18	2.08
Taxable social security	0.95	1.15	1.06	1.05
Unemployment compensation	0.71	0.63	0.59	0.64
Net business and farm	4.79	4.87	5.30	4.99
Positive net business and farm	4.69	4.72	5.14	4.85
Negative net business and farm	0.11	0.14	0.16	0.14
Schedule E Income	0.33	0.32	0.38	0.34
Other Income	0.87	0.61	0.63	0.70
		Aggregate Amounts (Billions US$)		
Wages	199.4	230	266.8	232.1
Interest	0.6	0.5	0.4	0.5
Dividends	1.9	1.6	1.4	1.6
Capital gains	1.4	1.1	0.5	1.0
Pensions	34.0	40.0	37.9	37.3
Taxable social security	8.3	10.3	9.3	9.3
Unemployment compensation	3.0	2.8	2.7	2.8
Net business and farm	94.0	95.8	110.7	100.2
Positive net business and farm	95.0	97.4	113.3	101.9
Negative net business and farm	-1.0	-1.6	-2.6	-1.7
Schedule E income	5.8	8.0	6.9	6.9
Other income	16.6	14.8	16.0	15.8
Total income	367.1	406.6	454.6	409.4
Adjusted Gross Income	356.3	395.4	442.5	398.1
Deductions	92.4	101.3	109.8	101.2
Exemptions	56.0	60.4	66.7	61.0
Taxable income	208.0	233.6	266.0	235.9
Tentative tax	33.9	38.1	44.9	39.0
Nonrefundable credits	2.1	2.4	2.5	2.3
Income tax	31.8	35.7	42.4	36.6
Self-employment tax	12.6	12.9	14.5	13.3
Total tax	44.3	48.6	56.9	49.9
Withholding	16.2	18.6	22.8	19.2
Estimated tax payments	0.3	1.2	1.0	0.8
Refundable credits	1.4	1.4	1.4	1.4
Balance Due (contribution to tax gap)	**26.4**	**27.4**	**31.7**	**28.5**

Census Bureau Disclosure Review Board release authorizations CBDRB-FY2021-CES005-020 and CBDRB-FY22-P2599-R9418.

Note that the 49.6 million potential nonfilers in Tax Year 2014 from Table 1 reduces to 10.6 million nonfiler tax units in Table 3. This is partly because around 2 million spouses were combined into one tax unit per couple, but mostly because we estimate that about 37 million potential nonfilers did not have a filing obligation. Notice also that under 20 percent of nonfiler tax units were married in Tax Year 2014 (as opposed to about 38 percent of filers[20]).

The total income of nonfilers is estimated in Table 3 to average about $409 billion for the Tax Year 2014 through 2016 period. This is about 57 percent of the corresponding total income of all potential nonfilers shown in Table 1 (the remaining 43 percent of the income is spread among the 37 million potential nonfilers—averaging about $8,300 per person, which was under the average gross income filing threshold of $10,267 per person). Finally, Table 3 indicates that the total tax balance due (contribution to the tax gap) of the not-filers (after withholding, estimated tax payments, and nonrefundable credits) is estimated to average $28.5 billion over this period.

2. Late Filers

In addition to not-filers, who don't file a tax return at all, late filers also make a significant contribution to the nonfiling gap since they have a lot of unpaid tax but did not meet the filing deadline. Compared with not-filers, however, they do not contribute as much to the tax gap because they pay a much larger portion of their tax liability on time, such as through withholding and tax credits. Unlike not-filers, of course, we have tax returns for the late filers, so estimating their contribution to the gap is much more straightforward. On the surface, the gap is their aggregate balance due. However, we adjust this amount to take into account income and payments that are not reported on the late returns but are reported to the IRS on third-party information documents.[21] All of the data needed to estimate the nonfiling gap due to late filers is present in IRS administrative data, and we estimate it from multiple large samples drawn from population data (to mitigate the effects of data errors). Our estimates are provided in Table 4. Because we have comprehensive administrative (n) data for IRS Processing (i.e., calendar) Years 2015 through 2019 that can be linked to the Census survey data, we are able to identify those who file a Tax Year 2014, 2015, or 2016 tax return up to three years late.[22]

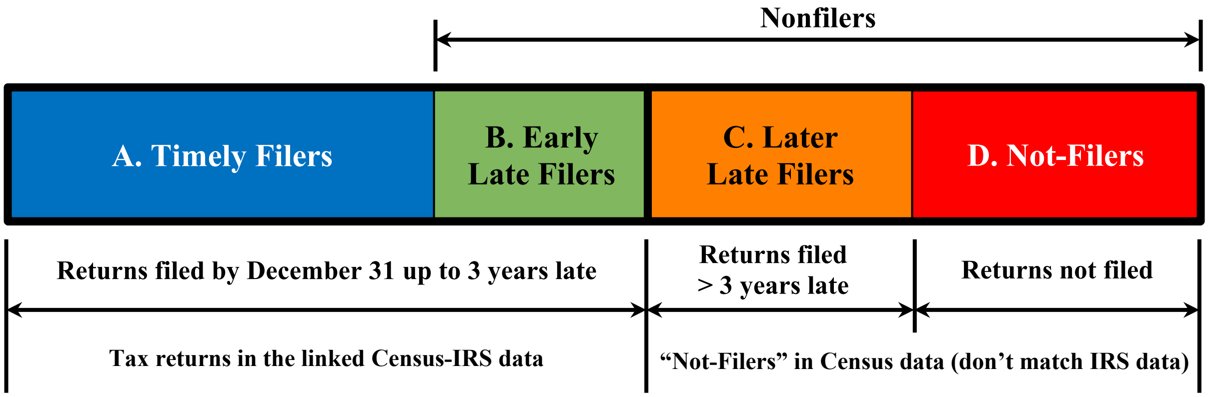
However, that means that the *later* late filers (those who filed more than 3 years late) appear as "not-filers" in the matched dataset, causing us to overstate the true not-filer portion of the gap. To avoid double-counting, we need to add only the *early* late filers to the Census-based estimate of not-filers. So, the total nonfiling gap is still the sum of the not-filer and late filer portions. See Figure 1.

[20] See IRS, SOI Tax Stats - Individual Statistical Tables by Filing Status

[21] See Section 2.1. We do not impute other kinds of income to them (such as from self-employment). However, late filers already report a significant amount of these kinds of income.

[22] We use December 31 of the relevant year (e.g., December 31, 2017 for Tax Year 2014) as the cut-off for distinguishing between late filers and not-filers.

Figure 1. The Role of Late Filers in the Census-Based Method


Nonfilers
A. Timely Filers
B. Early Late Filers
C. Later Late Filers
D. Not-Filers
Returns filed by December 31 up to 3 years late
Returns filed > 3 years late
Returns not filed
Tax returns in the linked Census-IRS data
"Not-Filers" in Census data (don't match IRS data)

Since the later late filers (C) are already accounted for among the Census "not-filers", only the early late filers (B) need to be added to the Census "not-filers" to estimate the nonfiling gap.

2.1 Method for Incorporating Third-Party Information for Late Filers

Like filers, some late filers do not report amounts consistent with the information reported on their behalf by third parties. We accounted for this for each late filer using the logic summarized in Table 4 for each line item on the return.

After accounting for additional income using the logic presented in Table 4, we re-calculated tax and the balance due for each return. We assumed that the total of all withholding for a given taxpayer that was documented by third parties on information returns was not more accurate than the amount reported by the taxpayer on his or her Form 1040.

2.2 Method for Handling Outliers in Population Data

A sampling method was applied to the Late Filer estimates to minimize the impact of administrative transcription errors and other outlier data issues that exist in the raw administrative data. The sampling method consisted of tabulating results for 100 to 125 one percent samples. The samples were ordered by aggregate balance due and the middle ten were selected and averaged to create our final estimates.

Table 4. Logic for Using Information Return Data to Adjust Items Reported on Late Returns

	Form	Line	Item	Adjustment Logic
A	1040	7	Wages	Let GIC = Max[(D-E+G), (J+I+H+F), 0] • If A>0 and (B+C)>0 and GIC>0 and -150<(B+C+L-GIC)<150, then: ○ Wages = (B+C) and ○ Schedule C net income = Max[K-(B+C), 0] • Else, if A>0 and (B+C)>0 and GIC=0 and -150<(A-(B+C))<150, then: ○ Wages = Max[A-L, (B+C), 0] and ○ Schedule C net income = L • Else: ○ Wages = Max[A, (B+C), 0] and ○ Schedule C net income = Max[K, (L-GIC)+K]
B	W-2	1	Wages	
C	W-2	8	Allocated tips	
D	Schedule C	1	Gross receipts	
E	Schedule C	2	Returns & allowances	
F	Schedule C	4	Cost of goods sold	
G	Schedule C	6	Other income	
H	Schedule C	28	Total expenses	
I	Schedule C	30	Business use of home	
J	Schedule C	31	Net profit (loss)	
K	1040	12	Schedule C net income	
L	1099MISC	7	Non-empl compensation	
M	1040	8a	Taxable interest	Interest income = Max[M, (N+O+P+Q+R)]
N	1099-INT	1	Interest income	
O	1099-INT	3	Interest on savings bonds	
P	K-1 (1041)	1	Interest income	
Q	K-1 (1120S)	4	Interest income	
R	K-1 (1065)	5	Interest income	
S	1040	9a	Ordinary dividends	Ordinary taxable dividends = Max[S, (T+U+V+W)]
T	1099-DIV	1a	Ordinary dividends	
U	K-1 (1041)	2a	Ordinary dividends	
V	K-1 (1120S)	5a	Ordinary dividends	
W	K-1 (1065)	6a	Ordinary dividends	
X	1040	9b	Qualified dividends	Qualified dividends = Min[X, Y] (The qualified dividends amounts from the Forms K-1 are not in our data.)
Y	1099-DIV	1b	Qualified dividends	
Z	1040	10	State tax refunds	State tax refund = Max[Z, Min[AA, AB]]
AA	1099-G	2	State tax refunds	
AB	Schedule A	5	Prior year deduction for S&L income taxes	
AC	1040	13	Capital gain (loss)	IRPCG = (AD+AE+AF+AG+AH+AI+AJ) Capital gain = Max[AC, IRPCG]
AD	1099-DIV	2a	Cap. gain distribution	
AE	K-1 (1041)	3	Net ST cap. gain (loss)	
AF	K-1 (1041)	4a	Net LT cap. gain (loss)	
AG	K-1 (1120S)	7	Net ST cap. gain (loss)	
AH	K-1 (1120S)	8a	Net LT cap. gain (loss)	
AI	K-1 (1065)	8	Net ST cap. gain (loss)	
AJ	K-1 (1065)	9a	Net LT cap. gain (loss)	
AK	1040	15a	IRA distributions	IRA and pension income combined to account for misclassification. If AK=0, then AK=AL If AM=0, then AM=AN IRA + Pension income = Max[(AL+AN), (AO-AK+AL), (AP-AM+AN)] AP=0 (to avoid double-counting pension income)
AL	1040	15b	Taxable IRA distrib'n	
AM	1040	16a	Pensions & annuities	
AN	1040	16b	Taxable pension, annuity	
AO	5498	3	Roth conversion amt	
AP	1099-R	2a	Taxable pension	
AQ	1040	18	Farm income or loss	Farm income = Max[AQ, (Max[AR,0] + Max[AS,0])]
AR	1099-G	7	Agricultural subsidy	
AS	1099-MISC	10	Crop insurance proceeds	
AT	1040	19	Unemployment comp.	Unemployment compensation = Max[AT, AU]
AU	1099-G	1	Unemployment comp.	
AV	1040	20a	Social security benefits	Social security benefits = Max[AV, AW]
AW	1099-SSA	3	SS benefits	

AX	1040	21	Other income	Line21Calc=AY+AZ+BA
				If (AX<0 and Line21Calc=0) or (Schedule C net income ≠ 0) or (Farm income ≠ 0) then: Other income = AX;
				Else: Other income = Max[AX, Line21Calc]

AY	W-2G	1	Gross winnings	
AZ	1099-C	2	Amt of debt cancelled	
BA	1099-G	5	ATAA payment	

	Form	Line	Item	Adjustment Logic
BB	1040	17	Schedule E net income	GrossE = (BC+BD+Max[(BE+BF), BG]+BH+BJ+Max[BI, 0]) If BB > GrossE, Then GrossE = BB Note: any negative amount from any of the following components is set to zero: Line17Calc = BK+BL+BM+BN+BO+BP+BQ+BR+BS+BT+BU+BV+BW+BX+BY Schedule E net profit (loss) = Max[BB, BB + (Line17Calc – GrossE)]
BC	Schedule E	23c	Total rents received	
BD	Schedule E	23d	Total royalties received	
BE	Schedule E	29a (g)	Passive income from partnership or S corp	
BF	Schedule E	29a (j)	Non-passive inc. from partnership or S corp	
BG	Schedule E	30	Passive + non-passive inc. from partn or S corp	
BH	Schedule E	35	Estate & trust income	
BI	Schedule E	40	Farm rental net income	
BJ	Schedule E	41	REMIC net income	
BK	K-1 (1065)	1	Ordinary business inc.	
BL	K-1 (1065)	2	Net rental real estate inc.	
BM	K-1 (1065)	3	Other net rental income	
BN	K-1 (1065)	4	Guaranteed payments	
BO	K-1 (1065)	7	Royalties	
BP	K-1 (1041)	5	Other portfolio income	
BQ	K-1 (1041)	6	Ordinary business inc.	
BR	K-1 (1041)	7	Net rental real estate inc.	
BS	K-1 (1041)	8	Other rental income	
BT	K-1 (1120S)	1	Ordinary business inc.	
BU	K-1 (1120S)	2	Net rental real estate inc.	
BV	K-1 (1120S)	3	Other rental income	
BW	K-1 (1120S)	6	Royalties	
BX	1099-MISC	1	Rents	
BY	1099-MISC	2	Royalties	
BZ	1040	64	Tax withheld	Total withholding = CB+CC+CD+CE+CF+CG+CH+CI+CJ+CK+CL+CM+CN Total prepayments = Total withholding + CA
CA	1040	65	Estimated tax payments	
CB	W-2	2	Income tax withheld	
CC	W-2G	2	Income tax withheld	
CD	K-1 (1120S)	13(Q)	Backup withholding	
CE	1099-B	4	Income tax withheld	
CF	1099-SSA	6	Income tax withheld	
CG	1099-RRB	10	Income tax withheld	
CH	1099-G	4	Income tax withheld	
CI	1099-DIV	4	Income tax withheld	
CJ	1099-INT	4	Income tax withheld	
CK	1099-MISC	4	Income tax withheld	
CL	1099-OID	4	Income tax withheld	
CM	1099-PATR	4	Income tax withheld	
CN	1099-R	4	Income tax withheld	

3. Nonfiling Gap Estimates

Our overall estimates of the individual income tax nonfiling gap, averaged over Tax Years 2014 through 2016 are provided in Table 5—adding the gap associated with late filers and not-filers. We average the estimates over the TY2014-2016 period to arrive at an estimate that is comparable to the underreporting gap estimates.

Table 5. Individual Income Tax and Self-Employment Tax Nonfiling Tax Gap Estimates ($ in Billions), Tax Years 2014-2016

Type of income	Not-Filers*			Late Filers**			All Nonfilers			TY14-16
	2014	**2015**	**2016**	**2014**	**2015**	**2015**	**2014**	**2015**	**2016**	**Average**
	Aggregate counts (Millions)									
Tax units	10.6	11.1	11.9	5.4	5.2	6.1	16.0	16.3	18.0	16.7
	Aggregate Amounts (Billions US$)									
Total income	367.1	406.6	454.6	360.8	347.1	404.4	727.9	753.7	859.0	780.2
Adjusted Gross Income	356.3	395.4	442.5	355.3	341.7	398.1	711.6	737.1	840.6	763.1
Taxable income	208.0	233.6	266.0	246.9	235.7	274.4	454.9	469.3	540.4	488.2
Tentative tax	33.9	38.1	44.9	51.7	48.9	56.2	85.6	87.0	101.1	91.2
Income tax	31.8	35.7	42.4	49.4	46.7	53.7	81.2	82.4	96.1	86.6
Self-employment tax	12.6	12.9	14.5	3.9	3.8	4.4	16.5	16.7	18.9	17.4
Total tax	44.3	48.6	56.9	53.3	50.5	58.1	97.6	99.1	115.0	103.9
Withholding	16.2	18.6	22.8	29.5	28.7	32.9	45.7	47.3	55.7	49.6
Estimated tax payments	0.3	1.2	1.0	9.2	8.2	9.2	9.5	9.4	10.2	9.7
Refundable credits applied to tax	1.4	1.4	1.4	4.5	3.8	4.7	5.9	5.2	6.1	5.7
Balance Due (contribution to the tax gap)	**26.4**	**27.4**	**31.7**	**10.1**	**9.8**	**11.3**	**36.5**	**37.2**	**43.0**	**38.9**
Income tax nonfiling gap ($B)							**30.3**	**31.0**	**35.9**	**32.4**
Self-employment tax nonfiling gap ($B)							**6.2**	**6.3**	**7.1**	**6.5**

* CPS-ASEC linked to administrative (n) data under Census Bureau Disclosure Review Board release authorizations CBDRB-FY2021-CES005-020 and CBDRB-FY22-P2599-R9418. Taxpayers who filed a TY2014 return by December 31, 2017, or a TY2015 return by December 31, 2018, or a 2016 return by December 31, 2019 are not included in the not-filing populations.

** Derived from population data tabulated on the IRS Compliance Data Warehouse. Filed late but within 3 years of the year originally due.

4. Differences in the Nonfiling Tax Gap Estimate Due to Changes in Methodology

As indicated in Section 1.1, our tax gap estimates for Tax Years 2011-2013 were based solely on IRS administrative data (using aggregate Census demographic data to place potential nonfilers into tax units). Our method for Tax Years 2014-2016 differs from our prior method in several important ways:

- The most important change was linking Census CPS-ASEC records to comprehensive tax administrative data, which allowed us to assign demographic information to each potential nonfiler so that they could be given the appropriate tax filing status, number of dependents, etc. Forming tax units in this way is much more accurate at the micro level than imputing filing status and dependents probabilistically. This allows us to analyze nonfiling within narrow portions of the population with greater confidence. Projecting results from the linked sample to the entire population of potential nonfilers was made possible by an improved weighting methodology (in lieu of the standard Census weights).
- The current estimates reflect an improved method for imputing self-employment income to potential nonfilers. In prior studies we trained our imputation model on the propensity of filers to report self-employment income. We now train our model on data from the IRS National Research Program (the results of audits of a representative sample of filed tax returns). This allows us to account for self-employment income that the NRP auditors detected even if the taxpayer did not report it. Our imputations, then, represent the amount of self-employment income that (in the judgment of the average NRP auditor) potential nonfilers *should have* reported on a tax return had they filed one—not the amount they *would have* reported if they did so with the same propensity as the average timely filer. This still assumes that potential nonfilers have the same level of self-employment income as similarly situated timely filers, but we have no means for modifying this assumption.
- We continue to distinguish between late filers and not-filers. However, our current estimates define late filers as those who file late, but before the end of the third year after the end of the tax year in question. Our estimates for TY2011-2013 defined late filers as those who file late, but before the end of the *fourth* year after the end of the tax year in question. That is, instead of defining late filers for Tax Year 2014 as those who file before December 31, 2018, we now define them as those who file before December 31, 2017. Those who actually filed in that fourth year are treated in the current methodology as not-filers. There are two reasons for this change: (a) the last year of tax administrative data available to us at Census was Processing Year 2019, providing only 3 years of late filing data for Tax Year 2016; and (b) although we could have applied the 4-year definition to Tax Years 2014 and 2015, we chose to apply the same definition to each year so that the three-year average represented one definition (and was therefore not subject to the relative accuracy of estimating someone's contribution to the tax gap depending on whether we treated them as a late filer or as a not-filer).[23]

The net effect of these methodological changes is a reduction in the three-year average nonfiling gap estimate by $0.3 billion, suggesting that the Administrative method is quite accurate when aggregated to the entire population.

[23] Preliminary analysis suggests that this one change made little difference on the overall estimate, however.

References

Erard, B. and Chih-Chin Ho (2001). "Searching for ghosts: who are the nonfilers and how much tax do they owe?," *Journal of Public Economics* 81, p. 25–50.

Internal Revenue Service (1996). *Federal Tax Compliance Research: Individual Income Tax Gap Estimates for 1985, 1988, and 1992*, Publication 1415 (Rev. 4-96).

Internal Revenue Service (2012). *Federal Tax Compliance Research: Tax Year 2006 Tax Gap Estimation,* at http://www.irs.gov/pub/irs-soi/06rastg12workppr.pdf.

Internal Revenue Service (2019). *Federal Tax Compliance Research: Tax Gap Estimates for Tax Years 2011–2013,* Publication 1415 (Rev. 9-2019).

Hertz, Thomas, Pat Langetieg, Mark Payne, Alan Plumley, and Margaret Jones (2021). "Estimating the Extent of Individual Income Tax Nonfiling," *2021 IRS Research Bulletin*, Publication 1500.

Jones, Maggie R. and Amy O'Hara (2014). "Do Doubled-Up Families Minimize Household-Level Tax Burden?" *2014 IRS Research Bulletin*, Publication 1500, p. 181-203.

Langetieg, P.T., Payne, J.M., and Plumley, A.H. (2016). The Individual Income Tax and Self-Employment Tax Nonfiling Gaps for Tax Years 2008-2010. Research, Applied Analytics & Statistics Technical Paper

Langetieg, P.T., Payne, J.M., and Plumley, A.H. (2017). "Counting Elusive Nonfilers Using IRS Rather Than Census Data," *2017 IRS Research Bulletin*, Publication 1500, p. 197-222. 17resconpayne.pdf

Wagner, D. and Layne, M. (2012). "Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications' Record Linkage Software," Washington, DC: Center for Administrative Records Research and Applications Internal Document, U.S. Census Bureau.