What Drives Filing Compliance?

Brian Erard (*B. Erard and Associates*) and John Guyton, Pat Langetieg, Mark Payne, and Alan Plumley (*IRS Research, Analysis, and Statistics: Office of Research*)[1]

ABSTRACT: IRS estimates that nonfilers were responsible for $25 billion in unpaid federal individual income tax liabilities in tax year 2006, or about 8.5 percent of the overall individual income tax gap. In this paper, we attempt to identify the factors that drive this form of noncompliance. The decision whether to file a tax return is essentially a participation decision, and there is a long history in empirical research of applying qualitative choice models, such as logit and probit specifications, to model participation outcomes. However, standard qualitative choice models assume that one has access to a representative data sample of participants and non-participants, including an indicator for the participation status of each subject in the sample. While IRS has detailed tax return information for filers (or participants), it lacks comparable information on nonfilers (or non-participants). To fill this void, we supplement IRS tax return information for filers with Census survey information from the general population of filers and nonfilers. However, this latter data source does not identify which respondents are filers and which are nonfilers. We therefore develop and apply a novel econometric methodology ("calibrated qualitative choice analysis") to estimate the drivers of filing compliance. Intuitively, whereas a standard logit or probit analysis relies on differences between the characteristics of participants and non-participants to infer what drives behavior, our methodology relies on differences between the characteristics of participants and the overall population of participants and non-participants.

Our IRS data source on filers is the Individual Returns Transaction File (IRTF). Many households have no legal filing obligation because their income is below the filing threshold and they do not meet certain other filing criteria, such as a need to report taxes on tip income. Some of these households do file, however, to claim refunds of withheld earnings or to claim a refundable tax credit, such as the Earned Income Credit. Since our focus is on filing compliance, we restrict our IRTF sample to households with a legal filing obligation. This is achieved by applying an algorithm to check whether a given return satisfies any of the various conditions (such as gross income above the relevant filing threshold or net self-employment earnings in excess of $400) that trigger a filing requirement. Our supplementary sample of filers and nonfilers is drawn from the Current Population Survey Annual Social and Economic Supplement. In past research, we have found that certain income sources are understated in this survey. Therefore, in order to more accurately identify households with a legal filing obligation, we follow Erard, Langetieg, Payne, and Plumley (2014) in imputing additional income across the sample.[2] To assign household members to tax returns, we also impute tax filing status. The CPS-ASEC is a stratified random

---

[1] Contract: Brian Erard, BEandAssoc@Aol.com, 703-390-9368.

[2] Erard, Brian, Pat Langetieg, Mark Payne, and Alan Plumley (2014). "Missing Returns vs. Missing Income: Estimating the Extent of Individual Income Tax Filing Noncompliance from IRS and Census Data", *Proceedings of the 107th Annual Conference on the National Tax Association*, Santa Fe, New Mexico (November 13-15, 2014). *https://www.ntanet.org/wp-content/uploads/proceedings/2014/030-erard-langetieg-payne-plumley-missing-returns-vs-missing-income.pdf*

sample; however, the stratification criteria are not publicly available. A desirable feature of our econometric methodology is that we are able to effectively control for the stratified nature of the sample simply by applying the sample weights.

We begin by estimating cross-sectional models of filing behavior for selected years. We find that the impact of taxpayer reporting burden on filing compliance depends on whether a household's income is close to the filing threshold. For those with income well above the threshold, the reporting burden is positively associated with filing compliance, presumably because taxpayers with higher income and greater tax complexity have greater opportunities for legal tax avoidance, which comes at the price of a greater filing burden. In contrast, a high reporting burden is found to discourage filing among households with income only slightly above the filing threshold. Our results also indicate that filing compliance is sensitive to various socio-economic factors, including income level, marital status, and geographic location.

We then extend our framework to permit analysis of a time series of cross-sections, which enables us to explore how filing compliance responds to year-specific factors. Our results indicate that the Economic Stimulus Act of 2008 provided a large, albeit temporary, stimulus to filing compliance, indicating that the financial incentive of the program was sufficiently strong that some households with a legal filing obligation overcame their reluctance to file a return.