Does Giving Tax Debtors a Break Improve Compliance and Income? Evidence from Quasi-Random Assignment of IRS Revenue Officers

William C. Boning Joel Slemrod Ellen Stuart Alex Turk

This version: January 10, 2023

This paper uses the quasi-random assignment of IRS Revenue Officers to tax debtors' cases as an instrumental variable to identify the causal effects of suspending debt collection on tax compliance and future income. In contrast to uninstrumented estimates, we find no statistically significant evidence that putting off attempts to collect debt reduces compliance with future tax obligations or future reported income. Among marginal hardship cases, pausing collection instead increases future income, specifically W-2 earnings by the taxpayer's spouse. Even after conditioning on certain case aspects, there is evidence of non-random assignment of Revenue Officers, raising concerns about bias in the estimates.

Keywords: Tax administration; debt collection.

JEL classification: H20

Boning: Office of Tax Analysis, U.S. Department of the Treasury (corresponding author, email: William.Boning@treasury.gov); Slemrod: University of Michigan; Stuart: The University of Sydney; Turk: Internal Revenue Service. We are indebted to Kelly Bellomo, Dean Chavez, John Guyton, Ron Hodge, Mike Martyn, Alicia Miller, and Stacy Orlett for many helpful conversations about the collections process. We thank Sutirtha Bagchi, Andreas Hagemann, Tom Hertz, Charles Loeffler, Janet McCubbin, Mike Mueller-Smith, and participants at the 2018 NTA Annual Conference on Taxation for helpful comments and discussions. We are particularly grateful to Ugo Troiano, who contributed significantly to the early stages of this project. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors and do not necessarily reflect the views or the official positions of the U.S. Department of the Treasury or the Internal Revenue Service. Any taxpayer data used in this research was kept in a secured Treasury or IRS data repository, and all results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

Optimally allocating limited tax agency resources is crucial for tax administration, but obtaining credibly identified estimates of the effects of enforcement policies to inform this resource allocation is fraught with difficulty. Tax agency resources are generally targeted toward cases that the tax agency believes are likely to generate revenue, either immediately due to collections or eventually via expanding perceived (and actual) deterrence of future noncompliance. Because of the targeting of resources, there is a problematic selection bias: the targeted taxpayers are different on important dimensions, and so their future behavior is arguably different from a control group of untargeted taxpayers even absent any policy intervention.

In recent years, statistical methodologies designed to overcome such problems have been applied to tax enforcement. The most prominent example is the partnership of governments and academic researchers to implement randomized controlled trials of enforcement policies, in which an untreated group of taxpayers is, on average, identical to the treated group, thus comprising a natural control group that allows unbiased estimation of treatment effects. This burgeoning literature is surveyed in Slemrod (2019).

Another modern technique that holds promise to understand the causal effect of tax enforcement initiatives is what is known as the "judges model" or "examiner assignment design." The idea is that individuals assigned to adjudicate cases vary in their inherent tendency for leniency. If cases are assigned randomly to examiners, then there is random variation in how leniently the cases are handled. While this variation will not (and should not) impact the final result in most cases, it could change the outcome for marginal cases. In other words, cases where a "yes" or "no" decision is on the margin could be decided by the assigned examiner's inherent tendency toward leniency. If one has a measure of this inherent tendency, it can be used to identify the local average treatment effect of the enforcement action. This research design, which utilizes operational rather than randomized experimental data, has been used to estimate local average treatment effects for a wide variety of policy interventions, including receipt of Social Security Disability Insurance (Maestas et al. 2013), bankruptcy protection (Dobbie and Song 2015; Dobbie et al. 2017), incarceration (Kling 2006; Loeffler 2013; Mueller-Smith 2015), pre-trial detention (Dobbie et al. 2018), and eviction (Humphries et al. 2019).

Whether this research design can provide credible causal estimates in settings such as tax enforcement rests critically on whether the examiners are in fact randomly assigned (as in Mueller-Smith (2015)), or at least randomly assigned conditional on observed variables. To shed light on this question, we apply the examiner assignment design to one important example of IRS intervention in tax debt cases: whether Revenue Officers designate cases as Currently Not Collectible (henceforth CNC). When working on a case of outstanding tax debt, if a Revenue Officer determines that the taxpayers income and assets are insufficient to meet basic living expenses, the IRS can pause collection efforts by designating the debt CNC. Whether to designate a given case CNC is a decision that rests with the judgment of the assigned Revenue Officer.

In this paper, we apply the examiner assignment design to learn about the effects on future taxpayer behavior of having their debt designated CNC. We use variation in Revenue Officers' propensities to designate similar cases CNC to study how suspending collection affects taxpayer behavior. Our approach is a version of the examiner assignment design developed in Doyle (2007) and Doyle (2008), which used quasi-random appointment of child protection investigators to cases in order identify the causal effects of foster care placement. We gauge a Revenue Officer's propensity to designate a case CNC using a residualized "leave-one-out" measure based on the proportion of their other cases the Revenue Officer has determined to be CNC, adjusted for any observable characteristics of the case that could alter whether a CNC designation is appropriate. This measure functions as an instrument for whether or not a case was actually deemed CNC, as it does not in any way depend on that case's characteristics. In other words, we compare cases where Revenue Officers who differ in their inherent leniency might disagree on whether the taxpayers hardship warranted the CNC designation.

After we construct our instrument, we ask the question: What is the impact on taxpayers of having their debt designated as CNC? How does it affect their future reported income, tax remittances, and tax return filing? Because debt collection is suspended primarily because of economic hardship, it is difficult to separate the effects of suspending collection efforts from the consequences of the economic hardship that might qualify a taxpayers case for suspension. This sample selection problem means that ours is a natural setting for using an examiner assignment research design if, but only if, the assignment of Revenue Officers to cases is sufficiently random for this design to be applied fruitfully.

At face value, it seems an examiner assignment design should work well in our setting. On paper, Revenue Officers are randomly assigned to cases. However, our conversations with case managers and Revenue Officers suggests that there are factors that influence case assignment, including the difficulty of the case and where the taxpayer is located. While we are able to control for most of the factors in case assignment that systematically affect case assignment (in theory leaving us with the random component to construct our instrument), tests of conditional random assignment suggest that our instrument does not remove all of the bias in naïve OLS estimates that ignore the inherent selection problem. The leave-one-out instrument based on the overall decision-making tendency of Revenue Officers is correlated with some pre-treatment covariates, conditional on observable factors that might influence case assignment, suggesting that the exclusion restriction may be violated. The imbalance is small in absolute terms, but even a small imbalance is cause for concern when interpreting the results of an instrumental variables approach.

We also pursue an instrumented difference-in-differences approach. This approach allows us to control for unobservable differences between the treatment and control groups that do not change over time, and for any time-invariant bias in our instrumental variables approach. We do not observe trends in the instrumented difference-in-difference estimates in periods before treatment for most outcomes, suggesting that these estimates may provide causal evidence of the effect of suspending collection efforts due to hardship on taxpayer compliance and income. However, our point estimates from this exercise are underpowered and largely insignificant, with confidence intervals that often include the naïve OLS estimate.

These two research designs lead us to conclude that suspending collection due to hardship causes taxpayers to report larger W-2 earnings in future years, which is driven by an increase in W-2 earnings by the spouse of the taxpayer with whom the case is associated. This result contrasts with estimates that ignore the sample selection issue, in which being designated CNC is associated with declines in W-2 earnings and tax filing behavior.

Overall, the examiner assignment design casts doubt on the impact that would be estimated from a model that ignores the sample selection problem often inherent in analyzing data regarding operational IRS enforcement, but does not generate tight point estimates of the true causal impacts. Perhaps most concerning is the correlation between the overall decision-making tendency of Revenue Officers and some pre-treatment covariates, a correlation that survives holding constant observable measures of factors that might affect Revenue Officer assignment to cases. Such correlation implies that the examiner assignment design generates estimates of the impact that retain some bias. Thus, in the CNC setting, Revenue Officers are not enough like judges who are randomly assigned cases by a computer for the examiner assignment model to generate reliably unbiased estimates. Further research is required to sharpen the conclusions that can be drawn from quasi-random, but not literally random, assignment of IRS personnel to tax cases.

2 The collections process

2.1 Tax debt

Each year, millions of U.S. taxpayers do not fully remit the taxes they owe, resulting in billions of dollars of uncollected tax debt. Attempts to collect this debt loom large in taxpayers' lives. If collecting debts from taxpayers whose economic resources are insufficient to meet basic living expenses reduces their incentive to work or invest, postponing collection could induce higher future collections.

On the other hand, collecting tax debts deters tax evasion by increasing the effective penalties for failing to report taxes. Taking into account whether tax debts are collected adds nuance to the classic Allingham and Sandmo (1972) model of tax evasion. While the original model assumes all debts are collected, this is not true in practice. If not all debts are collected, then changes in the collection rate affect both the expected present value of the tax remitted per dollar of reported income, and of the penalties remitted per dollar of unreported income discovered in an audit. Taxpayers choose both how much income to report and how much tax to remit (and when), given penalties both for underreporting and underpayment.

There is little evidence about the consequences of collection attempts writ large, including on how pursuing or forgoing collection affects taxpayers' future compliance and income. One exception is Miller et al. (2014), who find that, conditional on a later suspension of collection, assigning a case to an IRS field collection agent ("Revenue Officer") earlier in the collections process is associated with larger amounts collected. Other work on tax debt collection finds that letters highlighting financial penalties (Cranor et al. (2020)) or potential social stigma (Perez-Truglia and Troiano (2018)) modestly increase payments. Miller and Nikaj (2016) find that selling property tax liens to investors prompts higher payments, but less so in times of economic distress.

In this paper, we exploit an institutional feature of the debt collection process to provide evidence of the causal effects of tax collection on taxpayers' subsequent behavior. To provide causal evidence, we leverage the fact that the Internal Revenue Service (IRS) pauses collection when it determines that a taxpayer's income and assets are insufficient to meet basic living expenses, and designates their debt currently not collectible (CNC). The extent to which the IRS should grant such hardship designations is an open policy question; for example, the National Taxpayer Advocate listed among 2018's most serious tax administration problems the IRS's policy to not proactively provide hardship relief without taxpayer contact.

Because debt collection is suspended primarily on the basis of economic hardship, it is difficult to separate the effects of suspending collection efforts from the consequences of the economic hardship that qualifies a taxpayer's case for suspension. We use variation in IRS Revenue Officers' propensities to designate similar cases CNC to study how suspending collection affects taxpayer behavior. To our knowledge, this is the first time an examiner assignment design has been applied to study the causal impact of a tax enforcement policy.

We analyze de-identified administrative data linking over 123,000 tax debt cases to information about taxpayers' incomes, tax return filings, and contemporaneous information about the Revenue Officers who worked each case. We track three outcomes: payments on outstanding tax debt, future tax return filing, and W-2 earnings. We estimate local average treatment effects of tax collection suspensions due to hardship stemming from variation in Revenue Officer assignment. In other words, we compare cases where Revenue Officers who differ in their inherent leniency might disagree on whether the taxpayer's hardship warranted the CNC designation.

2.2 Delinquency designation

The IRS sends taxpayers who fail to file or to remit their known tax liability a bill notifying the taxpayer of their outstanding debt.[1] These notices are the beginning of the collection process.[2] Many taxpayers contact the IRS after receiving a notice in order to pay off some or all of the debt, to dispute it, or to explain that they are unable to pay. The IRS may negotiate an installment agreement or extend the due date for a taxpayer, depending on the taxpayer's circumstances. A small fraction of taxpayers settle their outstanding tax debt for less than the amount they owe through the Offer in Compromise (OIC) program or through

[1]Third-party information reports provide information on the income of taxpayers who fail to file.

[2]This process is outlined at https://www.irs.gov/newsroom/the-collection-process and in Appendix A.

"partial pay" installment agreements.[3]

When taxpayers do not respond to the initial (or subsequent) notices, their accounts are considered delinquent. At the end of Fiscal Year (FY) 2017,[4] there were approximately 14 million delinquent accounts with a total of over $131 billion owed in taxes, penalties, and interest (Internal Revenue Service (2017), Table 16). Delinquent accounts may be handled either by a call site (via the Automated Collection System (ACS)) or by a field office. In the first case, ACS personnel will try to contact the taxpayer by correspondence and by phone to negotiate a payment solution.[5] In the second case, a Revenue Officer from a local collection field office will work with the taxpayer to try to resolve the outstanding debt. Accounts that are assigned to ACS may ultimately be transferred to a field office if its attempts to resolve the debt are unsuccessful.

2.3 Revenue Officer assignment

Our understanding of the process by which cases are assigned to Revenue Officers is based on the Internal Revenue Manual (IRM), which specifies IRS administrative procedures, as well as discussions with two group managers and a former Revenue Officer. These conversations highlighted that many factors play a role in case assignment, including some based on professional judgment, although the Internal Revenue Manual (1.4.50.10 9) explicitly states that "[p]rofessional judgment should play a limited role in case selection."

The process proceeds as follows. A delinquent case is assigned to the local collection field office geographically closest to the taxpayer. The group manager of the office then assigns the case to a Revenue Officer. The first factor in case assignment is the grade of the

[3] In 2017, taxpayers made about 62,000 Offers in Compromise, of which the IRS accepted only 25,000 (Internal Revenue Service (2017), Table 16). These 62,000 offers are less than one half of one percent of all delinquent cases at the beginning of 2017. One condition of the IRS accepting less than the outstanding tax liability is that the taxpayer must remain compliant. "Partial pay" installment agreements occur when the taxpayer reaches an installment agreement with the IRS where the 10-year statute of limitations will run out on some of the debt before the taxpayer remits it. After the statute of limitations expires, the taxpayer is no longer responsible for the debt.

[4] The fiscal year for the IRS is the same as the United States federal government: Oct. 1 - Sept. 30.

[5] The IRS does not contact taxpayers by phone without first attempting to contact them by mail. See https://www.irs.gov/newsroom/phony-irs-calls-increase-during-filing-season.

case (9, 11, 12, or 13), which reflects the expected difficulty of closing the case. The case's priority is another factor in case assignment. Group managers are required to assign high priority cases before lower priority cases. High priority cases include those with particularly high balances, or where the statute of limitations for the debt is about to run out. Revenue Officers are more likely to receive new cases in zip codes where they already have active cases.[6] The group manager also has access to other aspects of the case that may impact assignment. These include the taxpayer's history of interacting with the IRS and factors influencing how likely the individual will be able to remit the debt (e.g., older taxpayers have fewer working years ahead of them and therefore may have lower potential earnings).

Case assignment also depends on the General Schedule (GS) grade of the Revenue Officer and the current inventory of the Revenue Officer.[7] Generally speaking, grade 09 cases are assigned to GS grade 09 Revenue Officers, grade 11 cases are assigned to GS grade 11 Revenue Officers, and so on. Revenue Officers with fewer workload-adjusted cases in their current inventories are more likely to receive new cases. Average caseload varies with GS grade. GS grade 09 Revenue Officers have an average case inventory of 72 cases. For GS grade 11 Revenue Officers, the average inventory is 61 cases. GS grade 12 and 13 Revenue Officers have average caseloads of 39 and 37 cases, respectively.

Once a case is assigned to a Revenue Officer, the Revenue Officer contacts the taxpayer, conducts research to ascertain the taxpayer's ability to make payments toward their outstanding debt, and takes steps to close the case.[8] Depending on IRS guidelines and the Revenue Officer's judgment, possible steps range from seizures, liens, and levies to full payments, agreements in which the taxpayer will repay the debt in installments, and designating the case CNC, which suspends collection efforts. More than two-thirds of cases close through

[6] Instances when professional judgment is appropriate in case selection include efficiently allocating the resources of the Service (e.g., geographically clustering cases to minimize Revenue Officer driving time) and addressing "the developmental needs of the Revenue Officer" (i.e., making sure the officer is exposed to a variety of case types to develop expertise in order to meet standards for promotion).

[7] The General Schedule (GS) Pay Scale is a wage schedule for federal employees. Revenue Officers may have a GS grade of 05, 07, 09, 11, 12, or 13. Grades 05 and 07 are related to training. Starting in 2017, the IRS discontinued use of the GS 09 grade for Revenue Officers.

[8] A case is closed when when a Revenue Officer is no long actively working the case.

full payment, an installment agreement, or a CNC designation, as Table 1 shows. Cases may be deemed CNC for a variety of reasons, including hardship and inability to locate the taxpayer. In some cases, a taxpayer may pay off some portion of their debt and the remaining debt may be deemed CNC, for example, due to hardship.

Table 1: Distribution of how cases were closed

Description	Number of cases	Percent of cases
Installment agreement	49,041	34.1%
CNC	43,783	30.5%
Other	40,396	28.1%
Full pay	10,443	7.3%
Total	123,396	100.0%

Notes: Includes all cases that meet our sample restriction criteria (discussed in Section 3) between November 2014 and December 2018. Other ways a case might be closed include Offers in Compromise, abatement, payment tracer (which resolves situations of missing and misapplied payment issues), and being flagged for handling outside of the usual system and procedures.

2.4 Currently Not Collectible (CNC)

The vast majority of CNC cases for individual taxpayers are either "hardship" (49%) or "unable to locate" (39%).[9] Per the Internal Revenue Manual, "[a] hardship exists if a taxpayer is unable to pay reasonable basic living expenses" (IRM 5.16.1.2.9(1)). Outstanding tax liability that is designated CNC is still legally owed to the United States government.[10] When a taxpayer's debt is deemed CNC, the IRS sends the taxpayer a letter notifying them of the change in status and reminding them that the taxpayer must remit future tax liability. Changes in the taxpayer's circumstances may cause the IRS to re-evaluate whether a CNC designation is appropriate.[11] For example, when taxpayers with unpaid tax that

[9]Most of the remaining 12% of CNC cases are due to the taxpayer being deceased.

[10]The statute of limitations on unremitted debt is ten years, after which the IRS may no longer pursue the debt.

[11]Receipt of future returns from the taxpayer automatically initiates a review to see whether the taxpayer's circumstances have changed (e.g., a change in address, income, a new levy source). If there has been a change, the case can be re-activated and the CNC status may be revoked. When this happens, the case will either be sent back to ACS or directly to a group manager's queue. The taxpayer is not directly informed that their case has been re-activated, but the taxpayer may receive, e.g., a letter from ACS or a notice of a levy,

is designated CNC file returns, the designation may be rescinded if the taxpayer's income becomes sufficiently large. Interest and penalties continue to accrue on the outstanding balance while collection efforts are suspended. If the taxpayer is owed a refund on a future tax return, the IRS will retain the refund to offset the tax debt in CNC status. Even if a taxpayer has outstanding tax debt in CNC status, new unpaid tax undergoes the notification process described in Section 2.2.

A priori reasoning does not lead to clear predictions about the direction of changes in tax reporting and compliance behavior by those whose debt is designated CNC. A CNC designation suspends efforts to collect some or all outstanding tax debt, such as attempts to garnish wages and assets, and thereby reduces the incentive for the taxpayer to voluntarily repay outstanding debt. In addition, a CNC designation signals that the IRS can be lenient, and might lead the taxpayer to conclude that future noncompliance will also be met with leniency. This reduction in the expected costs of noncompliance has a negative effect on compliance. In contrast, maintaining a CNC designation requires that a taxpayer comply with the requirements to file tax returns and remit tax payments for new tax years' liability. Failure to comply with these requirements leads to the resumption of active collection, resulting in a positive effect on future compliance from a CNC designation.

The effect of a CNC designation on the incentive to earn income is also theoretically unclear. Having some or all tax debt designated CNC provides a nonnegative wealth effect (i.e., it is "good news" from the taxpayer's perspective). A permanent CNC designation would reduce the taxpayer's expected tax rate, because future earnings and assets will not be garnished to meet the tax debt. However, a CNC designation is not permanent (unless the 10-year statute of limitations is reached), and the IRS retains the right, if the taxpayer's financial situation improves sufficiently, to put the debt back in active collection status. Thus, a taxpayer may perceive that his or her marginal tax rate is reduced by the CNC designation for future incomes up to some unknown level that triggers the CNC status revocation, above

which would indicate that the IRS has resumed pursuit of the outstanding balance.

which the marginal tax rate absent the CNC designation is restored. The combination of a negative effect on future earnings from the wealth effect and a generally positive substitution effect means that the net effect on earnings is of indeterminate sign.

Consequently, both the magnitude and the sign of the impact of a CNC designation on future behavior can only be determined via empirical analysis. We turn next to this task.

3 Data and sample frame

To assess the effects of decisions made during the collection process, we combine information from the collection case management system with caseload information about Revenue Officers and administrative taxpayer information, including a history of filing status (i.e., whether or not the taxpayer filed a return), monthly data on unpaid tax liability, and information from individual tax returns. Data on Revenue Officers include their GS grades, inventory levels at the end of each month, and their decisions in past collection cases. We supplement the taxpayer data with information about local allowable living expenses and whether they live in an urban or rural area.[12]

We focus on individual taxpayer cases and therefore exclude business entities other than sole proprietorships. We begin with about 283,000 individual taxpayer cases. Approximately 123,000 taxpayers remain after excluding cases where we suspect case assignment may not have been random,[13] cases designated CNC because the Revenue Officer was unable to locate the taxpayer, and cases worked by Revenue Officers who handled fewer than 20 cases between November 2014 and December 2018.[14] We perform this final exclusion to ensure that we observe sufficient decisions made by each Revenue Officer to obtain a reliable estimate of their tendency to designate a case CNC, what we call their "leniency." Figure 1 shows the

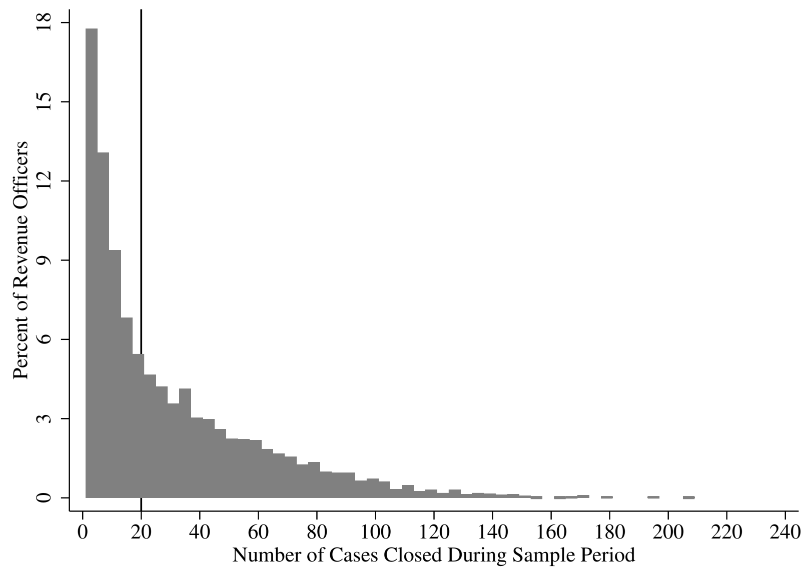
[12] Appendix B explains the construction of these variables and the outcome variables in detail.

[13] We discuss situations when case assignment may not be random in Section 3.1.

[14] The mean (standard deviation) fraction of cases designated CNC is 0.24 (0.24) among Revenue Officers who worked fewer than 20 that meet our selection criteria, and 0.30 (0.12) among Revenue Officers who worked at least 20 cases that meet our selection criteria. As a robustness check, we repeat our analysis using alternative cut-offs of 10 cases and 30 cases, and the results are qualitatively unchanged. Appendix C shows the IV results for cut-offs of 10, 20, and 30 cases.

distribution of the number of cases closed by a Revenue Officer between November 2014 and December 2018. Out of 4,808 Revenue Officers included, 2,345 (49%) meet our sample restriction criteria.

Figure 1: Distribution of number of cases closed per Revenue Officer over the sample period



Notes: 4, 808. Includes all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018.

3.1 Field data

To assemble our sample for analysis, we begin with the universe of cases with an assessed balance due assigned to a field collection Revenue Officer between November 2014 and December 2018. The data include the numeric ID of the Revenue Officer who worked each case, the date the case was assigned to field, the date the case was assigned to the Revenue Officer, and the date the case was closed. We drop cases where it appears that the Revenue Officer reported working zero hours on that case, cases that were not closed, cases where the taxpayer lives outside of the United States or on a military base, and cases flagged as handled outside the usual system and procedures. Because we are ultimately interested in the impact of a CNC designation on future outcomes, we drop cases designated CNC because the Revenue Officer was unable to locate the taxpayer.

In addition, we exclude certain categories of cases in which the Revenue Officer's manager might exercise deliberate discretion in the assignment process. These categories were identified based on interviews about the case assignment process with field collection group managers and an individual who worked for several years as a Revenue Officer. Cases where the statute of limitations is close to expiration are dropped,[15] as these cases may be assigned to a Revenue Officer who works quickly. We use only the first time we observe a case being assigned within a given group, which eliminates cases where there is already an existing case involving the same taxpayer. As a result, individuals only appear once in our final sample.

New cases in a group manager's queue that involve the relevant taxpayer in an ongoing case are usually assigned to the Revenue Officer working the existing case. This also removes cases that have been returned to the group's queue. The group manager is better able to anticipate the type of work the case will require when cases return to the group's queue, particularly if the case is an installment agreement default or a review of a previous CNC decision, which may influence the assignment decision.[16] Finally, we exclude cases that are reassigned to a different Revenue Officer within the group (or in a different group), as reassignment may reflect unobserved characteristics of the case. In contrast, the group manager assigns cases that the group has not seen before on the basis of limited information about the case and the group's Revenue Officers that is observable in administrative databases.

3.2 Revenue Officer data

The Revenue Officer data includes monthly-level information about factors used to determine the workload to assign to the Revenue Officer, including the Revenue Officer's GS grade, monthly inventory and inventory by case grade, and the fraction of the Revenue Officer's case load that is above their GS grade level. We match this data to the field data using the

[15] We consider the statute of limitations close to expiration if the expiration date is within a year.

[16] As is standard in literature using examiner assignment to identify causal estimates, we do not exclude individuals who appear more than once in our data but are assigned to a different local IRS office, or "group," who, because they are in a different location, may not have a previous history with the taxpayer. Less than 3% of our sample appear in the data more than once.

Revenue Officer id number and relevant month: if a case in the field data was assigned to a Revenue Officer in, e.g., May 2017, we match the Revenue Officer's information from the previous month.

We make two additional exclusions from the data based on the Revenue Officer data. First, we exclude Revenue Officers whose GS grade is listed 5 or 7, which are training grades. Second, we exclude group-month combinations when only a single Revenue Officer in the group is assigned any cases. This second exclusion is necessary because there will be no variation in proclivity to designate a case CNC in that group during that month.

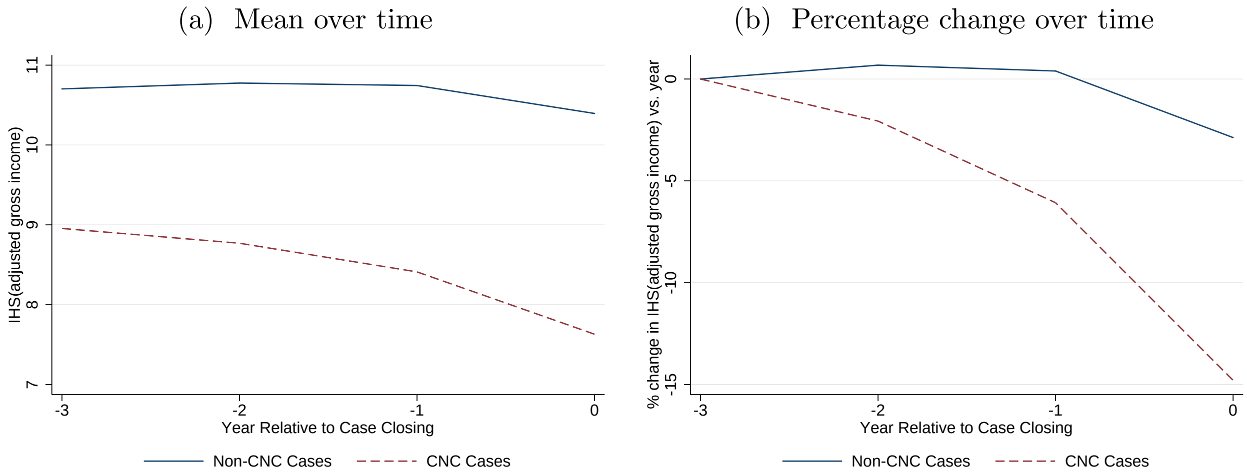
3.3 Taxpayer data

Our taxpayer panel spans January 2009 to December 2018. This panel contains the taxpayer's year of birth, outstanding tax debt and payments, information from annual tax returns, and information about the taxpayer's and spouse's income from third-party information reports. We measure behavior at the level of the taxpaying unit, combining the third-party-reported values for the taxpayer and, if the taxpayer filed a joint tax return in the year the case closed, the spouse listed on that return. The measures of tax compliance we construct are payments toward outstanding tax debt and an indicator for filing a return conditional on having third-party-reported earnings. We measure income using total W-2 earnings as well as W-2 earnings broken out by whether the taxpayer or their spouse was the earner. We take the inverse hyperbolic sine transformation of each of our measures of income. We then merge this panel with the field and Revenue Officer data described above.

4 Identification strategy

The fundamental challenge in assessing the causal effect of a CNC designation on future taxpayer outcomes is disentangling the effect of the CNC designation itself from the effects of the hardships that might lead to such a designation. For example, a sole proprietorship might become unprofitable, fail to remit taxes on the prior year's income, and then liquidate, leaving

Figure 2: IHS(adjusted gross income) over time (relative to year case closed)



Notes: N = 138,452. Panel (a) shows the mean of the inverse hyperbolic sine (a transformation similar to the natural logarithm) of adjusted gross income over time across two subgroups: tax units whose cases were designated CNC in dashed red and tax units whose cases were not designated CNC in solid blue. The x-axis indicates the year relative to when each case closed, e.g., -1 is the year before the taxpayer's case closed. Panel (b) shows cumulative percentage changes in this measure relative to the value three years before the case closed. Adjusted gross income adjusted for inflation to 2017 values.

the proprietor with substantial tax debt and little income. In this example, determining the effect of a subsequent CNC designation is complicated by the effect of the business failure itself. Taxpayers whose cases are designated CNC experienced substantially larger decreases in adjusted gross income in the years leading up to their cases being closed (shown in Figure 2). This means that taxpayers whose cases are not designated CNC are, as a whole, a poor comparison group for taxpayers receiving a CNC designation. In what follows, we address this sample selection issue with an instrumental variable for being designated CNC based on the propensity of the Revenue Officer assigned to a case to designate other cases CNC as an instrument for CNC designation.

4.1 Source of variation and instrument

Variation in CNC status across otherwise similar cases comes from randomness in which Revenue Officer is assigned to a case, combined with differences in Revenue Officers' propensity to designate cases CNC. An intuitive method for estimating Revenue Officer j's propensity to designate cases CNC would be to determine what proportion of cases Revenue Officer j designated CNC. Simply using this measure to predict whether Revenue Officer j designated case i as CNC would be biased, however, since case i would have been included in the calculation. To avoid this bias, we construct a case-specific instrument using a "leave-one-out" measure of the propensity of Revenue Officer j to designate case i CNC.[17] Specifically, the "leave-one-out" measure is

$$Z_{ij}^{S} = \frac{\sum_{k=1}^{n_j} CNC_{kj} - CNC_{ij}}{n_j - 1}. \tag{1}$$

The numerator of this expression is the number of cases designated CNC by Revenue Officer j, less one if case i was designated CNC (i.e., $CNC_{ij} = 1$). The denominator is the total number of cases handled by Revenue Officer j less one.

As described previously, case assignment is not entirely random. Which cases are assigned to Revenue Officers may depend on case grade, case priority, taxpayer characteristics, and geographic considerations. For example, if higher grade cases are less likely to receive a CNC designation, then failing to control for case grade and Revenue Officer grade would make it seem like higher GS grade Revenue Officers were stricter. To address this issue, we follow Dobbie et al. (2018) and develop a residualized leave-one-out measure of the propensity to designate a case CNC that removes variation from our instrument that is driven by observable determinants of case assignment that may also impact CNC designation. As a result, our instrument is a measure of the Revenue Officer's tendencies controlling for the features of the case that may be correlated with a CNC designation. We discuss the empirical importance of residualization and potential pitfalls in Section 4.1.1.

We calculate the residualized leave-one-out instrument as follows. First, we regress true CNC status for each case on Revenue Officer and case characteristics measured before

[17] This version of Jackknife IV (Angrist et al. 1999) is used in, e.g., Doyle, Jr. (2007), Doyle, Jr. (2008); Maestas et al. (2013); Dobbie and Song (2015); Dobbie et al. (2017). Split-sample two-stage IV (Angrist and Krueger 1995) and limited-information maximum likelihood approaches can also address this bias.

assignment that may impact case assignment:

$$CNC_{ij} = \beta_0 + \beta \boldsymbol{X}_{ij} + u_i, \tag{2}$$

where $\boldsymbol{X}_{ij}$ includes an indicator for whether or not the case is high priority, case grade fixed effects, and case characteristics (the taxpayer's year of birth, estimated ability to pay,[18] an indicator of whether the taxpayer has had their case "in the field" since 2009, and indicator variables for whether the oldest debt associated with the case is more than 12 months old or more than 36 months old.)[19]

We include fixed effects for the geographic group of the assigned Revenue Officer.[20] To see why this might matter, consider two groups. Group A works cases in affluent Township A whereas Group B works cases in low-income Village B. We might expect that the cases in Village B are more likely to face a true financial hardship, and therefore more likely be given a CNC designation. We want to make sure the residual propensity to designate a case CNC reflects characteristics of the Revenue Officer rather than characteristics of the taxpayers who live where the Revenue Officer's cases take place. We also control for two relevant geographic variables: allowable living expenses and an urban indicator. In robustness checks, we also include total inventory and case grade by Revenue Officer GS grade fixed effects, which does not qualitatively change the results.

We then use the results of estimating Equation 2 to predict CNC status:

$$\widehat{CNC_{ij}} = \hat{\beta}_0 + \hat{\beta} \boldsymbol{X}_{ij}, \tag{3}$$

[18] "Estimated ability to pay" is the ratio of an estimated future income based on their previous adjusted gross income and age, and their outstanding balance with the IRS in the year before assignment.

[19] Appendix B explains how we derived the specific variables used in our analysis and provides sample statistics by CNC status.

[20] Groups are assigned cases from particular ZIP Codes in their local area. We are unable to control directly for ZIP Code fixed effects because there are too few cases per zip code.

and calculate the residual value of CNC_i^R by subtracting $\widehat{CNC_{ij}}$ from CNC_i:

$$CNC_{ij}^R = CNC_{ij} - \widehat{CNC_{ij}}. \tag{4}$$

The case-specific, residualized leave-one-out measure of Revenue Officer j's propensity to designate case i as CNC is computed using Equation 5:

$$Z_{ij}^R = \frac{\sum_{k=1}^{n_j} CNC_{kj}^R - CNC_{ij}^R}{n_j - 1}. \tag{5}$$

The numerator of this expression is equal to the sum of the residualized CNC designation for all cases covered by Revenue Officer j less the residualized CNC designation for case i. The denominator is equal to the total number of cases handled by Revenue Officer j less one. Intuitively, this measure indicates Revenue Officer j's residual propensity to designate cases as CNC (excluding case i), holding constant observed characteristics of the case. Conditional on observable characteristics and assuming conditional random assignment, Z_{ij}^R should be correlated with the decision in case i only if Revenue Officer j has a threshold level of leniency specific to case i.

Notably, the distribution of the residualized instrument is less dispersed than that of the simple leave-one-out instrument, with a standard deviation of 0.0826 versus 0.1119. This suggests that the intrinsic variation in Revenue Officers' leniency is less than the simple average of average case designations would suggest. This is illustrated in Figure 3, where Panel (a) shows how much more variation there is in the value of the simple leave-one-out instrument compared to the residualized leave-one-out instrument in Panel (b). The figure plots a slightly adjusted version of the instrument to include each Revenue Officer only once.[21]

[21]Panel (a) shows the fraction Z_j^{S*} of cases designated CNC by each Revenue Officer j, where

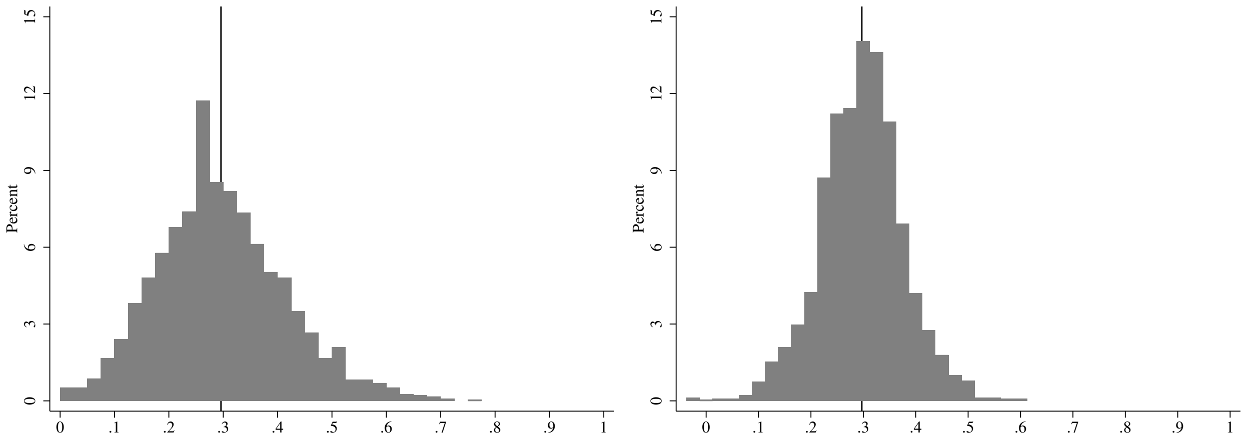
$$Z_j^{S*} = \frac{\sum_{n_j}^{k=1} CNC_{kj}}{n_j}. \tag{6}$$

Z_j^{S*} is equivalent to Z_{ij}^S (calculated in Equation 1) without omitting case i. Similarly, Panel (b) shows the

Figure 3: Distribution of the fraction of cases designated CNC by Revenue Officer

(a) Distribution of the fraction of cases designated CNC by Revenue Officer (Z_{ij}^{S*})

(b) Adjusted distribution of the fraction of cases designated CNC by Revenue Officer (Z_{ij}^{R*})



Notes: Unique at the Revenue Officer level ($N = 2,345$). Constructed using all cases closed by Revenue Officers between November 2014 and December 2018 who closed at least 20 cases that meet our sample restriction criteria during that time period. Panel (a) shows the fraction Z_{ij}^{S*} of cases designated CNC by each Revenue Officer, where Z_{ij}^{S*} is defined in Equation 6. Panel (b) shows the adjusted fraction R_{ij}^{S*} of cases designated CNC after accounting for residualization, as defined in Equation 7.

4.1.1 Tests of instrument exogeneity

Using Revenue Officer assignment to instrument for CNC status in order to measure its causal impact relies on the assumption that Revenue Officer assignment is uncorrelated with future outcomes, conditional on included covariates. This assumption allows for assignment of Revenue Officers based on certain observable characteristics (such as case grade) but, once those observable characteristics are controlled for, Revenue Officer assignment should be random. As mentioned previously, group managers are explicitly instructed that "[p]rofessional judgment should play a limited role in case selection" (IRM 1.4.50.10 9).

Our discussions with group managers underlined certain aspects of cases that might

adjusted fraction Z_j^{R*} of cases designated CNC by each Revenue Officer j, where

$$Z_j^{R*} = \frac{\sum_{n_j}^{k=1} \widehat{CNC_{ij}}}{n_j}. \tag{7}$$

Z_j^{R*} is equivalent to $Z_j^{S*} - \frac{1}{n_j}\sum_{n_j}^{k=1} CNC_{kj}^R$, where CNC_{ij}^R is calculated as in Equation 4. We plot Z_j^{R*} rather than $\frac{1}{n_j}\sum_{n_j}^{k=1} CNC_{kj}^R$ to facilitate a visual comparison: the resulting object has the same mean as true CNC status, as is illustrated by the vertical lines in the figure.

result in group manager professional expertise playing more than a limited role role in case assignment. Some of these factors are not included in the case data and therefore cannot be included in the creation of our instrument, raising the concern that the instrument may not be uncorrelated with future outcomes. We test whether Revenue Officer assignment is conditionally random by comparing pre-assignment case and taxpayer characteristics across Revenue Officers with high and low values of the average case designation. We regress pre-assignment characteristics on, in turn, an indicator for CNC designation, the simple instrument (Z^S_{ij}, defined in Equation 1), and the residualized instrument (Z^R_{ij}, defined in Equation 5), following Equation 8:

$$Y_i = \eta_0 + \eta Z_i + u_i, \tag{8}$$

where Y_i is the pre-assignment outcome for taxpayer i and Z_i is the value of the relevant CNC indicator or instrument.

We test for balance using five pre-assignment characteristics of the case:[22] model score (an estimate of how likely the taxpayer is to remit their outstanding liability), the taxpayer's average earnings (from Forms W-2 provided by employers), an indicator for whether the taxpayer filed a return in the year before their case was assigned to a Revenue Officer, and, if the taxpayer did file a return, the adjusted gross income reported.[23] We transform W-2 earnings by taking the inverse hyperbolic sine, which can be interpreted like a log transformation and enables us to include non-positive values.

[22]We cannot use the variables included in the creation of our residualized leave-one-out instrument, all of which were chosen based on our reading of the Internal Revenue Manual and our conversations with group managers and an ex-Revenue Officer. The five variables we test for balance are ones that were never explicitly mentioned as impacting case assignment. While we have no reason to think they directly affect case assignment, we understand that they may be correlated with things that do directly affect case assignment and therefore cannot rule out that they are indirectly associated with case assignment.

[23]One potential concern arises with using average adjusted gross income to calculate our estimate of the taxpayer's ability to pay off their balance, one of the variables used in the residualization process. This means that adjusted gross income appears on both the left- and right-hand sides of the regression for two of our balance tests. Any bias this step in the residualization might introduce is of minimal concern because the coefficient on these outcomes is statistically insignificant regardless of whether or not we residualize the instrument.

Table 2: Test of instrument exogeneity

	Model Score	Ave. W-2 Wages (IHS)	Ave. AGI (IHS)	AGI before assignment (IHS)	Filed before assignment
OLS	-0.065^{***}	-1.913^{***}	-1.505^{***}	-2.096^{***}	-0.160^{***}
SE	(0.001)	(0.037)	(0.030)	(0.037)	(0.003)
F-stat	2,263.104	2,744.591	2,527.897	3,264.277	2,369.145
IV (Simple)	-0.161^{***}	-2.057^{***}	-1.245^{***}	-1.727^{***}	-0.167^{***}
SE	(0.012)	(0.225)	(0.240)	(0.265)	(0.018)
F-stat	188.746	83.554	26.858	42.613	89.157
IV (Resid.)	-0.033^{*}	-0.822^{**}	-0.372	-0.569^{*}	-0.040
SE	(0.014)	(0.274)	(0.224)	(0.256)	(0.024)
F-stat	5.679	9.022	2.767	4.925	2.853
N	91,112	120,753	117,289	116,803	120,754

Notes: Standard errors, in parentheses, are clustered at the Revenue Officer level. Includes all cases that meet our sample restriction criteria between November 2014 and December 2018. Limited to cases worked by Revenue Officers who closed at least 20 cases that met our sample restriction criteria between November 2014 and December 2018. Variables are defined in Appendix B. All income variables are adjusted for inflation to 2017 values. Values for average W-2 earnings, average adjusted gross income, and and adjusted gross income in the year before case assignment are given using the inverse hyperbolic sine transformation. We measure adjusted gross income and W-2 earnings at the level of the taxpaying unit: if a taxpayer filed a joint tax return during the year in which their case was closed, we add the value for the taxpayer's spouse to the taxpayer's own value. If the taxpayer did not file a joint tax return in that year, the outcome value is equal to the value for the taxpayer alone.
*** Significant at the 0.1% level; ** significant at the 1% level; * significant at the 5% level.

As one would expect, cases that were designated not collectible had lower model scores (indicating a lower estimated probability of collection), lower W-2 earnings, lower adjusted gross income, and were less likely to file before assignment. These characteristics reasonably led Revenue Officers to be more likely to designate taxpayers' debt as currently not collectible. These associations, shown in the top panel of Table 2, are statistically significant at the 0.1% level.

The simple leave-one-out average of Revenue Officer CNC designations is designed to eliminate this sample selection problem, as it does not depend on the characteristics of the taxpayer in question. Notably, however, the second panel of Table 2 shows that the problem remains. All of the coefficients continue to be significant at the 0.1% level, and the magnitudes of the estimated coefficients do not change dramatically (the exception being the coefficient on model score, for which the absolute magnitude of the coefficient increases substantially). This indicates that cases are not randomly assigned to Revenue Officers, and their CNC designation tendencies are correlated with case characteristics.

The residualized Revenue Officer instrument reduces the imbalance substantially, but does not fully eliminate it. The magnitudes of the coefficients shown in the third panel of Table 2 for all five pre-treatment characteristics shrink considerably. The coefficients between the residualized instrument and model score, and between the residualized instrument and pre-assignment adjusted gross income, are statistically significant only at the 5% level. The estimated coefficient between the residualized instrument and average W-2 earnings is statistically significant at the 1% level.

Any violation of the exclusion restriction is cause for concern about biased estimates in the second stage coefficients.[24] Note, though, that these coefficients are small in economic terms. Assignment to a Revenue Officer whose value of the instrument is one standard deviation higher is associated with a model score that is, on average, 0.0027 lower, 6.7% lower W-2 earnings, and 4.7% lower adjusted gross income.[25] As the signs of the correlations with the residualized instrument are the same as the signs of the correlations with CNC status, one would expect any bias in the IV results to be toward the OLS estimates. Concerns about the remaining imbalance motivate our use of the difference-in-differences instrumental variables design detailed in Section 4.2.

4.2 Difference-in-differences instrumental variables design

Figure 2 shows that an OLS or uninstrumented difference-in-differences approach is invalid in this context, and Table 2 shows that an undifferenced instrumental variables approach does not resolve the problem. We next pursue a difference-in-differences specification in which the treatment is the instrumented value of CNC.[26] The advantage of the difference-in-differences

[24]For a brief discussion of how the standard IV assumptions apply in this context, see Appendix D.

[25]The standard deviation of our residualized instrument is 0.082 and the coefficient on average W-2 earnings in the balance test is -0.822: $-0.822 \times 0.082 = -0.067$ log points, about 6.7%. Similarly, the coefficient on adjusted gross income before assignment is -0.569 ($-0.569 \times 0.082 = -0.047$), and the coefficient on model score is -0.033 ($-0.033 \times 0.082 = -0.0027$).

[26]This approach is similar in spirit to, among others, Duflo (2001).

approach is that it removes any time-invariant bias in the instrumental variables estimates. It remains vulnerable to time-varying bias due to, for example, larger effects of the factors correlated with the instrumented value of CNC in periods after the CNC decision. Relative to difference-in-differences without instrumenting for the value of CNC, the instrumented approach avoids bias that would otherwise arise from correlations between omitted variables correlated with both CNC and trends in the outcome over time (but not correlated with the instrument).

We use the event-study version of the difference-in-difference model:

$$Y_{it} = \sum_k 1(t = k)[\phi_k \widehat{CNC}_i + \tau_k \boldsymbol{X}_{it} + \nu_k] + \iota_i + e_{it}. \tag{9}$$

where Y_{it} is the outcome variable, $\widehat{CNC}_i$ is the instrumented value of CNC, $\boldsymbol{X}_{it}$ are time-varying controls, ν_k are event-time fixed effects, ι_i is an individual fixed effect, and e_{it} is the error term. This approach allows us to test whether there is time-varying bias in the instrumental variables estimates by looking at the trends across time periods before the CNC decision. While a lack of trends before the CNC decision does not guarantee a lack of trends afterwards, the test still provides some information about the potential for bias.

As expected, in the first stage the instrument strongly predicts CNC status, with a point estimate of 0.489 that is statistically significant at the 0.1% level. The instrument easily passes the Cragg and Donald (1993) and Kleibergen and Paap (2006) tests of weak instruments.

5 Effects of Suspending Tax Debt Collection Due to Hardship

5.1 Research design choices

We now turn to the motivating research question: how suspending tax debt collection affects subsequent taxpayer compliance and income, explaining our definitions of a taxpaying unit and of event time before proceeding to test the pre-period trends and analyze effects after the CNC decision.

We measure behavior at the level of the taxpaying unit: if a taxpayer filed a joint tax return during the year in which their case was closed, we add the value for the taxpayer's spouse to the taxpayer's own value. If the taxpayer did not file a joint tax return in that year, the outcome value is equal to the value for the taxpayer alone. We include cases worked by Revenue Officers who closed at least 20 cases during our sample period.[27] We set the year before the year in which the case closed as Year 0. Coefficients are reported relative to Year 0. This ensures that all behavioral changes due to CNC status are properly included in the post period. For example, if a case closed in January of Year T, it may impact whether the taxpayer filed their Year T-1 tax return by April of Year T. One consequence of this approach is that behavioral responses may appear delayed: if a case closes in December, only a small fraction of Year 1 is post-case-closing, and the first full year in which CNC status can affect behavior is Year 2.

We estimate effects for years from two years before Year 0 (Year -2) to four years after Year 0 (Year 4). Years -2 and -1 allow us to visually assess whether or not the parallel trends assumption holds before treatment. For the outcomes we consider, coefficients in the years

[27] There is a critical trade-off in choosing this threshold. As shown in Figure 1, the number of Revenue Officers excluded from our analysis quickly increases as we raise the case count threshold. This reduction in sample size reduces the power of the estimates. Lowering the case count threshold means that we include Revenue Officers for whom we have a less precise measure of their latent tendency to designate cases CNC. Results are similar if we use alternative case count thresholds of 10 or 30 cases, although the standard errors are slightly different.

before treatment are not statistically significant.

5.2 Effects on compliance

If taxpayers whose cases are deemed CNC come to believe that their tax compliance is subject to less scrutiny than they had previously thought, they may be less compliant in the future. This reduced compliance could take the form of reduced payments to the IRS, or a lower probability of filing a tax return. On the other hand, if taxpayers try to avoid jeopardizing their CNC status and triggering renewed attempts to collect, they may be more compliant in the future.

5.2.1 Payments toward outstanding tax debt

Figure 4 shows changes in the dollar value of payments against outstanding tax debt made by taxpayers whose cases are designated CNC compared to taxpayers whose cases are not.[28] A standard difference-in-differences regression implies that payments from taxpayers whose cases are designated CNC fall substantially relative to other taxpayers with a decline that is largest in Year 2 before recovering partially. The IV regression results display a similar pattern in their point estimates but have 95% confidence intervals that include zero in each year, and are thus uninformative about whether payments towards outstanding tax debt decline following a CNC designation.

5.2.2 Filing tax returns

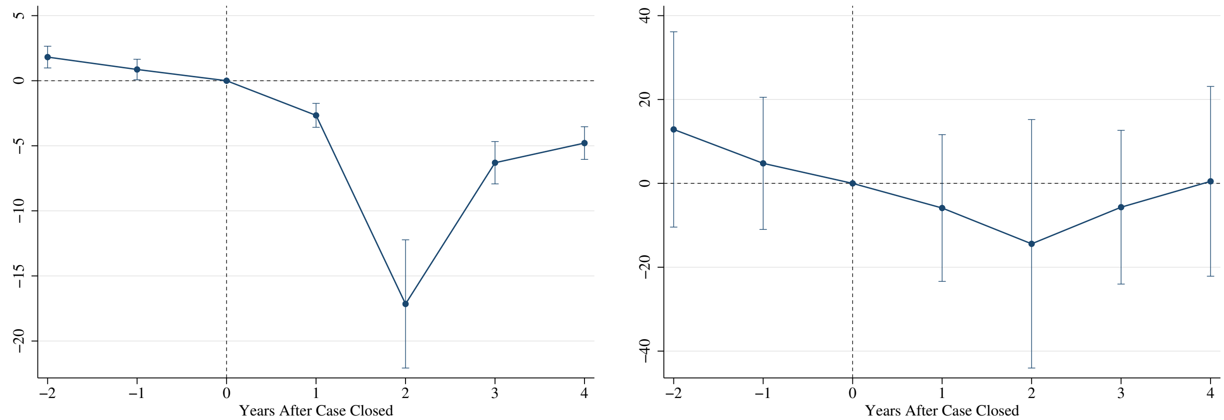
We also study the effect of a CNC designation on tax compliance as measured by filing a tax return. Taxpayers whose cases are assigned to the field for collection often have incomes below the threshold at which they would be required to file.[29] As a result, we examine filing

[28]This measure does not include payments against tax debt accrued after the case closed.

[29]For example, we observe that, even among tax units receiving earnings reported on Form W-2, about 40% had earnings below the filing threshold.

Figure 4: Effect of a CNC designation on payments toward outstanding tax debt ($1,000)

(a) Difference-in-differences (b) IV with residualized instrument



Notes: Number of cases is 139,246 for difference-in-differences and 121,923 for the IV specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Coefficients are shown in thousands. Payment values adjusted for inflation to 2017 values.

behavior among those with earnings reported on Form W-2, who are both more likely to be required to file and more likely to be detected if they fail to file.

The filing results are shown in Figure 5. While the standard difference-in-differences results suggest that a CNC designation leads to a reduced likelihood of filing, the IV results show that taxpayers who received a CNC designation were no less likely to file a return in the following years.

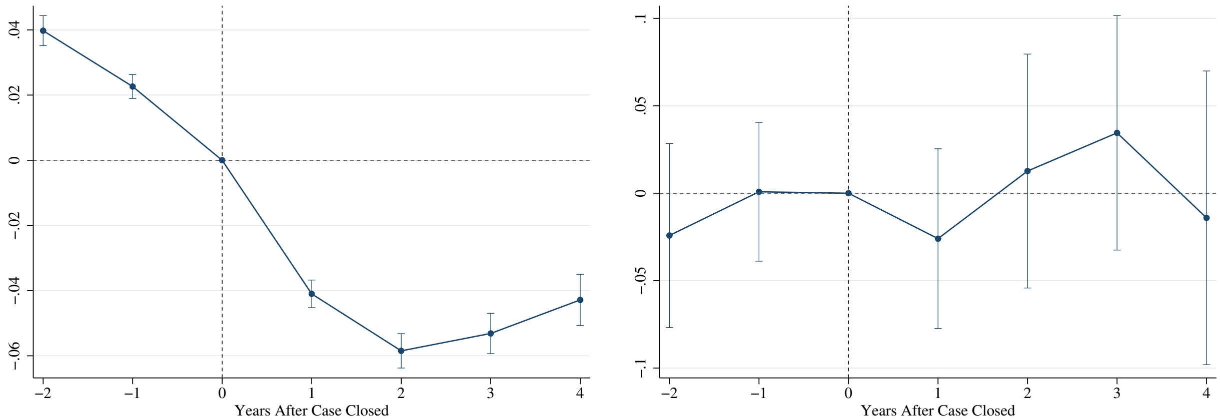
5.3 Effects on income

Suspending collection efforts due to hardship does not imply that a taxpayer's debt is forgiven. Taxpayers whose cases are given a CNC designation may use the relief to make investments to increase their future earnings, or increase labor supply with the expectation that the income they earn is less likely to go toward tax debts. On the other hand, a CNC designation could reduce labor supply through an income effect, as more of taxpayers' income becomes available for uses other than tax debt repayment.

We examine how a CNC designation affects W-2 earnings. We take the inverse hyper-

Figure 5: Effect of a CNC designation on filing a tax return

(a) Difference-in-differences (b) IV with residualized instrument



Notes: Number of cases is 139,246 for difference-in-differences and 121,923 for the IV specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018.

bolic sine of each of these measures of income.[30]

In standard difference-in-differences results CNC status is associated with large, lasting, and statistically significant decreases in W-2 earnings. Our instrumental variables difference-in-difference approach finds statistically significant increases in W-2 earnings as a result of a CNC designation. This increase is driven by an increase in W-2 earnings by the spouses of married taxpayers in years three and four after the case closes.

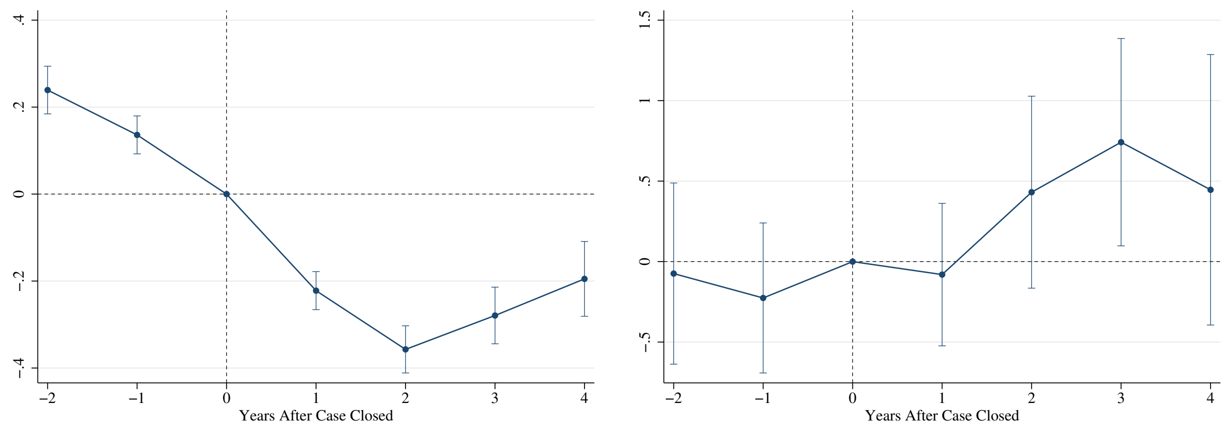
5.3.1 W-2 earnings

Figure 6 shows the effects of CNC designation on the sum of the taxpayer's and their spouse's earnings reported on Form W-2. The patterns in these results closely resemble the patterns for adjusted gross income: the standard difference-in-differences specification shows wage decreases after Year 0, while the instrumented difference-in-differences specification shows large increases in W-2 earnings. The increase in Year 3 is statistically significant at the 95%

[30]The inverse hyperbolic site transformation is: $y_{IHS} = \ln(x_i + (x_i^2 + 1)^{\frac{1}{2}})$. This is approximately equal to $\ln(2x_i) = \ln(2) + \ln(x_i) = 0.69 + \ln(x_i) \approx \ln(x_i)$ except for small values of $\ln(x_i)$. As a result, coefficients on these variables can be interpreted as we would for standard logarithmic dependent variables, but allows for the transformation of values equal to 0 and negative values. To transform the coefficients back into meaningful values, we use the approximation % change $\approx (\exp(\beta) - 1) * 100$ (see Bellemare and Wichman (2020) for more information about the inverse hyperbolic sine transformation).

Figure 6: Effect of a CNC designation on W-2 earnings (IHS)

(a) Difference-in-differences (b) IV with residualized instrument



Notes: Number of cases is 139,246 for difference-in-differences and 121,923 for the IV specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. W-2 earnings adjusted for inflation to 2017 values.

level.

To examine the source of the large rise in income, we split the tax unit's W-2 earnings into the earnings of the primary taxpayers associated with the tax debt (listed first on a joint return) and the taxpayer's spouses' earnings. The increase is attributable to taxpayers' spouses. Figure 7 shows that in the instrumented specification earnings do not rise for the individual whose case is designated CNC (Figure 7b), but there is a substantial increase in the earnings received by the spouses of married taxpayers (Figure 7d), which more than double 3 and 4 years after the case closed. There is a similar but much smaller in magnitude pattern in the standard difference-in-differences specification for spouses' earnings, although for individuals' own earnings the standard difference-in-differences shows a decline that begins well in advance of treatment.

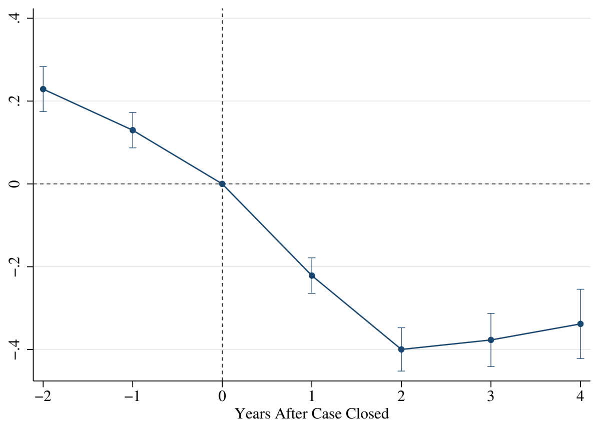
5.4 Summary

Suspending debt collection due to inability to pay basic living expenses leads to a large proportional increase in the taxpayer's spouse's (but not the taxpayer's own) W-2 earnings

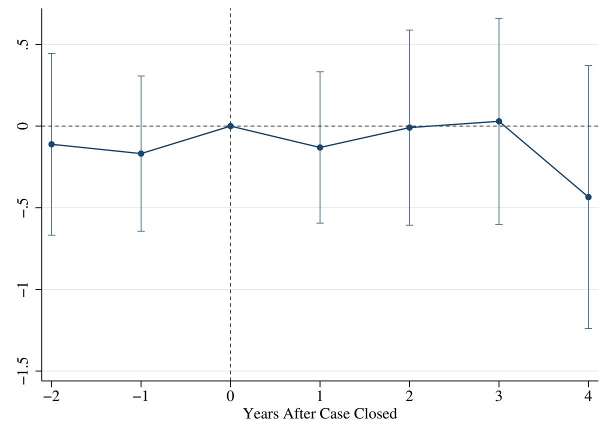
Figure 7: Effect of a CNC designation on W-2 Earnings (IHS): Taxpayer vs. Spouse

Taxpayer whose case was designated CNC

(a) Difference-in-differences

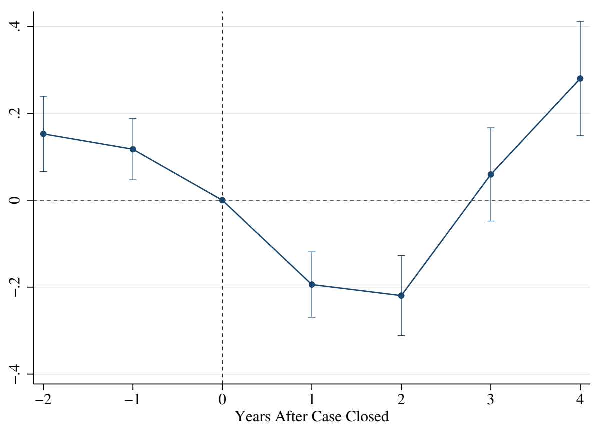


(b) IV with residualized instrument

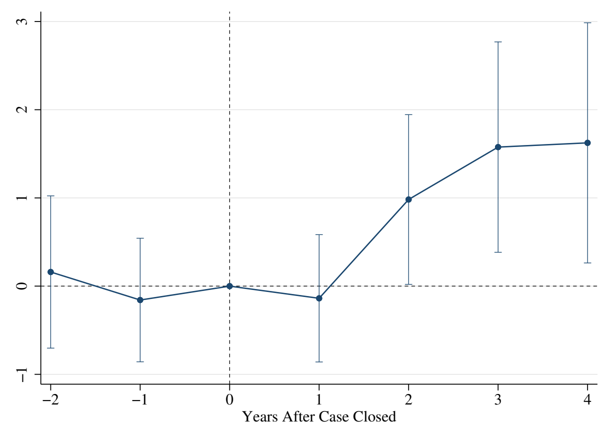


Spouse

(c) Difference-in-differences



(d) IV with residualized instrument



Notes: Number of cases for W-2 earnings is 139,246 for difference-in-differences and 121,923 for the IV specification, and the case numbers for the spouse's W-2 earnings are 85,492 and 70,171. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. Spouses include individuals married to taxpayers in the year in which their case was closed. W-2 earnings adjusted for inflation to 2017 values.

three and four years after the case closed. This increase in income is consistent with the incentives a CNC designation provides by reducing the effective marginal tax rate (at least over the range where additional income does not trigger the revocation of the CNC designation), and suggests that the increased keep rate on higher earnings has a greater effect on income than the wealth effect from setting debt aside. As the majority of the spouses of the married taxpayers in our sample are female, our results are consistent with the standard finding that married women's labor supply is more elastic than men's labor supply (Keane 2011). An alternative explanation for the increase in spouse's earnings is through a liquidity channel: by enabling investments that facilitate working (e.g. a second car or child care), the potential positive wealth shock from a CNC designation could increase labor supply in this population of taxpayers.

A CNC designation does not detectably change tax filing or payments toward outstanding debt. The lack of significant estimated effects on tax compliance could result from the imprecise nature of the IV estimates, from an offset between the increase in compliance stemming from the potential for debt suspension to be revoked and the decrease in compliance due to the perception that enforcement is lenient, or from an absence of such effects.

6 Conclusion

Random and quasi-random assignment of cases to administrative officers provides an opportunity to study the causal effects of policy interventions when administrative officers' discretion can determine who receives treatment. This paper uses such an approach to assess the effects of suspending attempts to collect a taxpayer's unpaid taxes. Variation comes from assignment of cases to Revenue Officers of differing "leniency," the inherent tendency to designate cases currently not collectible due to taxpayer hardship. To our knowledge, this is the first application of this research design to issues of tax enforcement. We find that, following the suspension of collection efforts, taxpayers have higher incomes, driven by increases in their spouses' earnings. Unlike naïve difference-in-difference results, in which suspending collection is strongly associated with declines in debt repayment, tax filing, and W-2 earnings, we find that these behaviors either increase or do not decline detectably.

In this setting, the ability of an examiner assignment design to elicit a causal effect of debt forgiveness not polluted by sample selection bias is tempered by the fact that the assignment of case managers to cases is, at best, only conditionally random and the average number of cases per Revenue Officer is not very large. These issues constrain the confidence one can have in the estimates reported here. Nevertheless, because quasi-random assignment of tax officers to cases is widespread, and the research availability of administrative data is growing rapidly, the potential for this research design to provide insight into the causal effects of tax enforcement actions is substantial. We look to future research applying it to other settings where the empirical caveats are less troubling, and hope that our initial foray into using the examiner assignment design to learn about the causal effects of tax enforcement will make the research path forward clearer.

References

Allingham, Michael G. and Agnar Sandmo, "Income Tax Evasion: A Theoretical Analysis," *Journal of Public Economics*, 1972, *1* (3), 323 – 338.

Angrist, Joshua D. and Alan B. Krueger, "Split-Sample Instrumental Variables Estimates of the Return to Schooling," *Journal of Business & Economic Statistics*, 1995, *13* (2), 225–235.

_ , Guido B. Imbens, and Alan B. Krueger, "Jackknife Instrumental Variables Estimation," *Journal of Applied Econometrics*, 1999, *14* (1), 57–67.

Bellemare, Marc F. and Casey J. Wichman, "Elasticities and the Inverse Hyperbolic Sine Transformation," *Oxford Bulletin of Economics and Statistics*, 2020, *82* (1), 50–61.

Cragg, John G and Stephen G Donald, "Testing identifiability and specification in instrumental variable models," *Econometric Theory*, 1993, pp. 222–240.

Cranor, Taylor, Jacob Goldin, Tatiana Homonoff, and Lindsay Moore, "Communicating tax penalties to delinquent taxpayers: Evidence from a field experiment," *National Tax Journal*, 2020, *73* (2), 331–285.

Dobbie, Will and Jae Song, "Debt Relief and Debtor Outcomes: Measuring the Effects of Consumer Bankruptcy Protection," *American Economic Review*, March 2015, *105* (3), 1272–1311.

_ , Jacob Goldin, and Crystal S. Yang, "The Effects of Pretrial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges," *American Economic Review*, February 2018, *108* (2), 201–40.

_ , Paul Goldsmith-Pinkham, and Crystal S. Yang, "Consumer Bankruptcy and Financial Health," *The Review of Economics and Statistics*, 2017, *99* (5), 853–869.

Doyle, Jr., Joseph J., "Child Protection and Child Outcomes: Measuring the Effects of Foster Care," *American Economic Review*, 2007, *97*, 1583–610.

_ , "Child Protection and Adult Crime: Using Investigator Assignment to Estimate Causal Effects of Foster Care," *Journal of Political Economy*, 2008, *116* (4), 746–770.

Duflo, Esther, "Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment," *American Economic Review*, 2001, *91* (4), 795–813.

Gross, Max and E Jason Baron, "Temporary Stays and Persistent Gains: The Causal Effects of Foster Care," *American Economic Journal: Applied Economics*, 2021.

Internal Revenue Service, "2017 Data Book," 2017. https://www.irs.gov/pub/irs-soi/17databk.pdf [accessed: June 4, 2018].

Keane, Michael P, "Labor supply and taxes: A survey," *Journal of Economic Literature*, 2011, *49* (4), 961–1075.

Kleibergen, Frank and Richard Paap, "Generalized reduced rank tests using the singular value decomposition," *Journal of econometrics*, 2006, *133* (1), 97–126.

Maestas, Nicole, Kathleen J. Mullen, and Alexander Strand, "Does Disability Insurance Receipt Discourage Work? Using Examiner Assignment to Estimate Causal Effects of SSDI Receipt," *American Economic Review*, August 2013, *103* (5), 1797–1829.

Miller, Erik, Stacy Orlett, and Alex Turk, "Uncollectible vs. Unproductive: Compliance Impact of Working Collection Cases That are Ultimately Not Fully Collectible," in "IRS-TPC Research Conference: Advancing Tax Administration" 2014.

Miller, Joshua J and Silda Nikaj, "The response of delinquent taxpayers to more aggressive collection," *National Tax Journal*, 2016, *69* (1), 77–102.

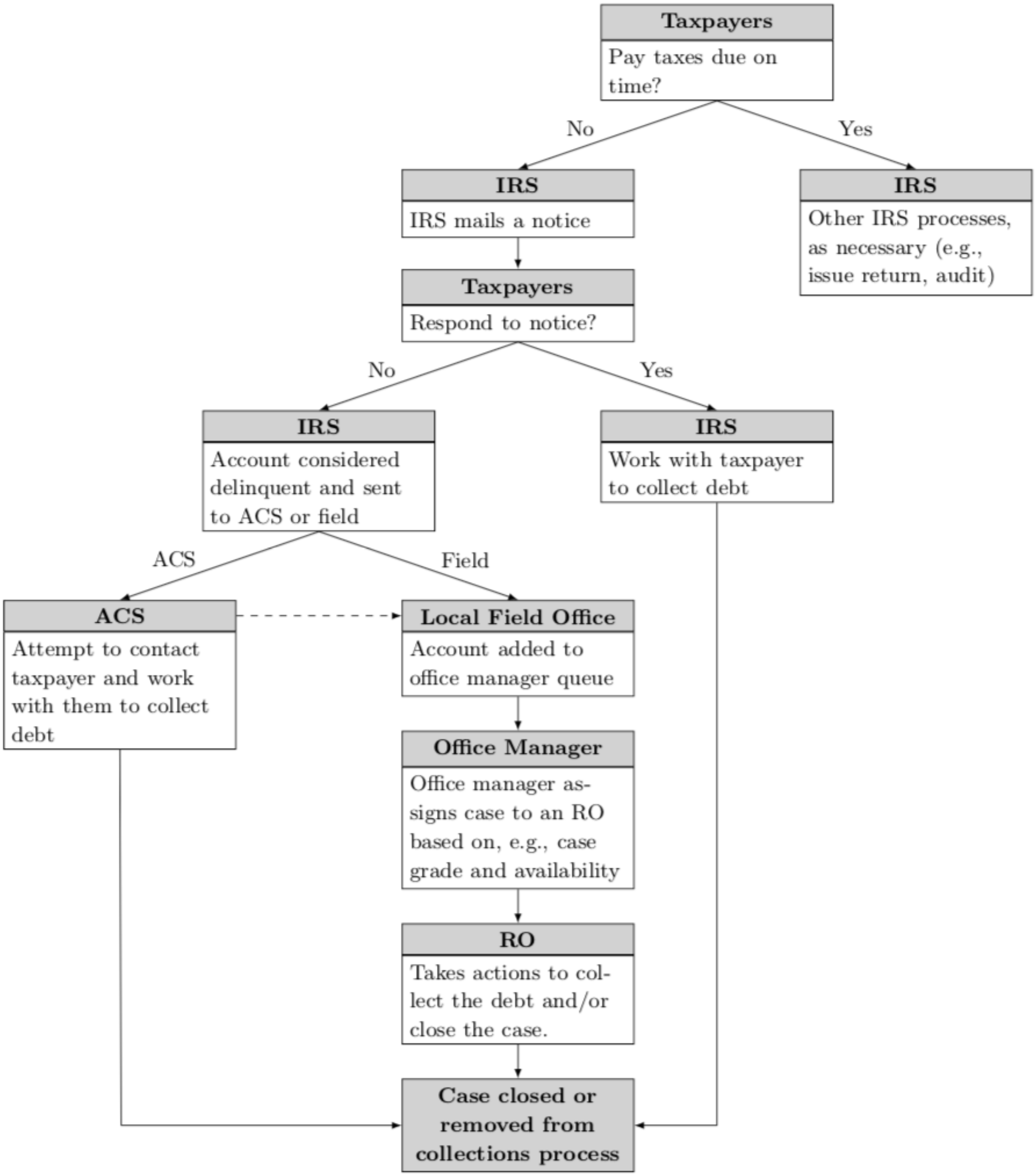
Mueller-Smith, Michael, "The Criminal and Labor Market Impacts of Incarceration," Working Paper 2015.

Norris, Samuel, Matthew Pecenco, and Jeffrey Weaver, "The Effects of Parental and Sibling Incarceration: Evidence from Ohio," *American Economic Review*, September 2021, *111* (9), 2926–63.

Perez-Truglia, Ricardo and Ugo Troiano, "Shaming Tax Delinquents," *Journal of Public Economics*, 2018, *167*, 120–137.

Slemrod, Joel, "Tax Compliance and Enforcement," *Journal of Economic Literature*, December 2019, *57* (4), 904–54.

Appendix A Overview of the collections process



Appendix B Data dictionary and sample statistics

B.1 Acronyms

ALE Allowable living expenses

ACS Automated Collection System

AGI Adjusted gross income

CNC Currently not collectible

EAP Estimated ability to pay

GS General schedule

IRM Internal Revenue Manual

IRS Internal Revenue Service

TPI Total positive income

B.2 Variable definitions

B.2.1 Variables included in the residualization process

Allowable living expenses The allowable living expenses data is provided quarterly at the county level by the IRS. We used the "housing and utilities" standard. We used the value for a "Family of 1" unless we had evidence that the taxpayer was married, in which case we used the value for a "Family of 2." When a zip code covered multiple counties, we took the average of the allowable living expenses given for those counties. We use the values from quarter 4 of Year 1. The zip-to-county conversion was done using data provided by the U.S. Department of Housing and Urban Development. We matched this to the administrative tax data by zip code.

Case grade Provided in the field data. Cases have a grade of 9, 11, 12, or 13. The assigned grade reflects the expected difficulty of closing the case.

Estimated ability to pay We develop an "estimated ability to pay" metric which is equal to a taxpayers average AGI for the three years before their case was assigned to a Revenue Officer, multiplied by the lesser of 10 and the number of years before the taxpayer turns 65, divided by the outstanding balance on the account when the taxpayer is assigned to a Revenue Officer. When AGI is missing for some year, we impute the filing threshold. Larger values suggest that the taxpayer would have greater income and therefore be more able to resolve their outstanding debt.

Group Provided in the Revenue Officer data. This is the group to which the Revenue Officer belongs. A Revenue Officer is assigned cases by the group manager.

High priority indicator Cases are assigned priority codes. The relationship between priority code and priority level is described in IRM 1.4.50.8.4 1. High priority cases include priority codes 99-108, with priority 99 and 100 cases being the highest priority cases. Medium priority cases include priority codes 201-208. Low priority cases include priority codes 301-303. This indicator is equal to 1 for cases with priority code 99 or 100.

Oldest debt more than 12 months old indicator This indicator is equal to 1 if the oldest debt on the case when it is assigned to field is older than 12 months old.

Oldest debt more than 36 months old indicator This indicator is equal to 1 if the oldest debt on the case when it is assigned to field is older than 36 months old.

Previously assigned to field indicator This indicator is equal to 1 if the taxpayer had modules assigned to the field before the case considered in the project. We consider debt starting in 2009 when constructing this variable.

Revenue Officer GS grade Provided in the Revenue Officer data. Revenue Officers may have a GS grade of 4, 5, 7, 9, 11, 12, or 13. A grade of 4 indicates a group manager. Grades 5 and 7 are training grades. As of 2017, IRS employees with a GS grade of 9 could no longer serve as Revenue Officers. Because our analysis focuses on cases closed before 2017, we observe Revenue Officers with GS grades 9, 11, 12, and 13.

Urban indicator The urban dummy is based on data provided by the U.S. Department of Agriculture Economic Research Service (USDA ERS). The data include the Rural Urban Continuum Code (RUCC) by county. We designate a location as not urban if the RUCC for the zip code is 7 or 9, which includes areas with populations less than 20,000 that are non-adjacent to metro areas. We match the county to zip code using the 2014 Q4 zip-to-county data provided by the U.S. Department of Housing and Urban Development. For a few zip codes not included in the data from USDA ERS we use population-by-zip code data from the 2010 Census. We categorize these zip codes as urban if the population is greater than 20,000. We matched this to the administrative tax data by zip code.

Year of birth The year of birth of the taxpayer. We censor at both ends: we set the year of birth equal to 1930 for taxpayers with year of birth older than 1930, and we set the year of birth equal to 1997 for taxpayers with year of birth earlier than 1997.

B.2.2 Outcome variables

Filing tax returns This indicator is equal to 1 if the taxpayer filed a return conditional on whether or not the IRS received a W-2 for the taxpayer.

Payments toward outstanding tax debt : Calculated as the sum of all payments remitted to the IRS against debt from previous tax years during the twelve months after the taxpayer's case was closed.

W-2 earnings We use the sum of earnings reported on all unique W-2s received for the taxpayer. We present this variable at the household, individual, and spouse level.

B.2.3 Balance variables

Model score An estimate generated by the IRS of the probability of repayment, with lower scores indicating a lower estimated probability of collection. This variable has 25$^{\text{th}}$, 50$^{\text{th}}$, and 75$^{\text{th}}$ percentile values of approximately 0.05, 0.09, and 0.2, respectively.

Average W-2 earnings The average of earnings reported through W-2s for the three years prior to assignment to a Revenue Officer.

Average AGI The average of Adjusted Gross Income (as defined above) for the three years prior to assignment to a Revenue Officer.

AGI before assignment The value of Adjusted Gross Income (as defined above) for the year prior to assignment to a Revenue Officer.

Filed before assignment This indicator is equal to 1 if the taxpayer filed a tax return in the year prior to assignment to a Revenue Officer.

B.3 Sample statistics

Table 3 shows the mean and standard deviation of the variables used in our analysis in Year 0 by CNC designation.

Table 3: Sample statistics (Year 0)

	Non-CNC	CNC	Total
Observations	86,083	37,313	123,396
Residualization variables			
High priority indicator	0.216	0.219	0.217
	(0.411)	(0.414)	(0.412)
Case grade = 9	0.142	0.142	0.142
	(0.349)	(0.349)	(0.349)
Case grade = 11	0.480	0.538	0.497
	(0.500)	(0.499)	(0.500)
Case grade = 12	0.334	0.306	0.325
	(0.471)	(0.461)	(0.468)
Case grade = 13	0.045	0.014	0.036
	(0.207)	(0.118)	(0.185)
Year of birth	1963	1961	1962
	(11)	(12)	(11)
Estimated ability to pay	54.566	15.733	42.824
	(148.502)	(65.649)	(130.406)
Previously assigned	0.413	0.425	0.417
	(0.818)	(0.806)	(0.814)
Oldest debt > 12 mo.	0.713	0.755	0.726
	(0.452)	(0.430)	(0.446)
Oldest debt > 36 mo.	0.598	0.701	0.629
	(0.490)	(0.458)	(0.483)
Allowable living expenses	1,860	1,731	1,820
	(585)	(524)	(570)
Urban indicator	0.978	0.973	0.976
	(0.147)	(0.163)	(0.152)
RO Inventory	54.330	54.881	54.497
	(16.518)	(15.883)	(16.331)
RO GS grade = 9	0.150	0.137	0.146
	(0.357)	(0.344)	(0.353)
RO GS grade = 11	0.429	0.482	0.445
	(0.495)	(0.500)	(0.497)
RO GS grade = 12	0.390	0.368	0.383
	(0.488)	(0.482)	(0.486)
RO GS grade = 13	0.031	0.014	0.026
	(0.174)	(0.116)	(0.159)

Table 3: Sample statistics (Year 0, continued)

	Non-CNC	CNC	Total
Observations	86,083	37,313	123,396
Outcome variables			
Payments toward outstanding debt	6,774	2,889	5,599
	(48,149)	(38,487)	(45,479)
Filed | W-2s	0.793	0.668	0.755
	(0.405)	(0.471)	(0.430)
W-2 wages	63,925	36,810	57,061
	(90,438)	(55,207)	(83,780)
W-2 wages, primary	81,609	48,491	73,269
	(99,555)	(62,130)	(92,704)
W-2 wages, spouse	11,801	6,047	10,165
	(30,817)	(19,626)	(28,212)
Balance variables			
Model score	0.170	0.105	0.150
	(0.162)	(0.098)	(0.148)
Pre-Ave. W-2 wages, HH	62,430	22,431	50,335
	(485,752)	(51,761)	(407,128)
Pre-Ave. AGI, HH	170,507	54,989	135,577
	(2,699,416)	(1,365,700)	(2,377,013)
AGI year before assignment, HH	209,803	50,231	161,551
	(5,323,174)	(330,781)	(4,450,411)
Filed before assignment	0.831	0.671	0.783
	(0.375)	(0.470)	(0.412)
Instruments			
Simple LOO instrument	0.290	0.331	0.302
	(0.108)	(0.115)	(0.112)
Residualized instrument	-0.010	0.022	0.000
	(0.079)	(0.084)	(0.082)

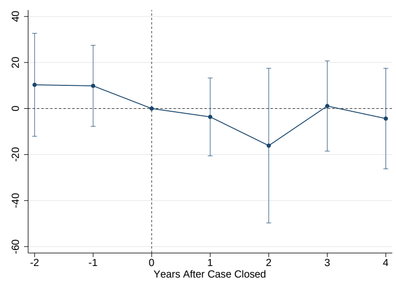
Notes: Includes all cases that meet our sample restriction criteria between November 2014 and December 2018. Limited to cases worked by Revenue Officers who closed at least 20 cases that met our sample restriction criteria between November 2014 and December 2018. All monetary values adjusted for inflation to 2017 values.

Appendix C Robustness to choice of case count cutoff

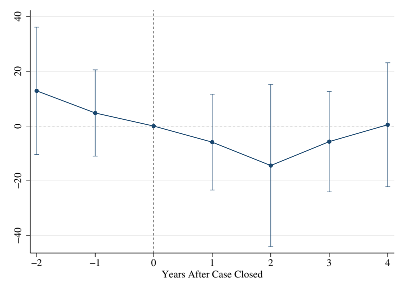
Figures 8 through 11 show our residualized IV results using the alternative Revenue Officer case count cut-offs of 10, 20, and 30.

Figure 8: Effect of a CNC designation on payments toward outstanding tax debt ($1,000), IV with residualized instrument

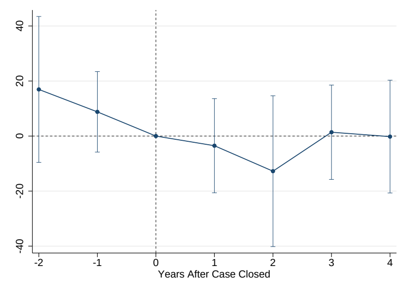
(a) 10+ Cases



(b) 20+ Cases



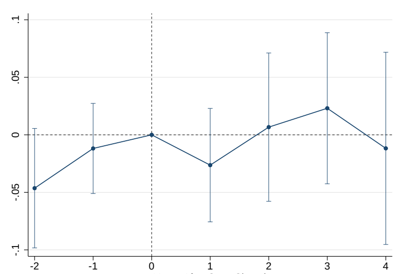
(c) 30+ Cases



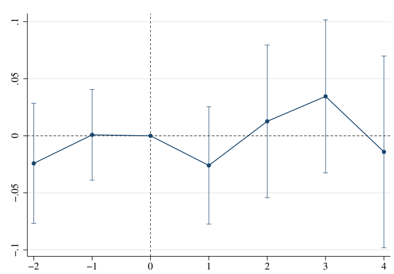
Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Coefficients are shown in thousands. Payment values adjusted for inflation to 2017 values.

Figure 9: Effect of a CNC designation on filing a tax return, IV with residualized instrument

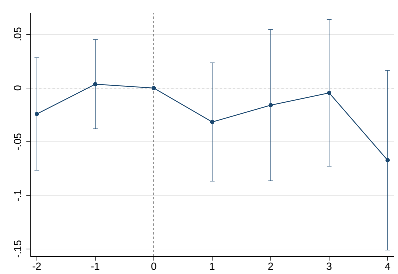
(a) 10+ Cases



(b) 20+ Cases



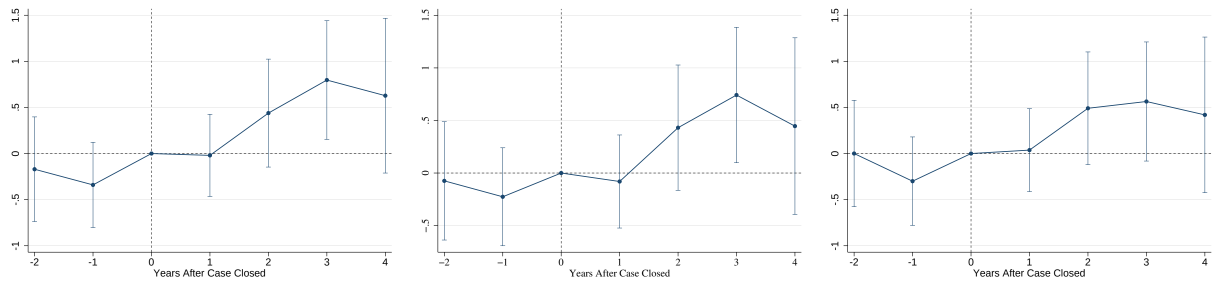
(c) 30+ Cases



Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018.

Figure 10: Effect of a CNC designation on W-2 wages (IHS), IV with residualized instrument

(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases

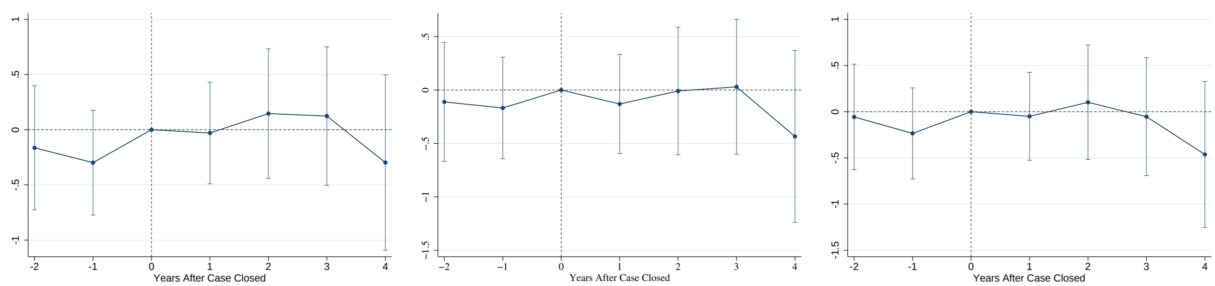


Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. W-2 earnings adjusted for inflation to 2017 values.

Figure 11: Effect of a CNC designation on W-2 Wages (IHS): Taxpayer vs. Spouse

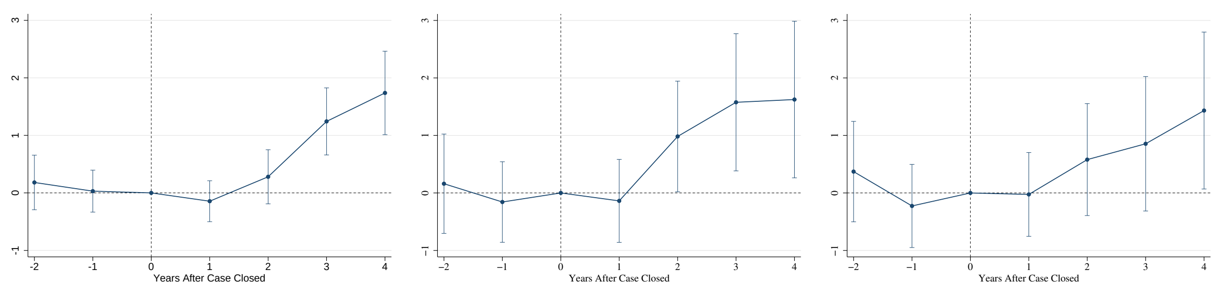
Taxpayer whose case was designated CNC

(a) 10+ Cases (b) 20+ Cases (c) 30+ Cases



Spouse

(d) 10+ Cases (e) 20+ Cases (f) 30+ Cases



Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. Spouses include individuals married to taxpayers in the year in which their case was closed. W-2 earnings adjusted for inflation to 2017 values.

Appendix D Discussion of standard IV assumptions

The primary assumptions in any instrumental variables specification are relevance, monotonicity, and the exclusion restriction.

Relevance means that Revenue Officer assignment must be correlated with CNC designation (formally, $\mathbb{C}\text{ov}[CNC_{ij}, Z_{ij}] \neq 0$). The correlation between CNC status and the residualized leave-one-out instrument is 0.102.

In our setting, monotonicity means that cases designated CNC by low-propensity (high-σ_j) Revenue Officers would always have also been designated as CNC by high-propensity (low-σ_j) Revenue Officers. Similarly, cases that were not deemed CNC by high-propensity (low-σ_j) Revenue Officers would also not have been deemed CNC by high-propensity (low-σ_j) Revenue Officers. This assumption would be violated if, e.g., some Revenue Officers were more likely to designate cases as CNC if the taxpayers were older, while some Revenue Officers were more inclined to designate cases as CNC if the taxpayer were younger. Monotonicity is also a concern in situations where the examiners have a multidimensional decision to make. In our setting, cases may be closed through full pay, installment agreements, and other decisions in addition to a CNC designation. For a discussion of monotonicity when there are multi-dimensional choices in these settings, see Norris et al. (2021) and Gross and Baron (2021).

Finally, the exclusion restriction means Revenue Officer propensity to designate a case CNC must only affect taxpayer outcomes through the variation in having a case designated CNC, that is, $\mathbb{C}\text{ov}[Z_i, h_i] = 0$. This assumption would be violated if Revenue Officer propensity to designate a case CNC is correlated with unobservable determinants of future taxpayer outcomes. This assumption is also violated if Revenue Officer propensity to designate a case CNC impacts future taxpayer outcomes through means other than CNC designation (e.g., if Revenue Officers that are more willing to deem a case CNC are also more likely to provide information to help a taxpayer avoid being in this situation in the future, or make more intense attempts to collect unpaid tax before designating a case CNC).