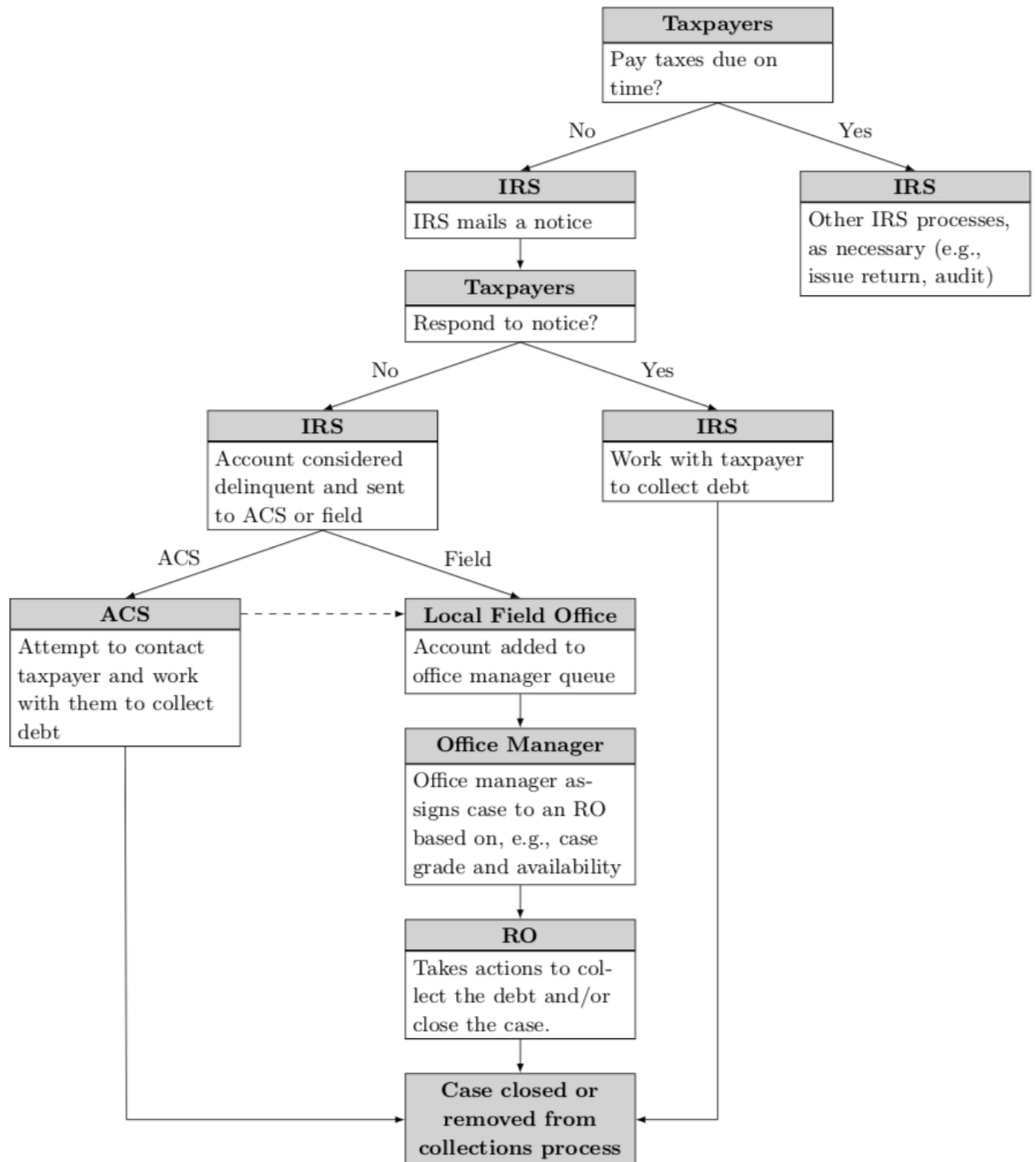


Online Appendices for “Does Giving Tax Debtors a Break Improve Compliance and Income? Evidence from Quasi-Random Assignment of IRS Revenue Officers”

Appendix A Overview of the collections process



Appendix B Data dictionary and sample statistics

B.1 Variable definitions

B.1.1 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.1.2 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.1.3 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 25th, 50th, and 75th 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.2 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
<i>Residualization variables</i>			
High priority indicator	0.216 (0.411)	0.219 (0.414)	0.217 (0.412)
Case grade = 9	0.142 (0.349)	0.142 (0.349)	0.142 (0.349)
Case grade = 11	0.480 (0.500)	0.538 (0.499)	0.497 (0.500)
Case grade = 12	0.334 (0.471)	0.306 (0.461)	0.325 (0.468)
Case grade = 13	0.045 (0.207)	0.014 (0.118)	0.036 (0.185)
Year of birth	1963 (11)	1961 (12)	1962 (11)
Estimated ability to pay	54.566 (148.502)	15.733 (65.649)	42.824 (130.406)
Previously assigned	0.413 (0.818)	0.425 (0.806)	0.417 (0.814)
Oldest debt > 12 mo.	0.713 (0.452)	0.755 (0.430)	0.726 (0.446)
Oldest debt > 36 mo.	0.598 (0.490)	0.701 (0.458)	0.629 (0.483)
Allowable living expenses	1,860 (585)	1,731 (524)	1,820 (570)
Urban indicator	0.978 (0.147)	0.973 (0.163)	0.976 (0.152)
RO Inventory	54.330 (16.518)	54.881 (15.883)	54.497 (16.331)
RO GS grade = 9	0.150 (0.357)	0.137 (0.344)	0.146 (0.353)
RO GS grade = 11	0.429 (0.495)	0.482 (0.500)	0.445 (0.497)
RO GS grade = 12	0.390 (0.488)	0.368 (0.482)	0.383 (0.486)
RO GS grade = 13	0.031 (0.174)	0.014 (0.116)	0.026 (0.159)

=====

Table 3: Sample statistics (Year 0, continued)

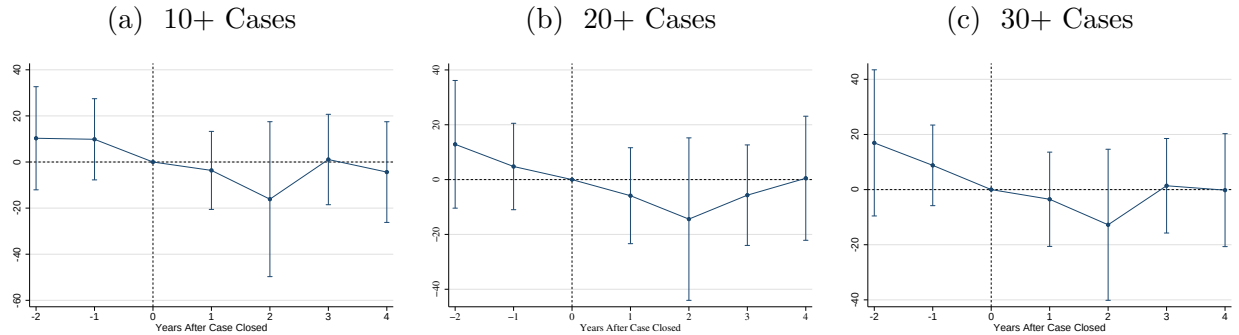
	Non-CNC	CNC	Total
Observations	86,083	37,313	123,396
<i>Debt variables</i>			
Balance at assignment, HH	129,529 (283,171)	160,218 (300,285)	138,809 (288,796)
Balance > Pre-Ave. AGI HH indicator	0.337 (0.473)	0.545 (0.498)	0.400 (0.490)
<i>Outcome variables</i>			
Payments toward outstanding debt	6,774 (48,149)	2,889 (38,487)	5,599 (45,479)
Filed W-2s	0.793 (0.405)	0.668 (0.471)	0.755 (0.430)
W-2 wages	63,925 (90,438)	36,810 (55,207)	57,061 (83,780)
W-2 wages, primary	81,609 (99,555)	48,491 (62,130)	73,269 (92,704)
W-2 wages, spouse	11,801 (30,817)	6,047 (19,626)	10,165 (28,212)
<i>Balance variables</i>			
Model score	0.170 (0.162)	0.105 (0.098)	0.150 (0.148)
Pre-Ave. W-2 wages, HH	62,430 (485,752)	22,431 (51,761)	50,335 (407,128)
Pre-Ave. AGI, HH	170,507 (2,699,416)	54,989 (1,365,700)	135,577 (2,377,013)
AGI year before assignment, HH	209,803 (5,323,174)	50,231 (330,781)	161,551 (4,450,411)
Filed before assignment	0.831 (0.375)	0.671 (0.470)	0.783 (0.412)
<i>Instruments</i>			
Simple LOO instrument	0.290 (0.108)	0.331 (0.115)	0.302 (0.112)
Residualized instrument	-0.010 (0.079)	0.022 (0.084)	0.000 (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

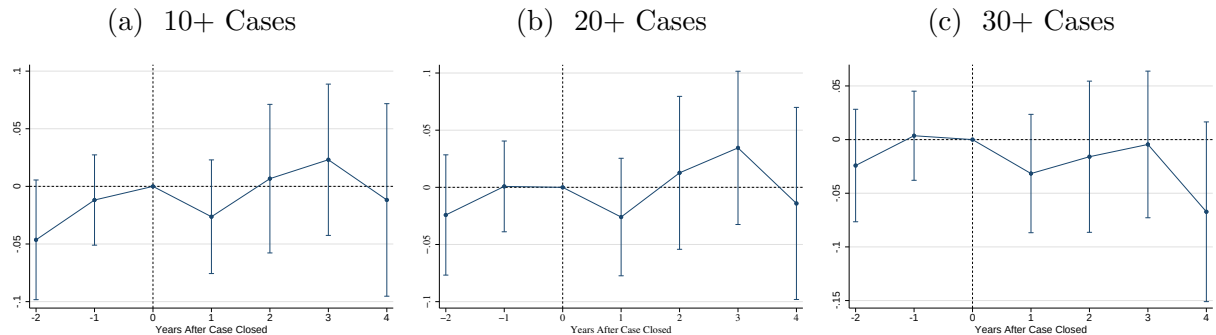
Figures 8 through 11 show our residualized IV results using the alternative Revenue Officer case count cut-offs of 10, 20, and 30.

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



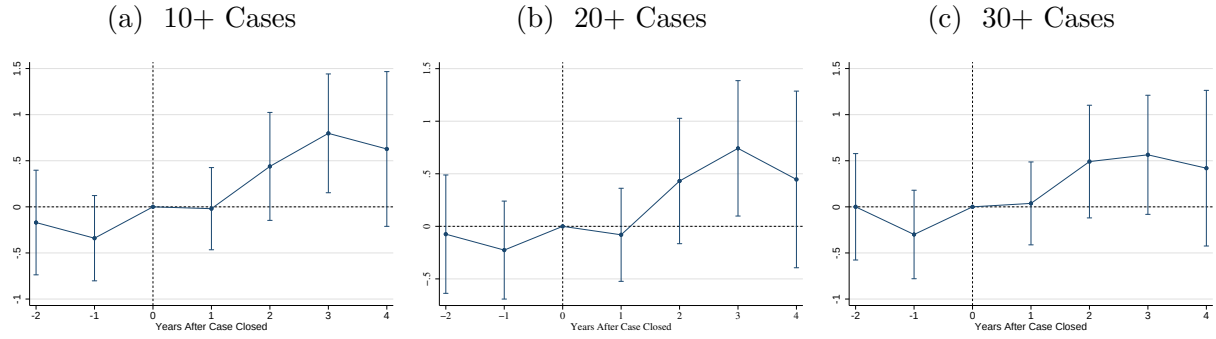
Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Coefficients are shown in thousands. Payment values adjusted for inflation to 2017 values.

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



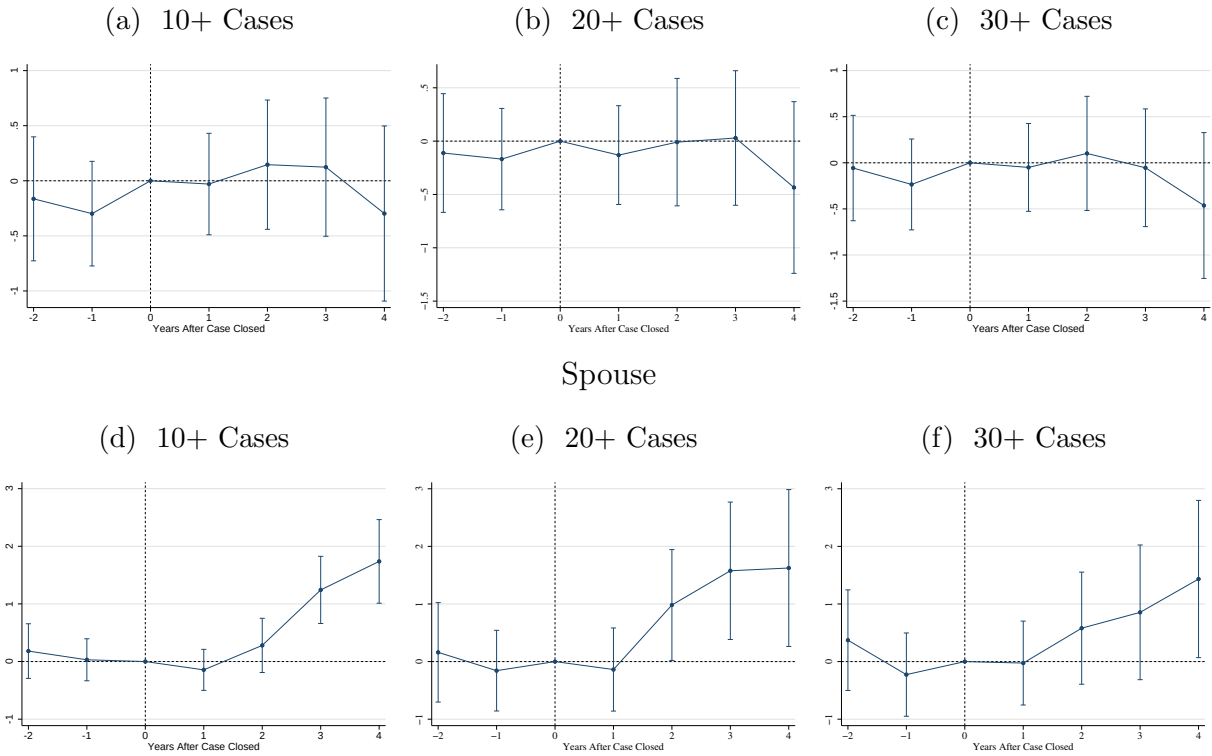
Notes: Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018.

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



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



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, $\text{Cov}[CNC_{ij}, Z_{ij}] \neq 0$). 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, $\text{Cov}[Z_i, h_i] = 0$. 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).

Appendix E Further Investigation of Potentially Systematically Non-random Case Assignment

The correlations between Revenue Officer leniency and pre-assignment case characteristics imply that case assignment is not random. In conversations with the group managers who assign cases, the only pattern of non-random case assignment that they mentioned was that Revenue Officers would on rare occasion be assigned a case in a higher grade for training purposes. Such training is accounted for by including Revenue Officer and case grades in the construction of our instrument. However, other patterns the group managers did not raise could drive case assignment.

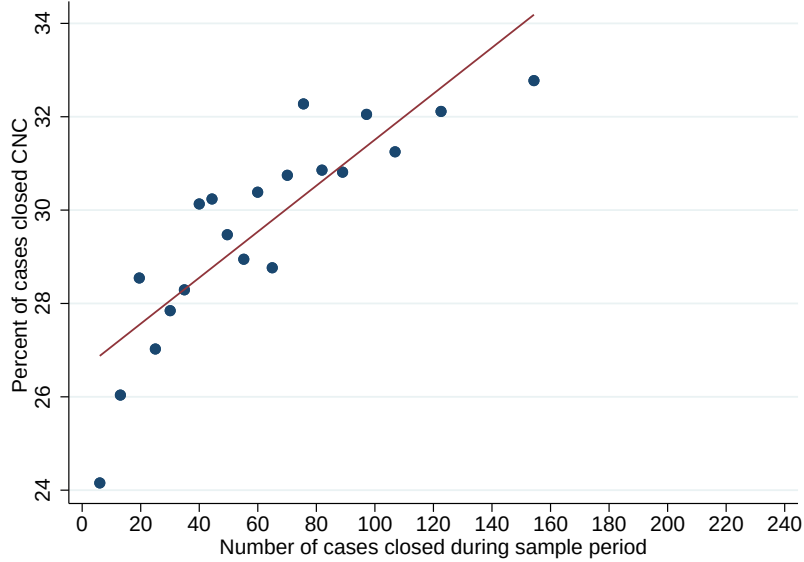
The most directly problematic case assignment pattern would be if managers were to assign cases where they expected the taxpayer's future income (and thus the potential revenue collected) to be high or low to Revenue Officers the manager expected to be particularly harsh or lenient in designating cases CNC. Note that the manager's expectation would have to be over and above the case characteristics for which we control in constructing the instrument, including an IRS model's prediction of the potential case revenue. For such an assignment pattern to be revenue maximizing, managers would assign cases with higher perceived potential future income to harsher Revenue Officers (and vice versa) to ensure that taxpayers who might be able to pay are not assigned to a Revenue Officer who is more likely to assign the case CNC. Such a pattern would bias our estimates of the effect of pausing collection on future income downward.

While we cannot directly test whether this better-revenue-potential-to-harsher-officer assignment pattern takes place, we find correlations between the likelihood a case is designated CNC and other Revenue Officer characteristics that are inconsistent with this taking place. These three correlations are as follows:

(1) As Figure 12 shows, Revenue Officers who close more cases total during the sample period also designate a higher fraction of their cases CNC. This could be consistent with the pattern of concern if CNC cases are worked less intensively than other cases because

they are assigned to less thorough Revenue Officers. However, other correlations in the data contradict that possibility.

Figure 12: Revenue Officers Who Close More Cases Designate More Cases as CNC

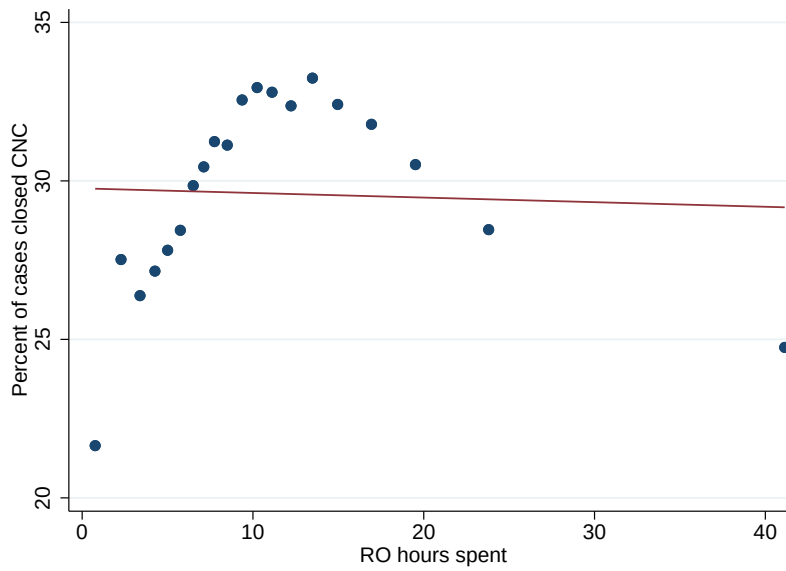


Notes: Each dot is an average of one twentieth of the Revenue Officers in our main analysis sample. The line is a linear best fit.

(2) The cases closed with the fewest hours spent by the Revenue Officer are less likely than average to be designated CNC. This conflicts with the pattern of concern because it is inconsistent with low-revenue-potential cases being designated CNC quickly. The cases most likely to be designated CNC are those in the middle of the distribution of hours spent on the case, as shown in Figure 13.

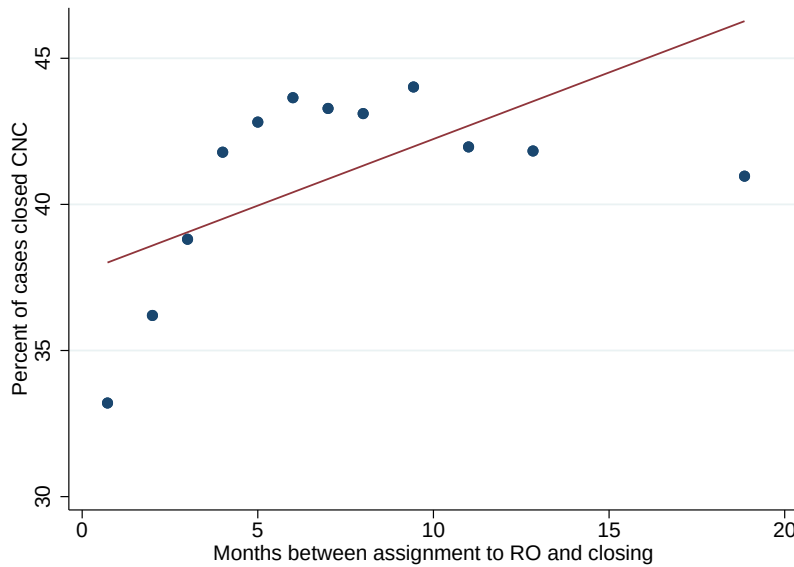
(3) Likewise, Figure 14 shows that cases designated CNC spend neither an especially long nor especially short time on Revenue Officers' desks, on average.

Figure 13: Cases with Middling Hours Spent are Most Likely to be Designated CNC



Notes: Each dot is an average of one twentieth of the cases in our main analysis sample. The line is a linear best fit.

Figure 14: Cases That Spend a Middling Amount of Time on Officers' Desks are Most Likely to be Designated CNC

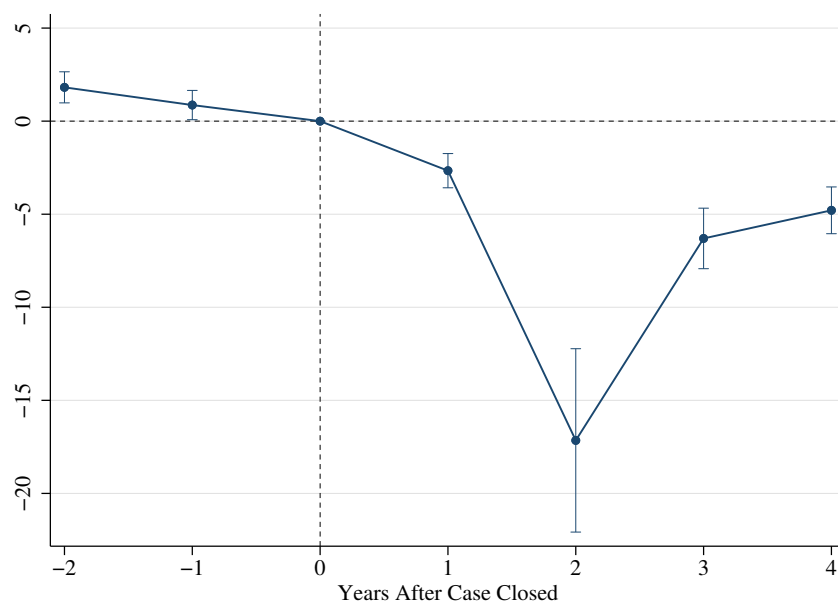


Notes: Each dot is an average of one twentieth of the cases in our main analysis sample. The line is a linear best fit.

Appendix F Results Using Difference-in-Differences

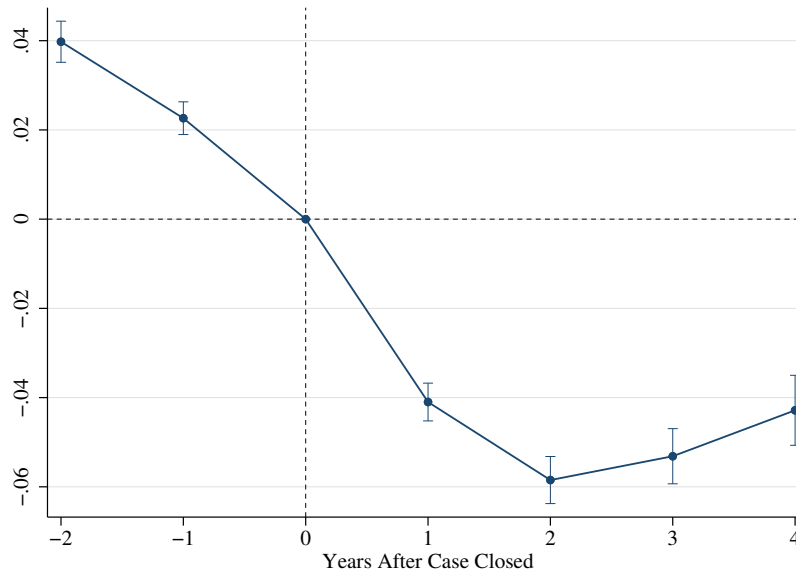
Figures 15 through 18 show the results if we run a standard difference-in-differences event study (i.e., if we do not use our residualized instrument).

Figure 15: Difference in differences:
Effect of a CNC designation on payments toward outstanding tax debt (\$1,000)



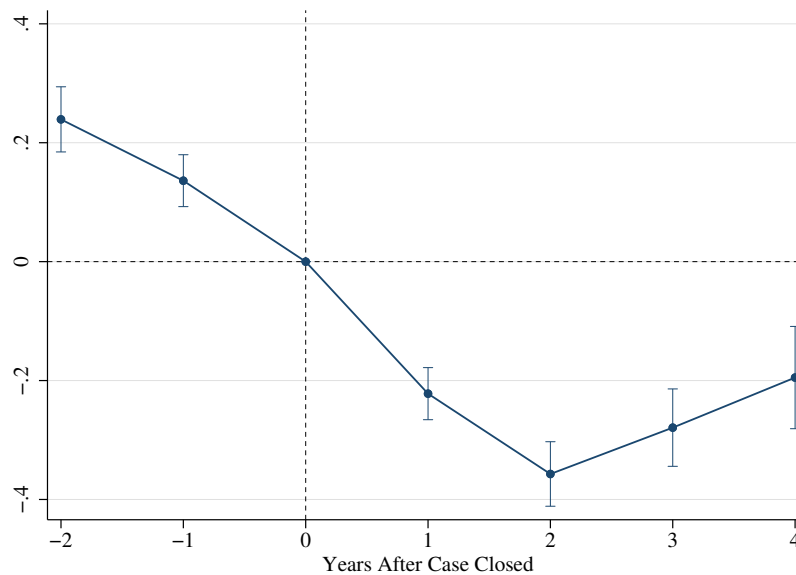
Notes: Number of cases is 139,246 for the difference-in-differences specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Coefficients are shown in thousands. Payment values adjusted for inflation to 2017 values.

Figure 16: Difference in differences:
Effect of a CNC designation on filing a tax return



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

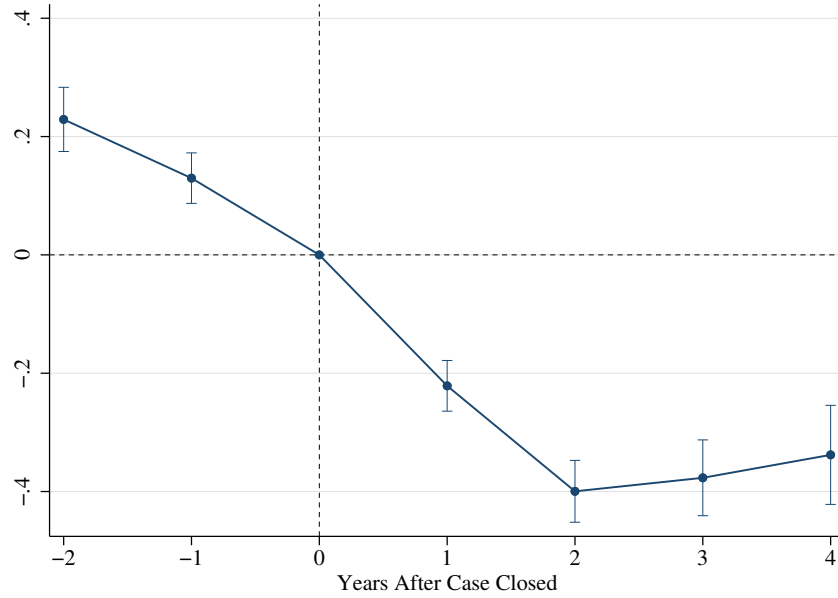
Figure 17: Difference in differences:
Effect of a CNC designation on W-2 earnings (IHS)



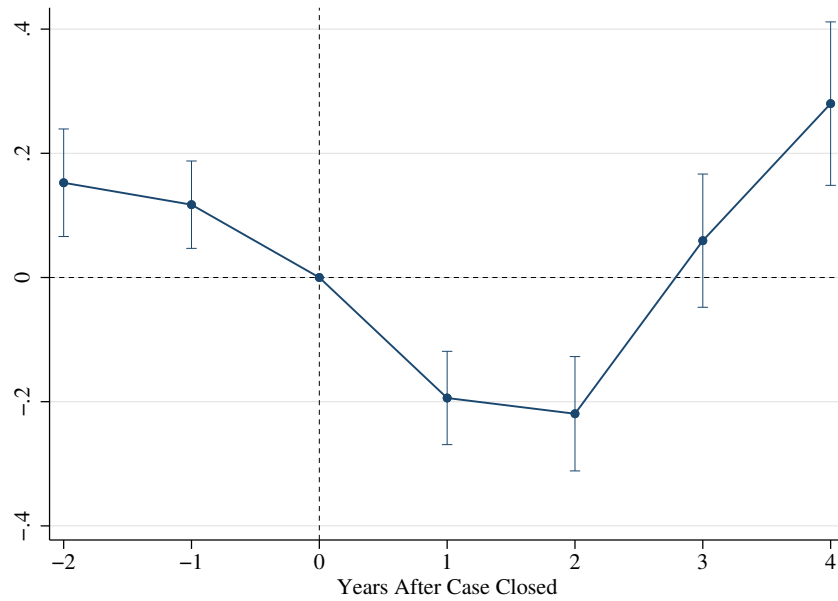
Notes: Number of cases is 139,246 for difference-in-differences specification. Includes cases worked by all Revenue Officers that meet our sample restriction criteria between November 2014 and December 2018. Values given are the inverse hyperbolic sine transformation of the outcome values. W-2 earnings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

Figure 18: Difference in differences:
Effect of a CNC designation on W-2 Earnings (IHS): Taxpayer vs. Spouse

(a) Taxpayer whose case was designated CNC



(b) Spouse



Notes: Number of cases for W-2 earnings for the difference-in-differences specification is 139,246 for panel (a) and 85,492 for panel (b). 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.

Additional References

Gross, M. and Baron, E.J. (2022) Temporary stays and persistent gains: the causal effects of foster care. *American Economic Journal: Applied Economics*, 14(2), 170-199. Available from: <https://doi.org/10.1257/app.20200204>

Norris, S., Pecenco, M., and Weaver, J. (2021) The effects of parental and sibling incarceration: evidence from Ohio,” *The American Economic Review*, 111(9), 2926-2963. Available from: <https://doi.org/10.1257/aer.20190415>