

NBER WORKING PAPER SERIES

NEW GIG WORK OR CHANGES IN REPORTING?
UNDERSTANDING SELF-EMPLOYMENT TRENDS IN TAX DATA

Andrew Garin
Emilie Jackson
Dmitri K. Koustas

Working Paper 32368
<http://www.nber.org/papers/w32368>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2024

We are grateful to Katharine Abraham, Dan Black, Naomi Feldman, John Guyton, Lawrence Katz, Elira Kuka, Kara Leibel, Alicia Miller, Bruce Meyer, Jacob Mortenson, Max Risch, Michael Weber, Andy Whitten, Eleanor Wilking, and seminar participants at the Office of Tax Analysis at the US Treasury, the National Tax Association meetings, MIT Sloan, and the Census Bureau for helpful discussions. We thank Emmanuel Saez for providing us with code used to implement the calculations in Chetty, Friedman, Saez (2013) in the restricted-access administrative data. We are grateful to the Russell Sage Foundation and the Sloan Foundation for generous support, and to Reigne Dadey for excellent research assistance. Garin gratefully acknowledges support from the OECD Future of Work Fellowship and from the Sloan Foundation for a post-doctoral fellowship at NBER during which part of this research was conducted. Jackson gratefully acknowledges support from the Peter G. Peterson Foundation for a post-doctoral fellowship at NBER during which part of this research was conducted. 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, the Internal Revenue Service, or the National Bureau of Economic Research. All results have been reviewed to ensure that no confidential information is disclosed.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2024 by Andrew Garin, Emilie Jackson, and Dmitri K. Koustas. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

New Gig Work or Changes in Reporting? Understanding Self-Employment Trends in Tax Data

Andrew Garin, Emilie Jackson, and Dmitri K. Koustas

NBER Working Paper No. 32368

April 2024

JEL No. H26,J21,J41,J46,J82

ABSTRACT

Rising self-employment rates in U.S. tax data that are absent in survey data have led to speculation that tax records capture a rise in new “gig” work that surveys miss. Drawing on the universe of IRS tax returns, we show that trends in firm-reported payments to “gig” and other contract workers do not explain the rise in self-employment reported to the IRS; rather, that increase is driven by self-reported earnings of individuals in the EITC phase-in range. We isolate pure reporting responses from real labor supply responses by examining births of workers’ first children around an end-of-year cutoff for credit eligibility that creates exogenous variation in tax rates at the end of the tax year after labor supply decisions are already sunk. We find that exposing workers with sunk labor supply to negative marginal tax rates results in large increases in their propensity to self-report self-employment—only a small minority of which leads to bunching at kink-points. Consistent with pure strategic reporting behavior, we find no impact on reporting among taxpayers with no incentive to report additional income and no effects on firm-reported payments of any kind. Moreover, we find these reporting responses have grown over time as knowledge of tax incentives has become widespread. Quantification exercises suggest that changes in taxpayer reporting behavior are a major driver of discrepancies between self-employment trends in self-reported and third-party reported data. Our findings suggest caution is warranted before deferring to self-reported tax data over other data sources when measuring labor market trends.

Andrew Garin
Heinz College
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213
and NBER
agarin@andrew.cmu.edu

Dmitri K. Koustas
Harris Public Policy,
University of Chicago
1307 E. 60th St. Room 3055
Chicago, IL 60605
dkoustas@uchicago.edu

Emilie Jackson
Michigan State University
486 W. Circle Dr
25D Marshall-Adams Hall
East Lansing, MI 48824
emiliej@msu.edu

1 Introduction

The emergence of new institutions and technologies over the past decade that have made it easier for firms to obtain labor from self-employed contract workers instead of employees has led to widespread speculation that the U.S. labor market is undergoing a fundamental transition towards a “gig” economy. The dramatic rise of online platforms like Uber and Lyft—which were relatively obscure in 2012 but engaged hundreds of thousands of workers by 2016—has helped to feed the narrative that there has been a rise in self-employment being driven by gig work (Katz and Krueger, 2019a). To date, major labor force surveys like the Current Population Survey (CPS) show no increase in the self-employment rate since 2000; however, there is reason to think that such surveys might miss a shift towards a gig economy if workers do not perceive themselves as contractors or are more likely to do side jobs that are not well-captured by standard questionnaires (Katz and Krueger, 2019a; Abraham, Hershbein, and Houseman, 2020; Abraham, Haltiwanger, Sandusky, and Spletzer, 2020). In sharp contrast to trends in the CPS, the share of individuals reporting self-employment income to the Internal Revenue Service (IRS) on their income tax returns rose dramatically between 2000 and 2014.¹ We present these two measured trends side-by-side in Figure 1. Are these income tax returns picking up on a deep change in the labor market that major surveys currently miss?

While tax records can provide a critical resource for understanding the labor market, they are not guaranteed to be more reliable indicators trends than surveys. Unlike in confidential surveys, individuals have incentives to be strategic about what they report on tax filings—and both those incentives and reporting decisions may be prone to change over time. This is particularly true in the case of self-employment earnings reported on U.S. income tax returns which, unlike employment income, can be purely self-reported without any third-party verification (Slemrod, 2016). A large literature has documented that self-employment reporting on income tax returns is highly sensitive to incentives in the U.S. tax code and there is substantial “bunching” of reported self-employment earnings at levels that maximize refundable tax credits like the Earned Income Tax Credit (EITC), even as bunching behavior is absent in survey data (LaLumia, 2009; Saez, 2010; Chetty, Friedman, and Saez, 2013; Kuka, 2014; Mortenson and Whitten, 2020). Moreover, this type of strategic reporting has become increasingly prevalent since 2000 (Chetty, Friedman, and Saez, 2013; Mortenson and Whitten, 2020). These earlier findings raise the possibility that the same forces driving an

¹This trend in self-employment tax filings has been well-documented in earlier work (Abraham, Haltiwanger, Sandusky, and Spletzer, 2013; Jackson, Looney, and Ramnath, 2017; Abraham, Haltiwanger, Hou, Sandusky, and Spletzer, Forthcoming) and has been interpreted as suggestive evidence of a rise in alternative work arrangements not captured in traditional surveys (Katz and Krueger, 2019a)

increase in bunching behavior impact the reporting of self-employment work more broadly, independent of any underlying change in the labor market.

Drawing on the universe of IRS tax returns, this paper shows that changes in pure reporting behavior—given fixed labor market behaviors—contribute substantially to the observed increase in self-employment individuals report to the IRS. First, we draw directly on the information returns issued by firms to self-employed independent contractors (of which online-platform-based “gig” workers are a subset) and show that the rise in *taxpayer-reported* self-employment since 2000, visible in Figure 1, cannot be accounted for by an increase in *third-party-reported* contract work, inclusive of the rise of new types of platform-based work. Rather, we find the dramatic increases in taxpayer-reported self-employment are concentrated specifically among individuals in the phase-in range of refundable tax credits who have a strict incentive to report additional earnings beyond their employer-reported wages. Second, using a novel regression discontinuity design that exogenously varies exposure to negative marginal tax rates after labor supply decisions are sunk, we find that these incentives lead to increased reporting of self-employment earnings by taxpayers even with labor supply held constant. In contrast to kink-bunching designs, our design directly rules out real changes in labor supply behavior—we find that, in practice, only a small fraction of these reporting responses lead to bunching at kink points. Third, we present evidence that the magnitude of the pure-reporting response to fixed tax incentives has grown over time, and the same factors that explain increased EITC bunching over time (Chetty, Friedman, and Saez, 2013; Mortenson and Whitten, 2020) explain increased self-employment reporting rates more broadly. Finally, we show that our estimates imply that changes in reporting behavior account for much of the increase in the rise of self-employment reported on tax-returns; particularly the increase that in excess of any growth in firm-reported contract work.

We begin by drawing on the information returns issued to independent contractors—including app-based workers in the Online Platform Economy (OPE)—to assess the extent to which tax trends in self-employment reporting are plausibly driven by a rise in “gig” arrangements (Harris and Krueger, 2015; Abraham, Haltiwanger, Sandusky, and Spletzer, 2018; Farrell and Greig, 2018; Katz and Krueger, 2019a). A unique feature of U.S. tax data is that payments to self-employed contract workers are directly observed on information returns reported by firms to the IRS. We follow previous work and break out the subset of these returns issued by OPE firms to directly measure their contribution to overall trends in contract work (Jackson, Looney, and Ramnath, 2017; Collins, Garin, Jackson, Koustas, and Payne, 2019; Lim, Miller, Risch, and Wilking, 2019). While we observe an increase in independent contract work reported by firms on 1099 returns between 2000 and 2004 that is consistent with income tax return reporting during that period, we observe *no* growth in

1099 contract work over the decade between 2005 and 2014, during which time the share of individuals reporting self-employment on tax returns rose rapidly.² Notably, we find that prevalence of Online Platform Economy (OPE) earnings has increased dramatically but only since 2014 after the trend in taxpayer self employment had plateaued.

We then show that self-employment reporting on income tax returns has only become more prevalent among individuals with strict incentives to report additional self-employment earnings. Importantly, tax filers with children and employer-reported wage earnings in the EITC phase-in range (below the first “kink” in the EITC schedule) face *negative* tax rates on the marginal dollar—and would therefore strictly benefit from reporting some self-employment earnings, rather than none. These individuals might therefore increase their tax refunds by choosing to report informal earnings they might otherwise not have reported to the IRS, or, in some cases, by fabricating the income entirely.³ We find that the entire increase in the propensity to report secondary self-employment income among wage employees is driven by filers in this incentivized group. Meanwhile, such increases are absent for those with firm-reported wages even slightly above the “kink” point where the marginal tax rate becomes positive. Likewise, we find no differential growth above or below the first EITC kink point for individuals with no children who face negative marginal tax rates. Moreover, we find no such rise in the probability of having *firm-reported* contract income among incentivized individuals with low wages and children, only a rise in the propensity to *self-report* self-employment.

Do these increased propensities to self-report self-employment reflect real increases in self-employment work, or simply changes in reporting behavior given fixed labor-supply behavior? We pinpoint pure reporting behaviors holding labor supply fixed by examining a sharp discontinuity in EITC eligibility based on the date of birth of an individual’s first child. An individual’s EITC benefit for tax year t is calculated based on the number of children in their household during year t . Hence, a first-born child born on or before December 31 of year t would count towards the EITC calculation for that year t , while a child born only a few days later at the start of year $t + 1$ would not. This cutoff creates sharp differences in incentives when reporting income for tax year t . This motivates a regression discontinuity design (RDD) comparing year- t reported earnings for parents with births of their first children occurring

²As Abraham, Haltiwanger, Hou, Sandusky, and Spletzer (Forthcoming) note, the early increase in 1099-reported work between 2000 and 2005 might not be reflected in survey data to the extent that workers do not perceive themselves as self-employed.

³While highly strategic agents should “bunch” and report total earnings at the precise level that minimizes net tax liability, in practice individuals may choose to report some self-employment to improve their tax position without bunching, especially if they have concerns about appearing suspicious. Roughly a third of Americans report having some type of informal income (Bracha and Burke, 2018); it is likely that individuals have some discretion over whether or not to report such informal earnings.

right before or right after the end of tax year t .⁴ Crucially, if parents with births near the discontinuity face *ex ante* uncertainty about which year the birth will occur in when deciding on year- t labor supply, then all labor supply decisions will be fully *sunk* by the time the child’s eligibility for the EITC is learned. In that case, any observed differences in self-reported earnings around the cutoff reflect pure reporting adjustments holding underlying work behavior fixed.

We find clear evidence of a discontinuity in the probability of reporting self-employment income around the December 31 cutoff. Our estimates indicate that having a child unexpectedly qualify towards a year’s EITC calculation increases the probability of reporting self-employment income by 1.5 percentage points in years since 2010. By contrast, we find no such impact on the probability of having *firm-reported* contract payments or wage/salary earnings, indicating the main result reflects pure reporting behavior and not any change in labor supply.⁵ Reassuringly, we find that the entire self-employment effect is driven by individuals with wages in or below the phase-in range who face negative marginal tax rates on the marginal dollar of self-reported self-employment earnings.

Moreover, we find that these reporting effects have grown steadily over time. Between 2000 and 2014, the baseline effect size has grown nearly threefold from 0.5 p.p. to 1.5 p.p., and the effect conditional on having wages in the phase-in range has grown over fivefold from 1 p.p. to more than 5 p.p. We show that changes in the generosity of the federal EITC over the period we study cannot plausibly drive the observed rise. To interpret these changes over time, we present a framework in which increased knowledge of refundable tax credit leads an increasing share of individuals to report self-employment earnings on their tax return even without any changes in underlying labor supply. We show that, in our empirical setting, such increases in knowledge should be reflected in larger RDD effects over time. This framework is motivated by prior research documenting that even during periods when tax policy remains fixed, *knowledge* of the incentives created by the EITC and related tax provisions differs across regions and spreads gradually over time (Chetty, Friedman,

⁴This design is inspired by earlier work by LaLumia, Saltee, and Turner (2014), who find little evidence that mothers strategically time first births around the end of the year to minimize their tax burden, but who do find that mothers with Schedule C income whose children are born in December are much more likely to bunch at the first EITC kink in that tax year than those with children born the following January. Feldman, Katuščák, and Kawano (2016) examine a similar discontinuity to study parent responses to children who age out of the Child Tax Credit at 17. A key difference in their setting is that their work studies changes in real labor supply behavior during the year *after* the age-out, whereas we focus on pure reporting behavior at the end of *same* tax year that eligibility changes.

⁵Unlike self-reported earnings, third-party-reported payments cannot be manipulated by individuals to avoid taxation; see, for example, Kleven, Knudsen, Kreiner, Pedersen, and Saez (2011). If individuals were strategically timing births to maximize EITC credits, then we should see individuals with births in December have firm-reported earnings closer to the one-child refund-maximizing level. However, we find no evidence that either firm-reported wage earnings or non-employee compensation vary around the threshold date.

and Saez, 2013). To test the plausibility of this channel we use sharp-bunching rates among individuals in a hold-out sample for each three-digit ZIP code in each year as a proxy for local knowledge of EITC incentives following Chetty, Friedman, and Saez (2013), and estimate an interacted version of our RDD specification. Consistent with our framework, we find that high-knowledge regions have dramatically larger effects than low-knowledge regions and that cross-sectional differences in knowledge and within-region changes in knowledge are associated with similar increases in the reporting effect. Our evidence implies that the behavioral impacts of spreading knowledge on self-employment reporting extend far beyond individuals engaging in highly strategic kink-bunching behavior.

Motivated by our conceptual framework, we conclude the paper with several counterfactual exercises examining the quantitative contribution of these reporting trends to the observed overall increase in self-employment tax filings since 2000. We find that our RDD estimates explain about a quarter of the rise in the share who self-report self-employment between 2000 and 2018, but account for nearly half of the increase after 2005 when trends in self-reported data and third-party reported data diverge most sharply. Because the short-run behavioral responses identified in our RDD results potentially underestimate longer-term adjustments in behavior, we also consider a more speculative scenario in which rates of self-employment reporting among individuals EITC-eligible children counterfactual increased by the same amount as for individuals without children within narrow wage-bins and demographic cells. This counterfactual scenario accounts for roughly half of the rise in self-reported self-employment rates between 2000 and 2018, and accounts for almost all of the increase in self-reported self-employment in excess of what is observed in third-party-reported data. Reassuringly, this counterfactual adjustment does not affect trends in third-party-reported data.

Our findings ultimately offer a cautionary tale about measuring labor-market trends with tax return data and other types of administrative data (Blank, Charles, and Sallee, 2009; Slemrod, 2016) without careful consideration of reporting incentives. While surveys are imperfect instruments, trends observed in some types of administrative data may be impacted by reporting incentives and should be interpreted accordingly. In our setting, we show that such incentives give the *appearance* of a shift in the nature of work in income tax return data that is belied by more reliable third-party reporting by firms themselves.

This paper is organized as follows: in the next section, we discuss the measurement of self-employment in administrative tax data. We also discuss important institutional details of the U.S. tax code that relate to incentives to report self-employment. Section 3 presents important motivating facts on self-employment trends. Section 4 introduces our main research design to estimate the response of self-employment to reporting incentives. Section 5

discusses knowledge of the tax code as a driver of the change. Section 6 provides new estimates of self-employment, which control for reporting trends. We also present an application of our new data contributions, using our insights to examine the response of self-employment to the business cycle. Section 7 concludes.

2 Data and Institutional Background

2.1 U.S. Tax Data on Employment and Self-employment Earnings

Our analysis draws on the universe of tax returns filed with the IRS (Internal Revenue Service 2022). In addition to the Form 1040 income tax returns filed by individuals, we also observe third-party reporting on information returns that are filed by firms whether or not an individual files an income tax return. Wage and salary earnings for each individual are reported directly to the IRS by employers on W-2 information returns; with few exceptions, all employment earnings individuals report on their income tax returns are covered by W-2 reporting. In contrast, self-employment proceeds are self-reported to the IRS by workers on their 1040 returns and are not necessarily covered by third-party information reporting. In the U.S. tax system, self-employment earnings are technically active income from a “sole-proprietorship” business. Self-employed workers must keep track of their revenues, expenses, and net profits, and report those amounts on Schedule C of their 1040 return.⁶ If an individual’s self-employment proceeds exceed a threshold level of \$433, they must file Schedule SE and pay Social Security and Medicaid (“SECA”) taxes—the SECA tax amount is equivalent to the total payroll tax that would be withheld from the same level of wage and salary earnings.⁷ Prior work typically measures self-employment rates in tax data by comparing annual self-employment earnings on Schedules C or SE to employment earnings on W-2 forms.⁸ In our analysis, we primarily focus on Schedule SE reporting because it is available at the individual level for all years in our data, whereas Schedule C items are only available at the tax-unit level prior to 2007.

While not all payments to self-employed workers are covered by third-party reporting, many such payments are. Specifically, when a self-employed worker performs services for a *firm*, the firm must report all gross payments to self-employed contractors in excess of \$600 on 1099-MISC Box 7 (replaced by 1099-NEC in 2020). While these returns do not capture

⁶Self-employment earnings from farming are reported on Schedule F instead of Schedule C.

⁷Self-employed individuals are responsible for paying the equivalent of both the employer and employee portion of payroll taxes, which together are 15.3 percent. Half of payroll taxes paid (the “employer”-share) is deductible. Thus, the effective marginal tax rate on self-employment earnings is $0.153 \times (1 - 0.0765) = 14.1\%$.

⁸For example, Abraham, Haltiwanger, Sandusky, and Spletzer (2020) focus on Schedule SE filers, while Jackson, Looney, and Ramnath (2017) focus on Schedule SE and Schedule C filers.

self-employment earnings derived from providing services or selling to individual customers, they are nonetheless useful in two important respects. First, they enable us to observe whether trends in self-employment are plausibly driven by firms shifting from hiring workers to contracting with self-employed freelancers.⁹ Second, 1099s provide information on a major class of self-employment revenues that is third-party reported and dependent on individual reporting behavior. However, unlike W-2 returns which report workers’ net employment earnings, 1099s report *gross revenues* rather than *profits after expenses*. Hence, recipients can still influence their taxable self-employment earnings through their self-reported expenses on Schedule C.

The tax data further enable us to pay special attention to a new and growing class of independent contract work mediated by online platforms. We refer to these arrangements—which are a subset of the broader “gig” economy—as the “online platform economy” (OPE). We focus specifically on labor platforms where workers directly provide services to others (for example, ridesharing or delivery) as opposed to platforms on which individuals sell goods or rent capital (for example, craft merchandise sites or homesharing). OPE earnings are reported to the IRS either as independent contractor earnings on 1099-MISC/NEC, or, in some cases, as vendor revenues on 1099-K. We identify 1099 forms issued from OPE companies using the method in Collins, Garin, Jackson, Koustas, and Payne (2019); we discuss several important measurement issues in the Data Appendix.

Following earlier work, we focus primarily on the self-employment rate defined as the share of individuals with labor earnings reported on either W-2 or Schedule SE who have positive annual self-employment net earnings reported on Schedule SE.¹⁰ One should note that this self-employment rate may not correspond exactly to the self-employment rate obtained from workforce surveys like the Current Population Survey (CPS). In the CPS, workers self-identify as employed or self-employed based on their predominant activity in a reference week. “Point-in-time” measures will generally lead to lower estimates of the prevalence of self-employment work—in particular, they under-count secondary, informal self-employment and firm-facing self-employment (National Academies of Sciences and Medicine, 2020; Abraham, Hershbein, and Houseman, 2020; Abraham, Haltiwanger, Sandusky, and Spletzer, 2020). Even surveys that record retrospective annual earnings like the CPS Annual Social and Economic Supplement (ASEC) may nonetheless under-count self-employed workers if such workers fail to report small supplemental earnings or if independent contractors incorrectly

⁹We explore this workforce in depth in earlier work (Collins, Garin, Jackson, Koustas, and Payne, 2019; Garin, Jackson, Koustas, and Miller, 2023).

¹⁰We focus on Schedule SE reporting both because Schedule SE information is available at the individual level for our full sample period while Schedule C information is not, and for comparability to analysis of Schedule SE trends in Abraham, Haltiwanger, Hou, Sandusky, and Spletzer (Forthcoming).

perceive themselves as employees.

In theory, the information reported on tax returns should be more comprehensive than information on surveys, since individuals face significant penalties if they misreport their income. In reality, enforcement is imperfect and individuals have considerable scope to strategically under-report—or over-report—income to minimize their tax burden and maximize refundable credits. Thus, changes in strategic reporting behavior over time can potentially drive trends in tax data, whereas confidential surveys are often immune to such concerns.

2.2 Incentives in the Tax Code to Report Self-Employment

Certain provisions in the tax code may incentivize certain individuals to report self-employment profits. In particular, the Child Tax Credit (CTC) and Earned Income Tax Credit (EITC) phase-in with higher earnings up to a threshold level where there is a first “kink” in the EITC schedule as the maximum credit level is obtained. Both this threshold level and the effective subsidy rate depend on the number of children claimed on a tax return.¹¹ The phase-in rate of the EITC is 7.65 percent for childless households, 34 percent for families with 1 child, 40 percent for families with two children, and 45 percent for families with three children. The CTC further increases the effective phase-in rate by 15 percentage points.

Individuals with children face net negative marginal tax rates in the EITC and CTC phase-in range, even after incorporating all federal taxes including payroll (FICA/SECA) taxes.¹² These subsidies are sometimes enhanced by additional credits provided by states. By contrast, the net subsidy disappears once earnings are above the applicable phase-in range, since additional earnings are subject to the payroll tax but no marginal subsidy.¹³ Moreover, households with no children *always* face a positive marginal tax rate on earnings, as the payroll tax rate exceeds the lower EITC phase-in rate.¹⁴

Crucially, these negative marginal tax rates create an incentive for some individuals to report self-employment income that may not have been reported otherwise. In particular, individuals with EITC-eligible children under 19 and W-2 wages below the first EITC kink

¹¹The CTC applies to children under 17, and the EITC applies to children under 19.

¹²Every dollar of self-employment earnings reported is subject to the 14.1 percent SECA rate (net of the deductible part of the taxes paid). In the EITC phase-in-range, the increase in Federal credits per dollar reported exceeds the increase in SECA liability, resulting in a net increase in one’s refund—in this case, SECA taxes simply reduce the net refund.

¹³Of course, an important incentive to report self-employment income to tax authorities and pay payroll taxes is to contribute to future Social Security benefits. However, these incentives are faced by all taxpayers, not just those facing negative marginal tax rates. Moreover, the increase in Social Security benefits may be less salient to myopic consumers, and since benefits are based on the highest 35 years of earnings, earnings for young workers are likely to have little impact on future benefits.

¹⁴Even with the most generous state-EITC for childless individuals in the District of Columbia, they nonetheless face a marginal tax rate of 0 in the phase-in range.

are strictly better off reporting self-employment income on their 1040 return. In theory, such individuals could report self-employment earnings to maximize their net subsidy—this sort of hyper-strategic behavior would result in “bunching” at the first EITC kink point (Saez, 2010).¹⁵ In practice, though, these tax incentives may lead individuals to report additional self-employment income without going so far as to report this “optimal” amount. For example, many individuals have actual informal income (Bracha and Burke, 2018) that they may choose to report in full that they may not have reported otherwise. Additionally, strategic taxpayers may avoid reporting at the refund-maximizing level if they think it might appear suspicious to auditors.¹⁶

Appendix Figure A.1 shows the effective tax rate on reporting a first dollar of self-employment income at different levels of W-2 wage earnings, taking into account all federal taxes. Households without children face a positive marginal tax rate on self-employment income across the wage distribution. In contrast, for families with children, additional income at amounts below the credit-maximizing kink point is taxed at a negative tax rate. Households with three children with wage income below the threshold for the minimum income to maximize the EITC (\$13,870 in \$2015), and above the minimum income threshold for the CTC, face a negative tax rate that implies a credit of up to 41 cents on the dollar.¹⁷ Similarly, tax units with two children face a marginal tax rate of up to -0.37 and tax units with one child face a negative marginal tax rate of up to -0.31. As we will show below, these incentives play an important role in whether households report self-employment income in practice.

2.3 Sample Construction

In our analysis, we use de-identified full-count tax records incorporating both filer-reported returns and information returns. Most of our analysis uses data from 2000-2018. While some microdata are available back to 1996, 2000 is when information returns are first available for 1099 independent contractors.

We determine family structure based on information in tax filings as well as links from

¹⁵While the first EITC kink point is the refund-maximizing level in most years, Mortenson and Whitten (2020) note that temporary tax credits shift the level in some years.

¹⁶We note that individuals that would have reported low levels of self-employment earnings in the credit phase-in range, regardless of incentives, face an incentive to report additional self-employment income. However, their additional self-employment earnings would not impact the extensive margin of self-employment, which is our focus. We also note that individuals with W-2 earnings close to the first EITC kink point face positive marginal tax rates on any self-employment earnings that exceed the kink level, which mitigates the incentive to report additional amounts.

¹⁷For three children, the phase-in rates on the EITC and CTC are 0.45 and 0.15, respectively. After deductions, the overall marginal tax rate for a tax unit with wage earnings above \$2,500 and below the standard deduction is: $(1-0.0765)*(-0.45-0.15+0.153) = -0.413$.

parents to children from the Social Security Administration. These child links from the SSA provide comprehensive coverage for children born since the early 1980s who were issued a Social Security number, providing a complete picture of the number of children under 18 years of age by 2000. More information on how we clean and process the data is provided in the Data Appendix.

3 Trends in Self-Employment Reporting

Can shifts among firms away from traditional employment to freelance and platform-based work explain the rising share of the workforce with self-employment earnings on Schedule SE shown in Figure 1? To answer this question, we can directly look at the information reported by firms on 1099 returns. In Figure 2, we compare the evolution of the share of the workforce with Schedule SE earnings to the share with payments for contract or platform-based work on 1099 returns.¹⁸ To highlight the contribution of the nascent online platform economy, the figure breaks out the share of 1099 recipients who only get such returns from online platforms.¹⁹

As highlighted in the introduction, the share of the workforce with self-employment earnings on Schedule SE rose consistently after 2000, climbing 2 percentage points (a 20 percent increase over the 2000 share) before plateauing after 2014. However, the trend in 1099-reported contract work tells a different story. For the first few years after 2000—well before the advent of app-based platform work—the rise in the share of workers with Schedule SE earnings was paired with a comparable rise in the share with payments reported on 1099 returns.²⁰ However, beginning around 2004, a stark divergence between the two series

¹⁸In Figure 2 we follow Collins, Garin, Jackson, Koustas, and Payne (2019) and expand our definition of the “workforce” to include all individuals who receive wage (W-2) earnings, contract earnings (1099-MISC or OPE 1099-K), and/or self-employment (Form 1040 Schedule SE) earnings. As in Collins, Garin, Jackson, Koustas, and Payne (2019), we only include individuals with 1099-reported earnings when they receive a W-2 or file a tax return.

¹⁹ After 2016, several online platform companies reported payments to gig workers on form 1099-K, adhering to the higher \$20,000 reporting threshold for that form. In Figure 2, the 2017 and 2018 levels of total 1099-reported work reflect an imputation of how trends in online platform work would have evolved if reporting thresholds had remained constant described in Garin, Jackson, and Koustas (2022) and Garin, Jackson, Koustas, and Miller (2023); this series reproduces the corresponding series in Figure 4 of Garin, Jackson, and Koustas (2022). This imputation based on trends in state-level 1099-K data in Massachusetts and Vermont with a lower \$600 threshold in those states; see Garin, Jackson, Koustas, and Miller (2023) for details on the construction of the imputed series. While the imputed series approximates the number of workers engaged in such work, it overstates the number of these workers for whom a 1099 return was actually filed. The firm-reported contract work series without this adjustment is shown in Fig A.2.

²⁰The rise in 1099-reported contract work between 2000 and 2005 is consistent with survey data from the CPS Contingent Worker Supplement, in which the share of workers reporting being independent contractors increased by 1 percentage point between 1999/2001 and 2005 (Appendix Figure A.3).

occurs—whereas the share with earnings on Schedule SE continued to rise through 2014, the share with 1099 payments remained virtually unchanged from 2004–2014. It is only in 2014 that we begin to observe the explosive rise in platform-based “gig” work, just as the Schedule SE share *stops* growing. Accordingly, the increase in Schedule SE reporting can be roughly broken out into three phases: a first phase through about 2004 during which self-reported self-employment largely reflects growth observed using information returns; a second phase from 2004 to 2014 prior to the advent of platform work during which self-reported self-employment became increasingly prevalent but 1099-reported work did not; and a third phase after 2014 characterized by the emergence of the platform economy and yet no increase in the prevalence of self-reported self-employment.

One immediate takeaway from Figure 2 is that new forms of “gig” work cannot explain the rising prevalence in taxpayer-reported self-employment earnings. Payments from platform-based “gig” work only become quantitatively important after 2014, after the large rise in Schedule SE reporting had already ceased. More generally, while the rise in Schedule SE reporting may reasonably have been driven by a shift towards freelance work reported on 1099 returns early on, the substantial increase in the Schedule SE share between 2004 and 2014 occurred despite the 1099 share remaining unchanged. This finding raises an important question: if there was no increase in freelance and other gig work reported by firms, did self-employment really become more common over this period, or might the trend be driven by other factors? One possible explanation is that customer-facing self-employment work that is not subject to 1099 reporting continued to grow even as 1099-reported work plateaued. Another possibility is that the propensity to *report* self-employment earnings on income tax returns changed over this period.

Figure 2 also raises the question: why does the surge in platform-based work after 2014 not result in higher levels of Schedule SE reporting? First, as documented in other work (Collins, Garin, Jackson, Koustas, and Payne, 2019; Garin, Jackson, and Koustas, 2022; Garin, Jackson, Koustas, and Miller, 2023), the vast majority of new platform workers are ride-hailing or delivery drivers who do platform work sporadically and are only paid small amounts of gross income over the course of the year.²¹ In addition, when such filers file Schedule C, they typically expense over half of their gross payments. As a result, many workers have net earnings below the Schedule SE filing threshold after deducting expenses on their Schedule C. Second, platform workers are less likely to file a Schedule C than other types of contractors with non-employee compensation on a 1099-MISC in the first place. On

²¹Moreover, as noted in footnote 19, the 2017 and 2018 levels of online platform mediated work in 2 are imputed to display the expected number of workers who would have been sent 1099 returns if reporting thresholds had remained constant. Many such workers may not have received a 1099 return in practice, and thus may not have been aware of an obligation to report those earnings on an income tax return as a result.

one hand, these low rates of Schedule C filing may reflect higher rates of non-compliance among platform workers; but, on the other hand it is also plausible that many workers with 1099s from gig platforms have sufficiently low net earnings after expenses that they do not need to file. Finally, we note that we *do* observe an increasing number of workers who have 1099s from gig platforms and report their earnings on Schedule SE, but this reflects a shift among the self-employed towards platform work and away from other types of self-employment such that the aggregate share of workers with Schedule SE does not rise on net.

To probe whether rising rates of Schedule SE reporting might be driven by reporting incentives, we next examine whether the growth in Schedule SE reporting was concentrated among taxpayers who have stood to gain from such reporting. As discussed in Section 2.2, individuals with dependents on their tax return and household W-2 wage earnings below the top of the phase-in range for the EITC and CTC have an incentive to report additional income in order to receive a larger refundable credit. As an initial piece of evidence, Figure A.4a shows a massive increase in the Schedule SE share among workers who reported dependents and claimed the EITC, but had no comparable increase in the share with contracts payments reported on a 1099 return. For all other workers, we find the two series evolve roughly in parallel through 2014. Likewise, in Figures A.4b and A.4c we observe a similar evolution of both the Schedule SE share and the 1099 share (excluding platform-based gig work) for all individuals in households with over \$50,000 in labor earnings regardless of their dependants and likewise all individuals without kids regardless of their earnings.²²

The evidence is more striking when we plot 2000–2014 changes in the Schedule SE share by number of children and within narrow bins of tax unit combined W-2 wage earnings (in constant 2015 dollars) in Figure 3 Panel A. We find dramatic growth in self-employment concentrated among households with children and combined wages below the first EITC kink point, indicated by the maroon and green lines on the figure—these are exactly the households that gain from reporting some small amount of self-employment rather than none. Remarkably, growth in the Schedule SE share disappears precisely at the level of W-2 earnings corresponding to the top of the EITC phase-in range *specific to the number of children on the tax return*. Meanwhile, we observe no increase in Schedule SE filing among workers with children and W2 earnings above these thresholds. As a point of comparison, Panel B reports the corresponding change in the share with firm-reported payments on 1099 returns. Unlike in Panel A, we find no sharp differences in growth rates around the top of the credit phase-in ranges indicated by the maroon and green lines. For individuals without children, we observe only small increases in the share with Schedule SE earnings in Panel

²²Figure A.4d reports additional trends by gender, showing that most of the increase occurs among women.

A; yet, in contrast to individuals with children, we observe nearly identical increases in the share with 1099-reported earnings in Panel B.

We further show in Appendix B that the changes in self-employment reported on 1040-SE occurs precisely at the time of one's first childbirth, when they first face reporting incentives, and continues to grow in subsequent years. Moreover, we show that this increase in self-employment at the time of childbirth has grown steadily since 2000. However, we find no change in *firm-reported* contract income around childbirth. While suggestive that households are motivated by EITC incentives to report self-employment income, these findings could also reflect real increases in types of self-employment activity not reported on 1099 forms. The need to distinguish between real labor supply changes and pure reporting behavior motivates our research design in the next section.

4 Isolating Pure Reporting Behavior

4.1 A Stylized Model of Reporting Behavior

The findings in the previous section suggest that tax code incentives are potentially an important driver of the rise in the share of workers with self-reported self-employment earnings. One possibility consistent with these results is that a growing share of workers are *reporting* self-employment earnings in response to tax code incentives without any underlying change in real behavior that would be captured on information returns. However, the results so far are also consistent with *real* increases in self-employment work driven by tax code incentives that is limited to work done serving or selling to individuals rather than firms and therefore not reported on 1099 returns. We now formalize each of these stories in a stylized model to examine the conditions under which one can distinguish between them empirically.

4.1.1 Set-up

We consider a static model with L individuals denoted $i = 1, 2, \dots, L$ who engage in wage work, self-employment work, or both. Each worker i optimally chooses an amount of wage earnings w_i and an amount of self-employment net earnings s_i based on their preferences, labor market opportunities, and perceptions of the tax code. Total earnings are subject to taxes. However, while wages w_i are reported directly to the government by employers, workers have discretion over amount of self-employment earnings z_i to self-report to the government.²³ Taxes are then assessed on total earnings reported to the government $r_i = w_i + z_i$.

²³For simplicity, we abstract away from 1099 reporting and consider the case where no self-employment earnings are subject to third-party reporting.

For all workers, total reported earnings r_i are subject to an income tax at a constant rate $\tau_0 > 0$, which is common knowledge. In addition, there is a phased-in refundable tax credit for individuals with children, but awareness of this credit is limited. For individuals with children, the credit is phased-in at a constant rate up to a threshold level of reported earnings $\hat{r}$, above which it is phased-out at a constant marginal rate. We assume the phase-in rate exceeds the baseline payroll tax rate, so that workers with children (denoted as “eligible” by $E_i = 1$) face an effective tax rate of $\tau_1 < 0$ on earnings up to $\hat{r}$ and an effective tax rate of τ_2 where $\tau_2 > \tau_0 > 0 > \tau_1$ for earnings above $\hat{r}$. However, only a share λ of workers are aware of this credit (these individuals are denoted “informed” by $I_i = 1$); all uninformed individuals ($I_i = 0$) perceive the tax system to be a flat rate $\tau_0 > 0$ regardless of the number of children one has.

We do not explicitly model labor supply decisions, but note that choices of w_i and s_i depend on both the tax code perceived by individual i as well as idiosyncratic labor supply preferences, and preferences may in turn depend on the number of children one has. In this set-up, individuals with a given number of children may choose a wide range of values of w_i and s_i . Moreover, we allow for the possibility that choices over w_i and s_i depend on the number of children one has. For purposes of exposition, we make two simplifying assumptions. First, we assume that true labor supply choices w_i and s_i are independent of compliance types as defined below; this would follow from the timing assumptions in Allingham and Sandmo (1972) in which agents choose reporting amounts given a predetermined earnings level. More specifically we assume that individuals make *actual* labor supply decisions w_i and s_i expecting to fully comply with the tax code.²⁴ Second, while we allow that becoming informed about the refundable credit may lead people to change their labor supply so that their true earnings move closer to the refund maximizing level $\hat{r}$ as observed in Chetty, Friedman, and Saez (2013), we assume that becoming informed does not lead any individual to switch from earning $w_i > \hat{r}$ to earning $w_i < \hat{r}$ or vice-versa.²⁵

4.1.2 Reporting Behavior and The Measured Self-Employment Rate

Individuals realize their compliance type, becoming either a *non-complier* (denoted by $N_i = 1$) with probability θ or a *complier* (denoted by $N_i = 0$) with probability $1 - \theta$. Honest types ($N_i = 0$) report their self-employment profits honestly $z_i = s_i$. Non-compliant individuals

²⁴This assumption limits the number of cases we need to consider. One should note that, even under this assumption, choices w_i and s_i are allowed to respond to changes in the perceived tax code.

²⁵This is a weak assumption; while individuals have a strict incentive to “bunch” their earnings towards the refund-maximizing level, there is no incentive to adjust incomes past that level. This assumption therefore does not bind in cases where individuals can frictionlessly choose earnings, but rather rules out cases in which adjustments are “lumpy” such that individuals might adjust earnings beyond the ideal point. This allows one to treat the share of individuals with earnings in the credit phase-in range as exogenous over time.

report z_i strategically and without regard to their true earnings level. This complier type N_i reflects factors such as preferences over voluntary compliance, perceptions of risk of detection, and varying degrees of risk-aversion in the face of potential fines.²⁶

We summarize the reporting behavior of non-complier types in the following proposition:

Proposition 4.1. *Given the incentives in the tax code, non-complier types report positive self-employment earnings if and only if they have children ($E_i = 1$), have wages below the top of the credit phase-in range ($w_i < \hat{r}$), and are aware of the tax credit ($I_i = 1$).*

Proof. See Appendix D. □

Going forward, it will be useful to denote individuals with children and wages below the “kink” in the tax schedule with the indicator $BTK_i = 1$ and to denote the share of workers with $BTK_i = 1$ by κ .

We denote the *true* share of workers with positive self-employment earnings by σ and the *reported* share of workers with positive self-employment earnings by ρ . We let σ^0 and the σ^1 denote the true self-employment shares among individuals with BTK_i equal to 0 and 1, respectively, and likewise for the tax-payer reported shares ρ^0 and the ρ^1 , such that the overall shares can be expressed as $\sigma^{tot} = \sigma^0(1 - \kappa) + \sigma^1\kappa$ and $\rho^{tot} = \rho^0(1 - \kappa) + \rho^1\kappa$. Given Proposition 4.1, these reported shares can be characterized as follows:

Proposition 4.2. *The share of workers with $BTK_i = 0$ reporting self-employment earnings is given by*

$$\rho^0 = (1 - \theta)\sigma^0 \tag{1}$$

The share of workers with $BTK_i = 1$ reporting self-employment earnings is given by

$$\rho^1 = (1 - \theta)\sigma^1 + \theta\lambda \tag{2}$$

The overall share of workers reporting self-employment earnings is given by

$$\rho^{tot} = (1 - \theta)\sigma^{tot} + \kappa\theta\lambda \tag{3}$$

Proof. See Appendix D. □

Several important observations follow from Proposition 4.2. First, in the absence of eligibility for or knowledge of the refundable tax credit, the reported self-employment share

²⁶Since our focus is on the extensive margin of reporting, we do not model the exact reporting amount in detail as in Allingham and Sandmo (1972) and instead simplify the analysis without loss of generality using a two-type model.

will typically fall *below* the true share with self-employment income due to the behavior of non-compliers. Hence, even with widespread knowledge of the credit, declines in the overall compliance rate θ have ambiguous effects on overall reporting because reporting rates should decline among the majority of workers with $BTK_i = 0$. In the following section, we argue that there is no evidence that θ has been changing over our study period. Second, while awareness of the credit will result in a higher reported self-employment share, it is ambiguous whether the reported share would ever exceed the true self-employment share in the aggregate. Finally, the total reported self-employment rate will tend to be counter-cyclical since rising employment rates and wage levels will reduce the share of the workers facing negative marginal tax rates κ .

Accordingly, increases in the self-reported self-employment rate that are concentrated among those with $BTK_i = 1$ as observed in Section 3 above are therefore likely to be driven either by i) changes in the true self-employment rate reported by honest types (σ^1) or ii) changes in credit awareness among eligible individuals (λ) that impact strategic reporting. In practice, it is difficult to empirically distinguish between these two stories since the factors that determine credit eligibility—namely, the presence of children in the household—likely also directly impact labor supply decisions. In order to isolate strategic reporting behaviors from changes in actual labor supply, one would ideally randomly vary credit eligibility *after* the end of the tax year (but before taxes are filed) once labor supply decisions are already sunk and then test for impacts on reported self-employment earnings. In the next section, we show we can approximate this ideal experiment using a regression discontinuity design—and, by implementing it over time, we can assess the contribution of increased credit awareness to rising self-reported self-employment rates.

4.2 Regression Discontinuity Design

To isolate reporting responses to refundable credits, we examine a sharp discontinuity in EITC and CTC eligibility based on the date of birth of an individual’s first child. An individual’s EITC benefit for tax year t is calculated based on the number of children in their household during year t . A first-born child born on or before December 31 of year t would count towards the EITC calculation for that year t , creating an incentive for low-wage parents to report additional self-employment income on tax day. By contrast, a child born only a few days later at the start of year $t + 1$ would not count towards the EITC calculation in year t and their parents would have no such incentive. Because exact birth dates are difficult to precisely forecast far in advance, parents expecting a first child close to the end of the year are uncertain about what year the birth will occur in—and the corresponding

EITC status for year t —until labor supply decisions for year t are sunk.

This motivates a regression discontinuity design (RDD) comparing year- t self-employment earnings for parents with their first birth right before and right after the end of tax year t . If, in the limit—examining births right before and after midnight on December-31—parents face complete *ex-ante* uncertainty about the year of birth, they should make identical labor supply decisions in year t . However, *ex-post* after the children are born, the two sets of parents face different returns to self-reporting self-employment. Accordingly, all differences in self-reported tax year t self-employment between parents on either side of the discontinuity will be due entirely to differences in reporting behavior.

In the parlance of the model above, true self-employment rates σ should be equalized for individuals with births just before or after the cutoff and all differences should reflect differences in strategic reporting among informed non-compliers. More specifically, Proposition 4.2 implies one should observe no RDD effect among individuals with $BTK_i = 0$ and that the RDD effect among individuals with $BTK_i = 1$ identifies the share of informed non-compliers $\theta\lambda$.²⁷ This latter claim assumes first-time parents' awareness of the tax credit is similar to that in the broader population. However, for any level of general credit awareness in the population, individuals may be entirely inattentive to the credit before they have children and may only partially internalize knowledge of the credit and reporting incentives in the immediate aftermath of a first birth—in that case, the RDD effect among individuals with $BTK_i = 1$ might under-estimate the value of $\theta\lambda$ that is representative of the broader population of individuals with $BTK_i = 1$.

The key identifying assumption is that true year- t labor supply is identical for parents on either side of the December 31 cutoff. This assumption would be violated if parents are able to schedule births around New Year's Eve to maximize their tax refunds given their *true* year- t

²⁷Formally, let $w_i(d)$ and $s_i(d)$ denote the true labor supply choices individual i would make during the tax year in the scenario the birth occurred d days from the end of that tax year. With uncertainty over the exact timing of birth, individuals with births close to $d = 0$, will have sunk all labor supply decisions by the time uncertainty is resolved, such that $\lim_{d \uparrow 0} w_i(d) = \lim_{d \downarrow 0} w_i(d)$ and $\lim_{d \uparrow 0} s_i(d) = \lim_{d \downarrow 0} s_i(d)$. Consequently the probability that any individual is truly self-employed is equalized at the limit $\lim_{d \uparrow 0} \sigma(d) = \lim_{d \downarrow 0} \sigma(d)$, where $\sigma(d)$ is the true self-employment rate for individuals with births d days from the end of the tax year. In combination with Proposition 4.2, this implies:

$$\beta^{RDD, BTK=0} = \lim_{d \uparrow 0} \rho^0(d) - \lim_{d \downarrow 0} \rho^0(d) = (1 - \theta) \lim_{d \uparrow 0} \sigma^0(d) - (1 - \theta) \lim_{d \downarrow 0} \sigma^0(d) = 0$$

and

$$\beta^{RDD, BTK=1} = \lim_{d \uparrow 0} \rho^1(d) - \lim_{d \downarrow 0} \rho^1(d) = (1 - \theta) \lim_{d \uparrow 0} \sigma^1(d) + \theta\lambda - (1 - \theta) \lim_{d \downarrow 0} \sigma^1(d) = \theta\lambda$$

where $\rho^0(d)$ and $\sigma^0(d)$ are the reported and true self-employment rates for individuals with $BTK_i = 0$ and births d days from the end of the tax year, respectively, and $\rho^1(d)$ and $\sigma^1(d)$ are corresponding entities for individuals with $BTK_i = 1$.

labor supply.²⁸ In our analysis, we test for strategic birth timing by examining discontinuities in *firm-reported* earnings, which should exist if individuals are sorting. Additionally, we plot the distribution of births around the end of tax years 2011-2018 (corresponding to our benchmark sample below) in Appendix Figure A.5 and find that the distribution of births is mostly smooth, with expected decreases around the Christmas and New Year’s holidays. To ensure that selective avoidance of births during the New Year’s holiday does not influence our results, we omit a three-day bandwidth “donut hole” from our baseline analysis sample and examine robustness to the inclusion of the donut hole and alternate bandwidths.

We estimate regressions of the form:

$$y_i = \alpha + \beta \mathbf{1}\{date_i \in December\} + f(date_i) + \epsilon_i \quad (4)$$

on the sample of childbirths in December and January. In this specification, $date_i$ is the running variable—which is the child’s birth date measured as days since December 31—and outcome y_i is measured in tax year t . In our baseline analysis, we include births in a fifteen-day bandwidth around midnight on New Year’s Eve, omitting births in the three-day bandwidth “donut hole” around the start of the new year in case of any potential shifting around the holiday, and let $f(\cdot)$ be a linear function allowing the slope to change around the cutoff.²⁹ We also examine alternative bandwidths and polynomial specifications as robustness checks. The coefficient β on the indicator $\mathbf{1}\{date_i \in December\}$ is the RDD effect of having a child in tax year t instead of tax year $t + 1$. We estimate Equation 4 on all parents with first births in the specified window as identified in Social Security birth records. This sample includes known parents listed in Social Security records, regardless of whether or not that parent claims the child on their tax return.³⁰

In our baseline analysis, we pool recent years (tax years 2011–2018) when self-employment reporting rates were highest. We subsequently estimate RDD effects for individual tax years

²⁸Early work by Dickert-Conlin and Chandra (1999) documented a correlation between tax rates and birth timing. More recent work by LaLumia, Sallee, and Turner (2014) tests for such behavior using administrative tax records and, while they find some response of the timing of births to tax incentives, the effect is concentrated almost entirely among second and later births—meanwhile, the effect of tax incentives on first-birth timing is negligible. However, LaLumia, Sallee, and Turner (2014) also find that self-employed parents with births in December are much more likely to have earnings for that year that are bunched around the first EITC kink point than parents with births in the following January, consistent with some difference in reporting behavior depending on the year of birth. Our design tests for broader reporting responses beyond highly-strategic bunching behavior.

²⁹We use robust standard errors rather than clustering standard errors by date. Kolesár and Rothe (2018) show that SEs clustered by a discrete running variable “do not guard against model misspecification, and [...] have poor coverage properties”.

³⁰We examine all parents with known births. However, some births in the data are missing identifiers for one or both parents. Missing identifiers are more common for fathers than for mothers in the birth data.

to examine changes in effects over time.

4.3 Baseline Results on Reporting Effects

We begin by examining the key identifying assumption that true labor supply decisions and other individual attributes are smooth across the end-of-year discontinuity. In Panel A of Table 1, we show that there is no sorting across the discontinuity by gender, age, or pre-period earnings. A stronger prediction of our identifying assumption is that there should be no effect on *third-party* reported income in year t . The results in Table 1 Panel B show no effect on the propensity to have self-employment income from independent contracting work reported on a 1099 return employment income reported on a W-2 return. We also find no effects on the *level* of W-2 reported employment earnings in the tax unit or whether those earnings are above the first EITC kink for single-child filers. To more directly test whether parents with the most to gain from claiming an EITC-eligible child systematically shift birth dates before the end of the year, we calculate a “predicted EITC credit” similar to Chetty, Friedman, and Saez (2013) that is the credit each individual would earn if their birth occurred in December of tax year t based on their and their spouse’s (if married filing jointly) W-2 wages from year t and test for differences around the end-of-year cutoff. If individuals are sorting to maximize their EITC, then the wages of parents immediately to the left of the discontinuity should predict higher credits than the wages of those to the right. In practice, the difference in predicted credits at the discontinuity displayed in Column (5) of Table 1 and graphically in Panel A of Figure 4 are negligible and statistically indistinguishable from zero.

Panel B of Figure 4 shows that there is nonetheless a sharp difference across the end-of-year threshold in incentives to self-report self employment income on one’s tax return. Taking the sample as a whole, 25 percent of individuals with December births—but none with January births—face a negative federal marginal tax rate (inclusive of SECA taxes) on a first dollar of self-reported self-employment income in excess of their W-2 earnings.³¹ We separately examine effects for individuals with W-2 wages above or below the first EITC kink point in Appendix Table A.1 and confirm that the change in reporting incentives occurs only for those with W-2 wages below the first EITC kink, among whom a December birth virtually always makes the MTR negative (and 30 percentage points lower on net).³²

³¹We calculate this for all individuals regardless of filing status based solely on tax-unit-level W-2 reported earnings and the date of birth of the child using TAXSIM.

³²While these reporting incentives exist in theory, the returns on reporting can only be actualized if an individual files a tax return and claims their child as an eligible dependent. We examine effects on claiming behavior in Columns 3-6 of Appendix Table A.1 and within narrower wage bins in Figure A.6. Notably, a first birth only increases their propensity to claim a child on their return by about 60 percentage points. This

While third-party-reported earnings do not vary across the end-of-year cutoff, we find clear differences in the propensity to self-report self employment earnings. Panel C of Figure 4 shows a clear break in the propensity to self-report self-employment income after December 31. In sharp contrast, there is no change in the propensity to have self-employment earnings reported by *firms* on a 1099 return around the same cutoff date in Panel D. Away from the discontinuity, we find that the relationship between date of birth and either outcome is completely flat, further suggesting that the result is not driven by local sorting around the cutoff date.

Table 2 displays point estimates of the RDD reporting effects. We find that, around the cutoff, having a qualifying child for EITC determination increases the probability of reporting self-employment income by 1.34 percentage points over a baseline rate of 7.19 percentage points.³³ In Appendix Figure A.7, we show that this effect size is not sensitive to the size of the “donut-hole” used, is highly stable across a wide range of bandwidths, and is robust to including a quadratic polynomial in the running variable with slope breaks around the cutoff.³⁴ Importantly, we find that while some individuals who begin reporting self-employment due to the incentive “bunch” their total reported earnings within \$500 of the first EITC kink (corresponding to the largest possible credit), this sort of “sharp-bunching” behavior accounts for less than one-fifth of the baseline SE reporting effect. Further, though there is no impact on 1099-reported contract payments, we nonetheless find substantial effects on the propensity of individual with 1099-reported payments to in turn self-report positive Schedule SE on their tax return, suggesting that even those with third-party reporting payments exert considerable discretion in how they self-report their net proceeds.

The estimates in Panels B and C of Table 2 confirm that these reporting responses are concentrated among individuals with a strict incentive to do so.³⁵ As predicted, we find

is partly because individuals are not able to claim children as dependents if they are claimed by another filer (e.g. another parent on a different return or a grandparent), and partly because some individuals already claimed children as dependents prior to the first birth identified in the SSA data. There is also a small effect on 1040 filing for individuals with wages slightly above the first EITC kink point, despite there being no *self-employment* reporting incentives or effects for those individuals. This is expected—these individuals qualify for a refundable EITC benefit given their wages so long as their children were born during the tax year, but must file a 1040 return in order to claim that benefit.

³³The magnitude of the reporting effect in Column (1) is remarkably similar to the post-2010 observed increase in self-employment around first births in our event-study analysis in Appendix B, suggesting that those changes likely reflect pure reporting behavior given fixed labor supply.

³⁴However, over-fitting the quadratic polynomials on a small number of days results in unstable point estimates in specifications with smaller bandwidths and wide donut holes.

³⁵In Columns (4) and (5), we find that individuals with wages below the first EITC kink report \$509 more in self-employment income (\$647 more including their spouse) if their child is born in December. How much do these individuals gain from reporting those self-employment proceeds? To quantify the tax benefit from reporting self-employment income, we calculate each individual’s tax burden first using their own and their spouse’s wages and self-employment earnings, and then based on wages alone. We then take the difference

no effect of self-employment reporting for individuals with wages above the EITC phase-in range. By contrast, we find that the presence of incentives increases the share reporting self-employment earnings by 4.6 percentage points (over a base rate of 10.8 percentage points) among individuals with W-2 wages below the first EITC kink (including individuals with no W-2 wages). In Figure 5, we estimate effects separately for individuals within \$2000 bins of year- t tax unit W-2 wages measured in constant 2015 dollars. The results in the figure show that the reporting effect diminishes as W-2 wages approach the first EITC kink and disappears completely above the kink, at which point individuals face strictly positive marginal tax rates on any reported self-employment earnings.³⁶

We noted above that the RDD effects might under-estimate the value of $\theta\lambda$ that is representative of the broader population if individuals do not fully internalize broader awareness of credits and reporting incentives in the immediate aftermath of a first birth. It is possible that continued exposure to those same incentives over time might lead to larger shifts in reporting behavior as individuals become aware of the incentives and how to optimally respond to them. One piece of suggestive evidence comes from examining differences in behavior in the *subsequent* tax year $t + 1$, when children born on either side of the threshold *both* count towards EITC and CTC determination in tax year $t + 1$. In year $t + 1$, individuals on both sides of the year t cutoff face *identical* incentives in the following year $t + 1$. However, if there is learning-by-doing—that is, if a prior year of eligibility better prepares one to maximize their refund in the future—then accumulated experience could lead individuals with December births in year t to still respond more than individuals with January births in $t + 1$.³⁷ Consistent with continued learning, we find statistically significant reporting effects in year $t + 1$ in Appendix Table A.2. The effect size is considerable, amounting to about one-quarter of the initial year- t reporting effect. However, unlike our baseline design, the analysis of outcomes in period $t + 1$ does not hold labor supply fixed and these effects might reflect either pure reporting behavior or real increases in self-employment. Yet we find no effect on *third-party-reported* self-employment work in period $t + 1$ in Appendix Table A.2,

to measure the net reduction in one’s net tax burden (inclusive of refundable credits). For this exercise, we use TAXSIM to estimate the federal tax (including FICA/SECA taxes) owed on combined tax unit W-2 and Schedule SE earnings, or W-2 earning alone, given each individual’s marital status and number of claimed dependents; for computational reasons, we do not incorporate additional information on deductions or other types of income. For individuals without EITC-eligible children, this “gain” is negative, as individuals pay higher taxes as a result of reporting additional earned income. However, the results in Column (6) show that the realized gains from reporting self-employment increase for those with wages below the first EITC kink and EITC-eligible children. On average, these individuals receive \$448.5 in net subsidies for the self-employment reported on their 1040.

³⁶By contrast, Appendix Figure A.8 shows there are no such effects on 1099-reported payments.

³⁷This comparison identifies learning that increases with years of eligibility (conditional on number of years as a parent) similar to Ramnath and Tong (2017). This comparison does not capture general learning that occurs the longer one has spent as a parent, since the RDD holds this fixed.

which is consistent with—but not conclusive of—pure reporting behaviors. In any case, the effects in $t + 1$ indicate that new parents do not reach full awareness of the credit in the immediate aftermath of a first birth.

5 Reporting Effects Over Time

5.1 Year-Specific Estimates

Our baseline estimates show that, holding actual labor supply fixed, individuals are more likely to report self-employment income on their tax return when the tax code creates incentives to do so. To what extent might changes in such reporting behaviors contribute to increases over time in the overall rate of self-employment that is self-reported on tax returns? In our framework, the reporting rate of low-wage individuals with children ($BTK_i = 1$) at any point in time depends on the true self-employment rate (σ_t^1), tax credit awareness among eligible individuals (λ_t) that impacts strategic reporting, and the broader noncompliance rate (θ_t), each of which might evolve over time (as indicated by the inclusion of time subscripts). However, at any point in time, the RDD effect among individuals with $BTK_i = 1$ identifies the share of informed non-compilers $\theta_t \lambda_t$. If θ_t is constant over time—which is plausible in our study period, as argued below—any change in RDD coefficient for the $BTK_i = 1$ group would reflect changing credit awareness over time.

We therefore estimate the main RDD specification separately for birth cohorts in each year—that is, births in December of each year t and in January of the corresponding year $t+1$. Results for both individuals in the EITC phase-in region ($BTK_i = 1$) and all others ($BTK_i = 0$) are presented in Figure 6. We find that reporting effects have increased significantly over time for individuals with $BTK_i = 1$, increasing approximately *fivefold* between 2000, when the effect size is 1 percentage point, and 2015, when the effect size is six percentage points. Yet, for this same group, we find no effect on having 1099-reported non-employee compensation in all years. We find no reporting effects for individuals with $BTK_i = 0$ in any year. These results indicate that the propensity to report self-employment *conditional* on one’s true labor supply is far more responsive to incentives in the tax code in recent years than two decades ago, consistent with increasing awareness of refundable credits over time (rising λ_t).

Our framework assumes that the structure of the tax code remains constant over time. An alternative hypothesis is that changing reporting behaviors reflect increased generosity of tax credits, which amplify the incentives to report self-employment. To test this explanation, we examine how the impact on post-wage MTRs has evolved over time by estimating the

specifications in Column (5) of Table A.1 for each year 2000–2018. The results, plotted in A.9, show that the only major change in self-employment reporting incentives for families with one child occurred in 2009 when the Child Tax Credit was expanded, which lowered the federal MTR for low-wage families from about -25% to about -30%. This result holds even when incorporating state EITC expansions in our analysis in Panel B. While this expansion may have impacted reporting behavior, we observe that the vast majority of the increase in reporting effects apparent in Figure 6 occurs from 2000 to 2007, prior to the announcement of the 2009 reform. Accordingly, it appears that changes in policy cannot account for the observed increase in self-employment reporting effects observed over this period.

To further assess whether spreading knowledge of refundable credits might drive changes in reporting behavior over time, we can exploit the spatial dimension of knowledge diffusion. Specifically, credit awareness may differ not just over time, but also across regions. Chetty, Friedman, and Saez (2013) documented that while hyper-strategic “sharp-bunching” behavior—reporting exactly the amount of self-employment income that qualifies you for the maximum tax benefit—has become more common over time, these behaviors have spread gradually across regions. Although our sharp bunching behavior accounts for only a small fraction of the reporting responses we document (Table 2), the broader prevalence of hyper-strategic sharp bunching in a region provides a useful proxy for the state of local knowledge at the time of one’s first birth.³⁸ Following Chetty, Friedman, and Saez (2013), we measure the share of sharp bunchers in each three-digit ZIP code in each year and use this as a proxy for local knowledge $\lambda_{t,z}$ (where z subscripts are added to reflect variation across ZIP codes).³⁹

We test whether geographic diffusion of knowledge can explain the increase in reporting effects over time by estimating an interacted version of the main RDD specification in Equation 4 that allows the effect to vary with our proxy for local knowledge, the share of sharp

³⁸Chetty, Friedman, and Saez (2013) show that the more common sharp-bunching is in a locale, the more elastic is real labor supply in response to EITC incentives, consistent with bunching being a meaningful proxy of local knowledge.

³⁹We replicate the sharp-bunching measure from Chetty, Friedman, and Saez (2013) and extend it through 2018. This measure is defined as the share of 1040 filers with children in a 3-digit ZIP code and year with AGI under \$50,000 in adjusted 2010 dollars who report self-employment earnings and total earned income within \$1,000 of the first EITC kink for that year. The resulting aggregate series is presented in Appendix Figure A.11. ZIP codes for individuals in the RDD sample are taken from W-2s and 1099-MISCs in the current year, or most recent of the prior two years if missing. To avoid any mechanical dependence, we tabulate the sharp-bunching shares in each ZIP3-year cell only among individuals *without* new births in the December at the end of the tax year or in the following January.

bunchers in each ZIP3 in each year:

$$\begin{aligned}\Delta y_{it} = & \alpha + \beta BunchShare_{z(i),t} \times \mathbf{1}\{date_i \in December\} + \delta BunchShare_{z(i),t} \\ & + \gamma \mathbf{1}\{date_i \in December\} + f(date_i, BunchShare_{z(i),t}) \\ & + \zeta_{z(i),m(i)} + \phi_{z(i),t} + \epsilon_{it}\end{aligned}\tag{5}$$

Here, t denotes the RDD cohort (the tax year corresponding to the year-end December-January pair), $z(i)$ is the ZIP code of individual i , and $m(i)$ is the binary birth month indicator for whether the birth occurred in December or January $\{date_i \in December\}$. We interact all terms, including the function of the running variable, with the bunching share for each ZIP3 and tax year. Since nationwide average bunching rates increase systematically over time, we include ZIP-by-calendar-month and cohort-by-calendar month fixed effects in some specifications to isolate only *within-ZIP* variation in bunching rates. These fixed effects absorb the RDD main effect within each ZIP (and thus γ) but do *not* absorb the interaction effect (β), which is identified from *differential* growth in RDD effect over time across ZIPs with higher or lower bunching growth. We estimate this specification on all observations in our sample with non-missing ZIP codes, pooling all birth cohorts in the December and subsequent January of years 2000 through 2018 to capture the full evolution of knowledge since 2000.

The results presented in Table 3 show that changes in new parents' reporting responses over time are tightly linked to the evolution of the knowledge proxy in their specific locality. In periods when sharp-bunching is relatively more common among parents with older children in the same region relative to other regions, new parents are significantly more likely to report self-employment income in response to changing incentives around their first birth. The results in Column (1) indicate that reporting effects in regions with low knowledge are only a small fraction of the baseline estimates in Table 2, but reporting effects grow significantly with greater local knowledge of tax incentives. The coefficients on the interaction terms are nearly the same in magnitude when including ZIP-month and cohort-month fixed effects, which indicates that the interaction term estimates reflect differential increases in knowledge within regions rather than aggregate time trends. The results in Columns (3) and (4) show that as knowledge spreads, new parents exposed to incentives are more likely to bunch at the first EITC kink, but increases in that type of bunching behavior accounts for less than one-third of the overall increase in self-employment behavior. Reassuringly, there is no RDD interaction effect for individuals with wages above the first EITC kink in Panel C, or any interaction effects on the propensity to have of 1099-reported payments or to sort around the cutoff based on the EITC value of wages in Columns (5)-(8).

The magnitude of the interaction term estimates in Table 3 are large enough to explain most of the increase in RDD effects over time in Figure 6. The estimates in Columns (1) and (2) imply that an increase in knowledge corresponding to a 3 percentage point rise in sharp bunching—approximately the national increase observed in Appendix Figure A.11—should increase the RDD effect among individuals with $BTK_i = 1$ by 2.5–3 percentage points. Accordingly, local knowledge spread can account for the majority of the roughly 4 percentage point rise in the RDD estimates from 2000 to 2018 in Panel B of Figure 6. Importantly, this exercise only reflects *localized* knowledge transmission; any other knowledge spread that is not mediated by ZIP codes (e.g. broad-based dissemination via the internet or broader social networks) would not be captured by our analysis. Thus gradual transmission of knowledge through localized and non-localized channels taken together plausibly explains the observed increase in reporting behavior over time.

These results support the claim that the increase in RDD effects over time in Figure 6 reflects increasing awareness of refundable credits, where aggregate awareness is a population-weighted average of local awareness levels $\lambda_{t,z}$. However, in our framework, the results in Table 3 are consistent with *either* regional changes in local credit awareness $\lambda_{t,z}$ or changes in broader noncompliance $\theta_{t,z}$ —in principle, increased noncompliance rates might also be reflected in larger share bunching shares. Yet, in contrast to credit awareness, an increase in general compliance should be associated with *less* self-employment reporting by individuals outside the EITC phase-in range ($BTK_i = 0$). In practice, when we run panel regressions of the local self-reported self-employment rate among individuals with no children or with higher wages in each 3-digit ZIP code and year on the local sharp bunching rate (Appendix Table A.3), we find no evidence that bunching is associated with lower self-employment reporting rates among these groups.⁴⁰ Meanwhile, when we run similar regressions of the share of the *all* individuals in the workforce with $BTK_i = 0$, we find that a one-percentage point increase in bunching is associated with a 3.6 percentage point rise in self-reported self-employment, over three times the increase in the interacted RDD specification estimated on the first-birth sample. While the estimates in Appendix Table A.3 do not hold labor supply fixed, we find no such association with 1099-reported self-employment earnings, pointing again to the possibility that increase in first-birth RDD effects might understate the overall impact of increased awareness of refundable credits on reporting behavior.

We observe additional evidence that credit-motivated reporting behavior is changing

⁴⁰In this analysis, ZIP-3 self-employment rates are defined as a share of all individuals in the workforce (among whom ZIP codes are always observed), not as a share of all individuals with children, as we do not observe ZIP codes for all individuals with children. For comparison, 93% of the RDD sample with non-missing ZIP codes in Table 3 are in the workforce by our definition. For comparability to the estimates in Column (2) of Table 3, we include ZIP and year fixed effects and weight by population.

(rising λ) but broader noncompliance is not (constant θ) in tax audits. We draw on data from audits of a representative stratified random sample of 1040 filers conducted in tax years 2001 and 2006–2014 as part of the IRS’s National Research Program (NRP) Individual Income Tax Reporting Compliance Studies. Appendix Figure A.10 plots the frequency with which individuals who initially report self-employment income on Schedule SE are found in audits to actually have no self-employment income and, conversely, the frequency with which people initially reporting no Schedule SE earnings are found in audits to actually have such earnings. We plot these series separately for individuals with and without an incentive to report self-employment.⁴¹ We find, first, that audits consistently find that 2–2.5 percent fail to report self-employment earnings on their tax return, but that share has been constant over time, consistent with constant θ . Second, we find that it has become increasingly likely for audits to find individuals reporting self-employment earnings when they actually have none—but that this increase has occurred *only* among individuals with $BTK_i = 1$, consistent with rising λ .

6 Quantification and Counterfactuals

6.1 Accounting for Changes in Reporting Behavior

Our empirical analysis finds that individuals with incentives to self-report self-employment on their tax returns have become increasingly likely to do so over time. In this section, we attempt to quantify how much changing reporting behaviors rather than changing labor market behaviors might drive the observed rise in the share of workers reporting self-employment on their tax returns.

In our conceptual framework, changes in the propensity to report self-employment earnings independent of true self-employment behavior are driven by increasing awareness of refundable tax credits (λ). Accordingly we quantify the extent to which reported self-employment rates have increased purely due to changes reporting behavior by considering counterfactual in which credit awareness remained constant at 2000 levels. Importantly, this exercise cannot recover the *true* self-employment rate, which depends on baseline noncompliance rates (θ)—while we have argued θ is likely to have been constant over this period, we make no assumption as whether the share reporting self-employment in 2000 is higher or lower than ground truth.

We first examine the contribution of changing reporting behavior to observed changed

⁴¹We classify individuals on the basis of their firm-reported W-2 income and the number of eligible children determined by the audit.

that is implied by our RDD estimates. To do so, we use the annual RDD estimates in Figure 6 as estimates of $\theta\lambda$ in each year and use Proposition 4.2 to back out counterfactual reporting rates at 2000 knowledge levels. Specifically, if first-time parents immediately internalize local credit awareness so that the RDD effect for individuals with $BTK_i = 1$ in any year identifies $\beta_t^{RDD, BTK=1} = \theta_t\lambda_t$, and if $\theta_t \equiv \bar{\theta}$ is further taken to be constant over time, then the difference in estimates for two year two years t and t' identifies:

$$\beta_{t'}^{RDD, BTK=1} - \beta_t^{RDD, BTK=1} = \bar{\theta}(\lambda_{t'} - \lambda_t)$$

Therefore, one can calculate what the self-reported self-employment rate would have been in year t if knowledge had remained constant at 2000 levels as:

$$\begin{aligned}\rho_t^* &= (1 - \bar{\theta})\sigma_t^{tot} + \kappa_t\bar{\theta}\lambda_{2000} \\ &= \rho_t - \kappa_t(\bar{\theta}\lambda_t - \bar{\theta}\lambda_{2000}) \\ \implies \rho_t^* &= \rho_t - \kappa_t(\beta_t^{RDD, BTK=1} - \beta_{2000}^{RDD, BTK=1})\end{aligned}\tag{6}$$

In the implementation of this adjustment, one challenge is the self-employment reporting rate in 2 is calculated as the share of the workforce with any positive labor earnings, not as a share of population, while our RDD coefficients are estimated using *all* first-births and not just among individuals in the workforce. To convert back to a measure as a share of the workforce, we first measure the counterfactual *number* of individuals reporting self-employment in each year as $SE_t - (\beta_t^{RDD, BTK=1} - \beta_{2000}^{RDD, BTK=1}) \cdot POP_t^{BTK=1}$, where $POP_t^{BTK=1}$ is an estimate of size (rather than the share) of the population with $BTK_i = 1$ and SE_t is the observed count of individuals reporting self-employment on Schedule SE. We use population-level estimates of individuals with children under 19 in their household from the Social Security birth records linked to own plus spousal W-2 wage earnings to identify the population in the $BTK = 1$ group.⁴² We then proceed analogously to estimate the counterfactual size of the workforce implied by the RDD effects on self-employment; specifically, we use annual estimates of the effects of self-employment reporting behavior on being included in our workforce definition (which is only relevant for those with no other W-2 or 1099-reported earnings) and the same $POP_t^{BTK=1}$ estimates to make this calculation.⁴³

The result of this exercise is presented as “Scenario 1” in Figure 7a. This adjustment reduces the share of the workforce reporting self-employment by about half of one percentage

⁴²To reduce noise, we impose zero change in reporting responses for individuals with $BTK = 0$, which is consistent with the evidence in Figure 6.

⁴³Putting these pieces together, we calculate the counterfactual share of the workforce reporting self-

point in 2018, about one quarter of the total increase since 2000. Our RDD estimates on their own can therefore account for a significant portion of the observed growth in the prevalence of self-employment reporting on income tax returns over time. While roughly three-quarters of the increase remains unexplained after this adjustment, two important points are worth noting. First, in the adjusted “Scenario 1” series, the majority of the post-2000 increase in the share with self-employment income occurs in the early years between 2000 and 2005 (rising about 1 percentage point), during which period the share of workers with *third-party-reported* payments for self-employment work on 1099 returns also grew substantially (over one-half of a percentage point). If the increase in Schedule SE reporting through 2005 largely reflects a real increase in self-employment work, this counterfactual exercise should not eliminate that increase. Rather, counterfactual adjustment has the largest impact in the decade after 2005, during which period the share self-reporting self-employment earnings continued to rise even as the 1099 share remained constant. Second, the adjustment based on our RDD estimates is only accurate to the extent that first-time parents immediately internalize local credit awareness λ ; if internalizing that information happens gradually over time, then our the increase in RDD effects overtime might understate the true rise in $\theta\lambda$. Our results above provided suggestive evidence that the RDD effects estimated around first births may not capture all shifts in longer-term reporting behavior.

As an alternative exercise to capture how changes in awareness of incentives may have impacted reporting behavior more broadly (in ways not necessarily captured by behavior around first births), we next consider how the aggregate share reporting self-employment income would evolved if, within demographic cells and narrow wage bins, the share of individuals with $BTK_i = 1$ who report self-employment had increased by the same amount as for individuals with $BTK_i = 0$ individuals within the same group. Formally, we suppose that, within demographics by wage earnings groups g , the increase in *true* self-employment rates over time the same for individuals with $BTK_i = 0$ and with $BTK_i = 1$: $\sigma_{g,t}^{BTK=0} - \sigma_{g,2000}^{BTK=0} = \sigma_{g,t}^{BTK=1} - \sigma_{g,2000}^{BTK=1}$.⁴⁴ If that condition holds, and if general noncompliance types θ are

employment as:

$$\frac{SE_t^*}{WF_t^*}^{RDD} = \frac{SE_t - (\beta_t^{\Delta SE, BTK=1} - \beta_{2000}^{\Delta SE, BTK=1}) \cdot POP_t^{k, BTK}}{WF_t - (\beta_t^{\Delta SE Only, BTK=1} - \beta_{2000}^{\Delta SE Only, BTK=1}) \cdot POP_t^{k, BTK}} \quad (7)$$

Where $\beta_t^{\Delta SE Only, BTK=1}$ are RDD estimates where the outcome is an indicator for reporting self-employment earnings on Schedule SE *and* having no W-2 wages i.e. being in the workforce because of being “self-employed only”.

⁴⁴Within wage bins, the variation in BTK_i reflects whether or not one claims a child on a 1040 return who is EITC-eligible. We use children claimed on 1040 returns rather than children on SSA birth records for this exercise because many children are claimed by individuals who are not their birth parents (for instance, children living with grandparents). We do not require the self-employment rates for $BTK_i = 0$ and $BTK_i = 1$ to be the same in levels, only to grow equally over time within bins.

constant over time and independent of $BTK_i = 1$, then the the observed increase in the self-reported self-employment for people with $BTK_i = 0$ in group g can be used to infer the counterfactual growth that would have occurred for those in the same group g with $BTK_i = 1$ in the scenario where λ had remained constant over time.⁴⁵

After calculating counterfactuals within each group g , we then calculate the overall counterfactual self-employment reporting rate ρ_t^* . To do so, we take the evolution of the number of individuals in the workforce with $BTK_i = 0$ or $BTK_i = 1$ in each group g as exogenously given, with the exception of the number of individuals in the workforce in the zero wage bin. While individuals with wages are part of our workforce definition regardless of their self-employment reporting, those with no wages are generally not included in our workforce unless they report self-employment earnings; thus, our counterfactual adjustment to SE reporting growth must change the number of individuals in the workforce with zero wages. Following the logic of the counterfactual, we assume that, in the absence of any change in reporting behavior, the share with no wages among individuals with children in workforce would have grown after 2000 by the same amount as the share with no wages among individuals with *no* children in workforce within the same demographic cell; we then calculate the implied number of individuals with children who are the workforce despite no W2 earnings for each year and use this to calculate ρ_t^* .

The result of this exercise is presented as “Scenario 2” in Figure 7a. Under this counterfactual, the share of the workforce observed with self-employment earnings would have been nearly a full percentage point lower in 2014—cutting the increase in those years by roughly half. Notably, this exercise suggests that, had reporting rates among those with children evolved the same as reporting rates for those without children in the same demographic-by-wage groups, there would have been essentially *no* overall increase in self-reported self-

⁴⁵Specifically, if λ were constant at 2000 levels (λ_{2000}), then the counterfactual increase in rate of self-reported self employment between 2000 and year t for individuals with $BTK_i = 1$ in group g would have been:

$$\begin{aligned}\rho_{g,t}^{1*} - \rho_{g,2000}^1 &= \frac{(1 - \theta)\sigma_{g,t}^1 + \theta\lambda_{2000}}{-((1 - \theta)\sigma_{g,2000}^1 + \theta\lambda_{2000})} \\ &= \frac{(1 - \theta)\sigma_{g,t}^1 - (1 - \theta)\sigma_{g,2000}^1}{(1 - \theta)\sigma_{g,t}^0 - (1 - \theta)\sigma_{g,2000}^0} \\ &= \rho_{g,t}^0 - \rho_{g,2000}^0\end{aligned}$$

where the last line is the observed increase in rate of self-reported self employment between 2000 and year t for individuals with $BTK_i = 0$. This in turn implies:

$$\rho_{g,t}^{1*} = \rho_{g,2000}^1 + (\rho_{g,t}^0 - \rho_{g,2000}^0)$$

This condition says that the counterfactual share of individuals with $BTK_i = 1$ in group g reporting self-employment in year t is given by the actual baseline share in 2000 $\rho_{g,2000}^1$ and the counterfactual growth inferred from the $BTK_i = 0$ individuals in the same group.

employment rates between 2005 and 2014—the period when trends in self-reported self-employment most dramatically from trends in third-party-reported earnings. Put differently, if the assumptions underlying this exercise hold, then the vast majority of the increase in self-employment reported on tax returns is explained either by increases in 1099-reported work up to 2005 or by changes in reporting behavior since 2000. However, we note that the assumptions in Scenario 2—in particular, the assumption that, within demographic-by-wage cells, *true* self-employment rates evolved in parallel for people with and without children claimed on tax returns since 2000—are significantly stronger than the minimal assumptions underlying the counterfactuals based on our RDD estimates. Nonetheless, there is reason to think this assumption is plausible. In particular, if this assumption was violated, one would expect rates of 1099-reported self-employment work grow differential for individuals with and without children. In practice, however, the same “Scenario 2” adjustments lead to no change in the 1099 series in Figure 7b, implying that the underlying assumption holds for 1099-reported self-employment work.⁴⁶

7 Conclusions

Taken together, the results in the paper show that the increase in the share of taxpayers reporting self-employment earnings since 2000 cannot be attributed to the recent rise of platform-mediated “gig” work. Rather, our findings show that individuals have become more likely to report self-employment earnings on tax returns when the tax code incentivizes them to do so—even when underlying labor supply does not change—and that changes in reporting behavior are a quantitatively important driver of observed trends.

More generally, our results caution against trusting trends in administrative data over trends in survey data without careful consideration of the process by which the data are generated. Our work highlights that even while self-reported earnings are sensitive to reporting incentives, third-party-reported earnings can provide a reliable benchmark for statistical analysis. Accordingly, we believe that information reported on 1099 returns will continue to be a crucial source of information about trends in contract and gig work even as taxpayer reporting behavior continues to evolve.

Nonetheless, we do not think 1099s can be a complete substitute for the information about self-employment earnings on individual tax returns (Form 1040 Schedules C and SE, in particular). There are two primary limitations to 1099 reporting of self-employment

⁴⁶While the assumption underlying Scenario 2 holds for 1099-reported self-employment work, we acknowledge that we cannot rule out the possibility that specific types of self-employment work that are primarily consumer-facing (such 1099s are not issued) are more amenable to individuals with children, and that the prevalence of such (real) work grew faster since 2000 for those with children than for those without children.

work. First, 1099 reporting does not apply to all self-employment work, but rather only to payments made by *businesses* to compensate for self-employment work. Many self-employed individuals do not get paid by a firm, but are rather paid directly by individual customers who have no 1099 reporting requirement. Second, 1099 reporting only contains information about *gross* payments. In many cases, we care about net earnings after expenses (this is the earned income that is ultimately directly taxable), which are only reported on Schedules C and SE. Our view is that those interested in attempting to get a full picture of self-employment participation and earnings will ultimately need rely on Schedule C and Schedule SE returns to some extent. In that case, we think it is useful to try to highlight and quantify the influence of specific reporting incentives over time. It is therefore reassuring that our quantification exercises are able to account for most of the divergence between the third-party-reported and self-reported series.

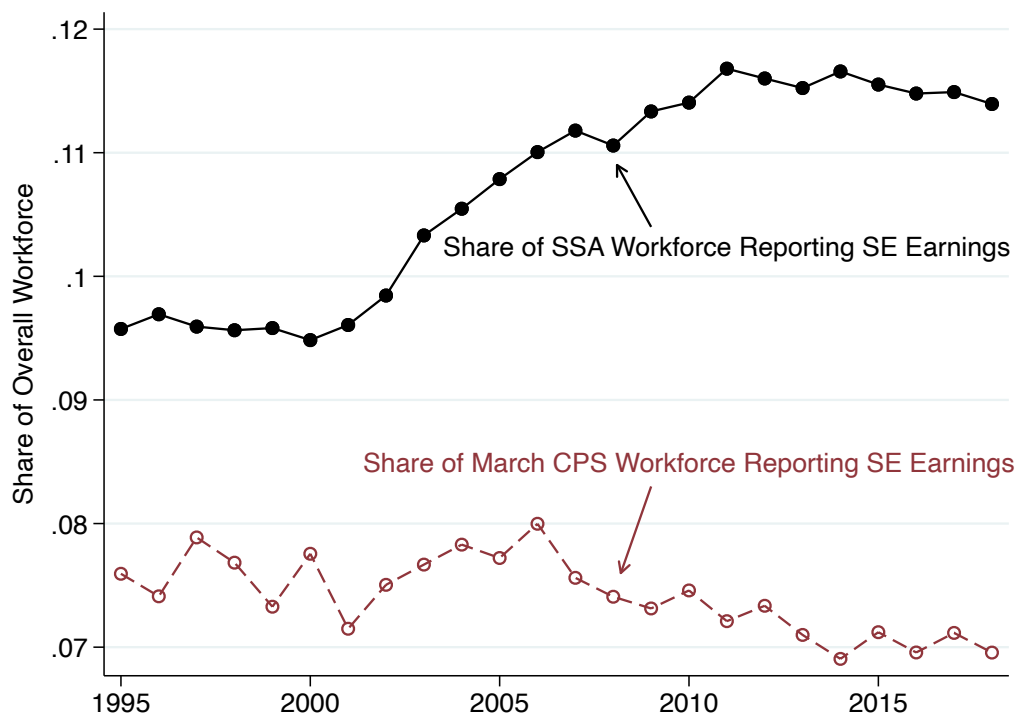
References

- ABRAHAM, K., J. HALTIWANGER, K. SANDUSKY, AND J. SPLETZER (2013): “Exploring Differences in Employment between Household and Establishment Data,” *Journal of Labor Economics*, 31(2), S129–S172.
- (2018): “Driving the Gig Economy,” .
- (2020): *Measuring the Gig Economy: Current Knowledge and Open Issues*, University of Chicago Press.
- ABRAHAM, K. G., J. C. HALTIWANGER, C. HOU, K. SANDUSKY, AND J. R. SPLETZER (Forthcoming): “Reconciling Survey and Administrative Measures of Self-Employment,” *Journal of Labor Economics*.
- ABRAHAM, K. G., B. J. HERSHBEIN, AND S. N. HOUSEMAN (2020): “Most Self-Employed Workers Are Independent Contractors,” *Policy Brief, W.E. Upjohn Institute for Employment Research*.
- ALLINGHAM, M. G., AND A. SANDMO (1972): “Income tax evasion: a theoretical analysis,” *Journal of Public Economics*, 1(3), 323–338.
- BLANK, R. M., K. K. CHARLES, AND J. M. SALLEE (2009): “A Cautionary Tale about the Use of Administrative Data: Evidence from Age of Marriage Laws,” *American Economic Journal: Applied Economics*, 1(2), 128–49.
- BRACHA, A., AND M. BURKE (2018): “How Big is the Gig?,” *Working Paper*.
- CHETTY, R., J. N. FRIEDMAN, AND E. SAEZ (2013): “Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings,” *American Economic Review*, 103(7), 2683–2721.
- COLLINS, B., A. GARIN, E. JACKSON, D. KOUSTAS, AND M. PAYNE (2019): “Is Gig Work Replacing Traditional Employment? Evidence from Two Decades of Tax Returns,” *SOI Working Paper*.
- DICKERT-CONLIN, S., AND A. CHANDRA (1999): “Taxes and the Timing of Births,” *Journal of Political Economy*, 107(1), 161–177.
- FARRELL, D., AND F. GREIG (2018): “The Online Platform Economy in 2018: Drivers, Workers, Sellers, and Lessors,” *JPMorgan Chase Institute*.
- FELDMAN, N. E., P. KATUŠČÁK, AND L. KAWANO (2016): “Taxpayer Confusion: Evidence from the Child Tax Credit,” *American Economic Review*, 106(3), 807–35.
- FLOOD, S., M. KING, R. RODGERS, S. RUGGLES, J. R. WARREN, D. BACKMAN, A. CHEN, G. COOPER, S. RICHARDS, M. SCHOUWEILER, AND M. WESTBERRY (2023): “IPUMS CPS: Version 11.0 [dataset],” Minneapolis, MN: IPUMS, (last accessed March 2021).
- GARIN, E. JACKSON, D. KOUSTAS, AND A. MILLER (2023): “The Evolution of Platform Gig Work, 2012-2021,” *NBER Working Paper #31273*.
- GARIN, A., E. JACKSON, AND D. KOUSTAS (2022): “Is Gig Work Changing the Labor Market? Key Lessons from Tax Data,” *National Tax Journal*, 75(4), 791–816.
- HARRIS, S. D., AND A. B. KRUEGER (2015): “A Proposal for Modernizing Labor Laws for Twenty-First-Century Work: The “Independent Worker,”” *The Hamilton Project Discussion Paper 2015-10*.

- INTERNAL REVENUE SERVICE (2022): “Individual Tax Files, 2000–2018,” United States Department of Treasury, (last accessed December 2022).
- JACKSON, E., A. LOONEY, AND S. RAMNATH (2017): “The Rise of Alternative Work Arrangements: Evidence and Implications for Tax Filing and Benefit Coverage,” *Office of Tax Analysis Working Paper 114*.
- KATZ, L., AND A. KRUEGER (2019a): “The Rise and Nature of Alternative Work Arrangements in the United States, 1995–2015,” *ILR Review*, 72(2), 382–416.
- KATZ, L. F., AND A. B. KRUEGER (2019b): “Understanding Trends in Alternative Work Arrangements in the United States,” *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 5(5), 132–146.
- KLEVEN, H. J., M. B. KNUDSEN, C. T. KREINER, S. PEDERSEN, AND E. SAEZ (2011): “Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark,” *Econometrica*, 79(3), 651–692.
- KOLEŠÁR, M., AND C. ROTHE (2018): “Inference in Regression Discontinuity Designs with a Discrete Running Variable,” *American Economic Review*, 108(8), 2277–2304.
- KUKA, E. (2014): “EITC and the Self-employed: Real or Reporting Effects?,” *Public Finance Review*, 42(6), 691–719.
- LALUMIA, S. (2009): “The Earned Income Tax Credit and Reported Self-Employment Income,” *National Tax Journal*, 62(2), 191–217.
- LALUMIA, S., J. M. SALLEE, AND N. TURNER (2014): “New Evidence on Taxes and the Timing of Birth,” *American Economic Journal: Economic Policy*, 7(2), 258–293.
- LIM, K., A. MILLER, M. RISCH, AND E. WILKING (2019): “Independent Contractors in the U.S.: New Trends from 15 years of Administrative Tax Data,” *SOI Working Paper*.
- MORTENSON, J. A., AND A. WHITTEN (2020): “Bunching to Maximize Tax Credits: Evidence from Kinks in the US Tax Schedule,” *American Economic Journal: Economic Policy*, 12(3), 402–32.
- NATIONAL ACADEMIES OF SCIENCES, E., AND MEDICINE (2020): *Measuring Alternative Work Arrangements for Research and Policy*. The National Academies Press, Washington, DC.
- RAMNATH, S. P., AND P. K. TONG (2017): “The Persistent Reduction in Poverty from Filing a Tax Return,” *American Economic Journal: Economic Policy*, 9(4), 367–94.
- SAEZ, E. (2010): “Do Taxpayers Bunch at Kink Points?,” *American Economic Journal: Economic Policy*, 2(3), 180–212.
- SLEMROD, J. (2016): “Caveats to the Research Use of Tax-Return Administrative Data,” *National Tax Journal*, 69(4), 1003–1020.
- SOCIAL SECURITY ADMINISTRATION (2022): “Annual Statistical Supplement to the Social Security Bulletin, 2022 (Table 4B),” SSA Publication No. 13-11700, (last accessed March 2024).

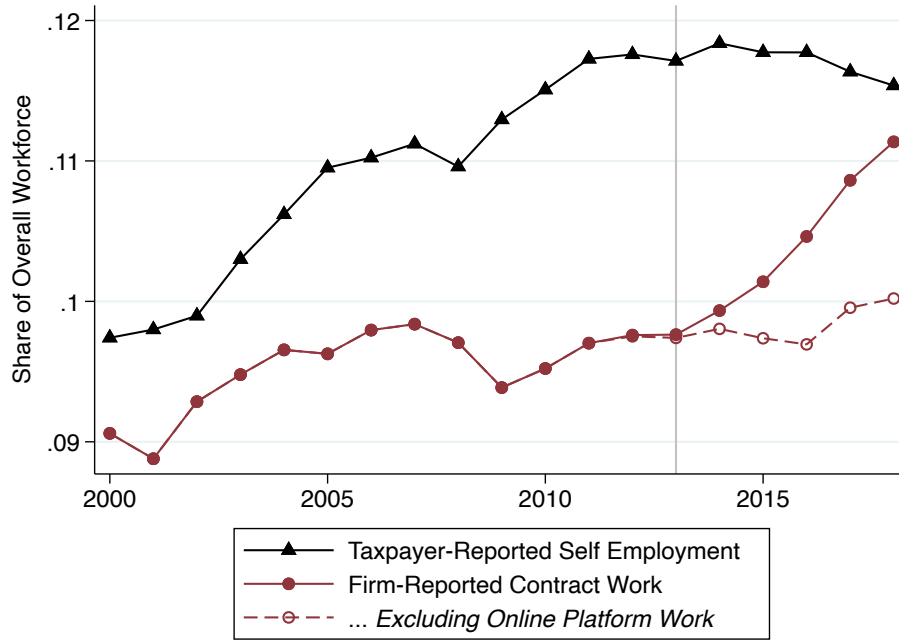
Figures

Figure 1: Share of Workforce with Self-Employment in Tax Returns and March CPS



Notes: Figure compares the share of the workforce—defined as all individuals with any employment earnings or self-employment earnings during the year—reporting positive self-employment earnings in tax records and in the March supplement to the Current Population Survey obtained from IPUMS (Flood, King, Rodgers, Ruggles, Warren, Backman, Chen, Cooper, Richards, Schouweiler, and Westberry, 2023). The black line reports this share in SSA data, where the workforce is defined as all individuals with positive earnings on either a W-2 return or on Schedule SE during the year, and individuals are classified as self-employed if they report *any* positive earnings on a Schedule SE. To extend the series before 1999, we draw on tabulations Social Security Administration (Social Security Administration 2022). The maroon line reports the share in the March CPS; for comparability to the tax data, the workforce is defined as all individuals reporting positive wage/salary earnings or self-employment earnings for the specified year (reported retrospectively in the following March), and individuals are classified as self-employed if they report *any* self-employment income for the year.

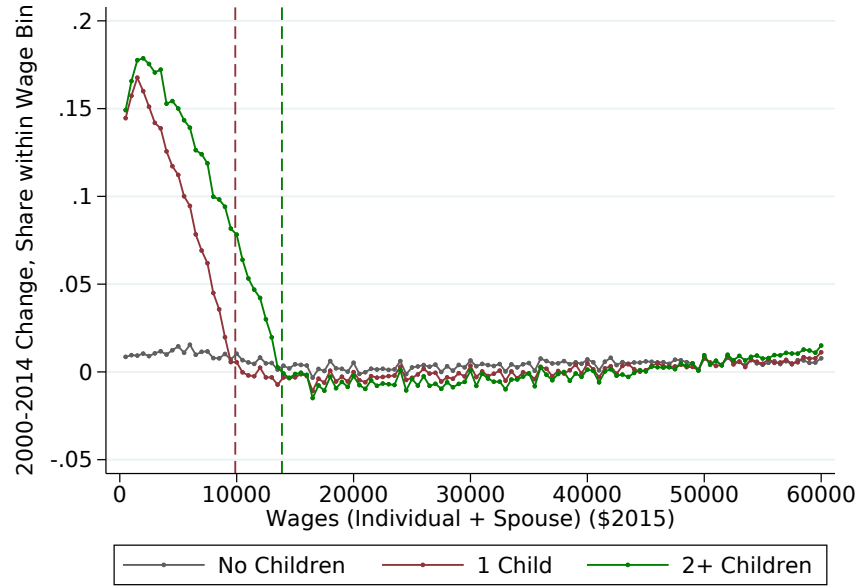
Figure 2: Self-Reported and Firm-Reported Self-Employment Earnings



Notes: Figure shows the share of individuals in the tax workforce reporting self-employment income on Form 1040 Schedule SE (comparable to Figure 1, but based on the IRS tax universe that also includes taxfilers without Social Security numbers) in each year (black line) and the share with firm-reported non-employee labor compensation exceeding \$600 on a 1099 Information Return (maroon line). In both series, the workforce is defined as all individuals with positive earnings on either a W-2 return or on Schedule SE during the year. Following the method in Collins, Garin, Jackson, Koustas, and Payne (2019), we separately break out the subset of independent contractors whose 1099-reported payments come exclusively from online platform economy firms. The 2017 and 2018 levels of total 1099-reported work reflect an imputation of how trends in online platform work would have evolved if reporting thresholds had remained constant described in Garin, Jackson, and Koustas (2022) and Garin, Jackson, Koustas, and Miller (2023). See Appendix D for additional details on data construction.

Figure 3: Growth in Reported Self-Employment 2000-2014, by Wage Income

(a) Any SE Filing

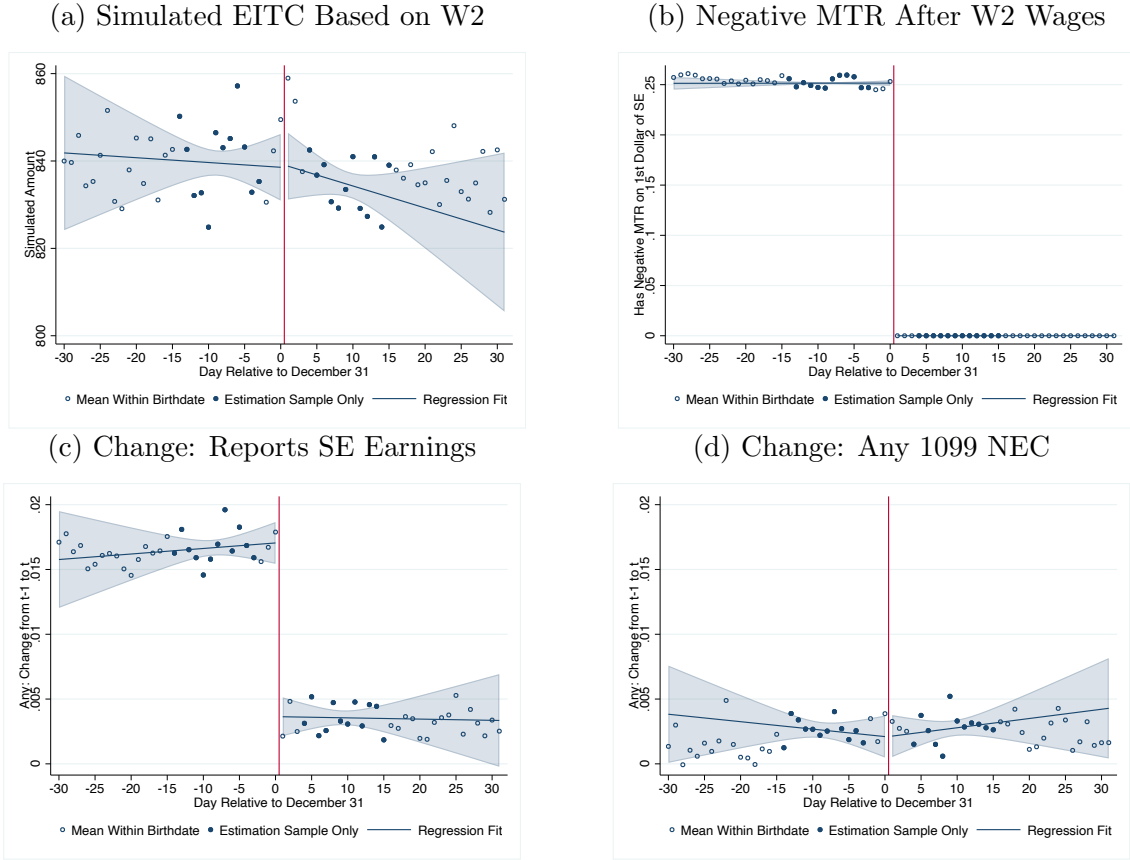


(b) Any 1099



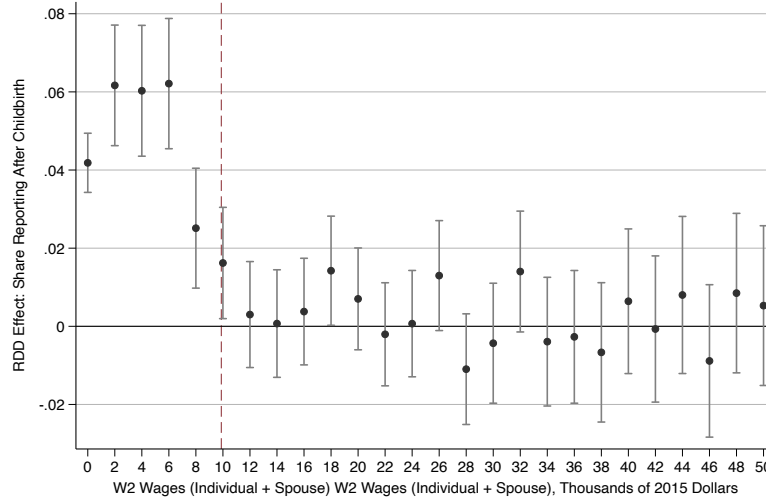
Notes: Figure shows the change in propensity to file SE (Panel A) or receive a 1099 Information Return (Panel B), for tax units with wage earnings. Wage earnings, in \$500 bins, are reported on the x-axis. Change is calculated between 2000 and 2014. Wages are determined based on W2 information returns. Number of children and spouse as reported on tax return. The area to the left of the vertical lines report the earnings where reporting an additional dollar of self-employment income would face a negative marginal tax rate due to the EITC, for households with 1 (maroon line) and 2 (green line) children.

Figure 4: Regression Discontinuity Design



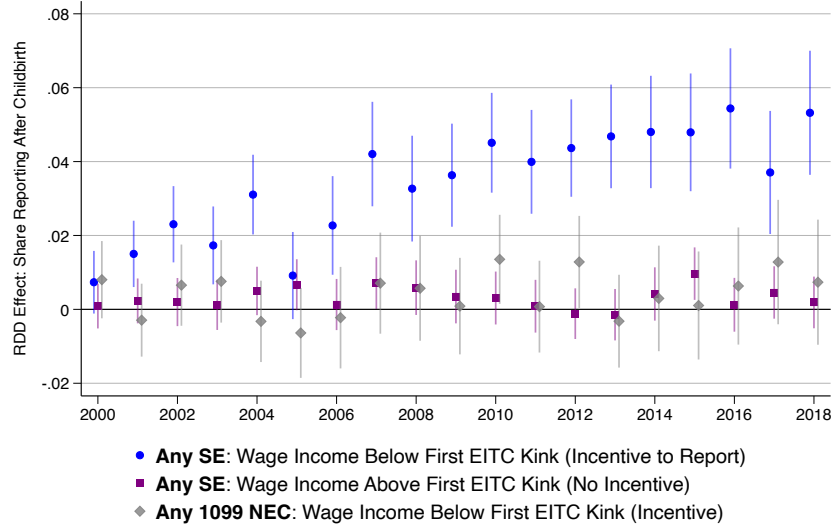
Notes: Figure graphically presents results from the baseline regression discontinuity design specification in Equation (4), pooling births in each December 2011-2018 and each subsequent January. Panel (a) examines the change in “simulated EITC” that is the credit each individual would earn if their birth occurred in December of tax year t based on their and their spouse’s (if married filing jointly) W-2 wages from year t . Panel (b) examines whether one would face a negative federal marginal tax rate (including SECA taxes) on a first dollar of self-employment earnings beyond one’s W-2 reported wage/salary earnings and those of any spouse in year t , given the year their child was *actually* born. Panel (c) examines the change in whether one reports any Schedule SE earnings in tax year t relative to the prior year $t - 1$. Panel (d) examines the change in having non-employee income reported on a 1099-MISC in tax year t relative to the prior year $t - 1$. Each dot is the average outcome for parents with births on the corresponding date, pooled across years. The solid dots are the calendar dates used in the main estimation window. The line segments are the regression fits allowing for the estimated discontinuity between December 31 and January 1, with 95 percent confidence bands from robust standard errors displayed.

Figure 5: RDD Estimates by Tax Unit W-2 Wage Earnings



Notes: Figure presents results from the regression discontinuity design specification in Equation (4) pooling births in each December 2011-2018 and each subsequent January, estimated separately for individuals within \$2000 bins of year- t tax unit (self plus spouse) W2 wages, measured in constant 2015 Dollars. The outcome is the change in whether one reports any Schedule SE earnings in tax year t relative to the prior year $t - 1$. The dashed maroon line is the amount where the first EITC kink occurs for families with one child based on the 2015 schedule.

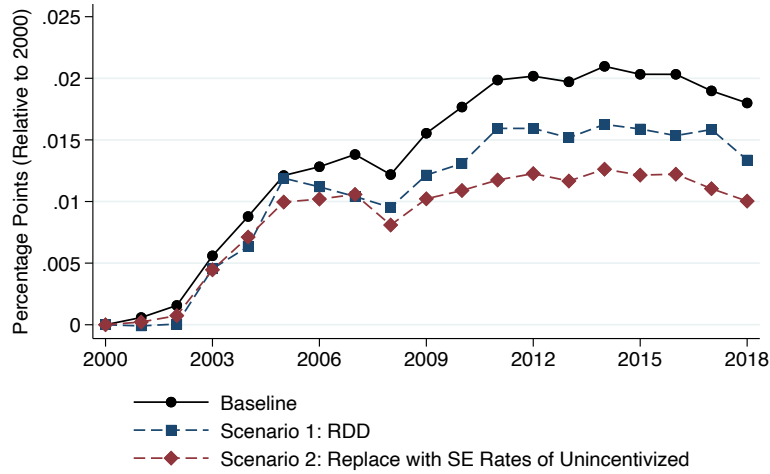
Figure 6: RDD Estimates by Year



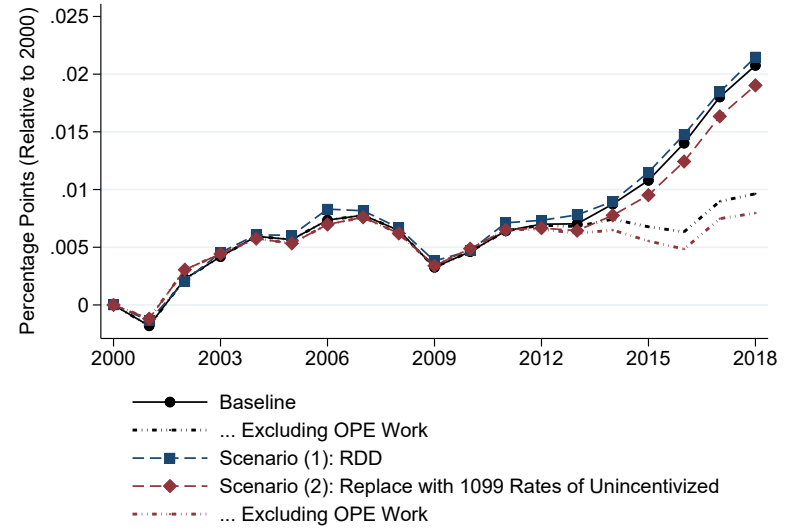
Notes: Figure reports our baseline RDD estimates from estimating Specification 4 in the text within individual cohorts. We report separate estimates for individuals with combined wages below the first EITC kink and those with wages above the first EITC kink point. Years correspond to the tax year t , at the end of which the births occur in the corresponding December or January.

Figure 7: Adjusted SE Shares Under Counterfactual Assumptions

(a) Adjusted Share of Workforce Filing Self-Employment



(b) Adjusted Share of Workforce with Non-employee Compensation



Notes: Figure presents the percentage point change in the share of the workforce with self employment income or 1099-reported contract payments relative to 2000, along with counterfactual increases in each series under alternative scenarios. Panel (a) applies these adjustments to self-employment, and Panel (b) to non-employee compensation. “Scenario (1)” adjusts self-employment downward using our annual RDD estimates reported in Figure 6, under the counterfactual assumption that the RDD effects remained constant in all years. “Scenario (2)” reports how overall rates would have evolved if trends for individuals who have incentives to report self employment followed those rates among comparable individuals without such incentives. Dashed lines in Panel (b) exclude the OPE.

Tables

Table 1: RDD Estimates, Placebos

<i>Panel A. Predetermined Outcomes</i>						
	Parent Is Female	Parent Age	Child Has Known Father	Tax Year t-1 Earnings		
	(1)	(2)	(3)	Has Any Schd SE	Sched SE Amount	W-2 Wages Amount
	(1)	(2)	(3)	(4)	(5)	(6)
<i>December Birth Effect</i>						
Coefficient	0.000729 (0.00237)	0.000960 (0.0297)	-0.000868 (0.00237)	-0.00101 (0.00120)	10.18 (17.93)	-38.30 (157.2)
N	1382740	1382740	1382740	1382740	1382740	1382740
<i>DV Mean Level, Jan Births</i>	0.507	28.48	0.495	0.0683	746.8	28945.9

<i>Panel B. Third-Party Reported Outcomes</i>						
	Δ Third-Party Reported Earnings			Tax Unit Wages, Tax Year t		
	Δ Any 1099 NEC	Δ Any Wages	Δ Wage Amount	Tax Unit W-2 Wages	Tax Unit Wages < 1 st Kink	Pred. 1-Child EITC W-2 Only
	(1)	(2)	(3)	(4)	(5)	(6)
<i>December Birth Effect</i>						
Coefficient	0.0000337 (0.00123)	0.00250 (0.00153)	1.113 (61.82)	46.07 (297.5)	-0.000889 (0.00206)	-0.761 (5.803)
N	1382740	1382740	1382740	1382740	1382740	1382740
<i>DV Mean Level, Jan Births</i>	0.0761	0.839	31433.7	55244.5	0.252	834.5

Notes: Table displays estimates from the baseline regression discontinuity design specification in Equation (4) on placebo outcomes, including both predetermined outcomes and third-party reported earnings. The sample is all individuals with births in the last fifteen days of December of each tax year t in 2011-2018 or the first fifteen days of January immediately following tax year t , omitting births within three days of the start of the new year. Placebo outcomes in Panel A are either earnings from the prior year (year $t - 1$) or fixed characteristics of the parent in the sample or their child. Placebo outcomes in panel B are changes in third-party-reported outcomes from year t relative to the prior year $t - 1$ or functions of tax-unit wages in year t as specified. Predicted 1-Child EITC Levels are calculated as the EITC amount one would receive if their first child had been born in year t (irrespective of when the true birth occurred) given only one's W-2 reported wage/salary earnings and those of any spouse reported on a 1040. For differenced outcomes in Panel B, we report mean year t levels of variable for individuals with first births in January of $t + 1$. Robust standard errors are displayed in parentheses.

Table 2: RDD Estimates: Main Effects on Self-Employment Reporting

	Δ Any SE (1)	Δ Any SE & Sharp Buncher (2)	Δ Has 1099 NEC & Reports SE (3)	Δ Individual SE Earnings (4)	Δ Tax Unit SE Earnings (5)	Tax Benefit from Reporting SE (6)
<i>Panel A. All Parents</i>						
Coeff	0.0134** (0.00118)	0.00254** (0.000326)	0.00388** (0.000950)	130.8** (11.57)	153.6** (19.53)	90.62** (9.372)
N	1382740	1382740	1382740	1382740	1382740	1382740
<i>DV Mean Level, Jan Births</i>	0.0719	0.00203	0.0406	855.4	1770.5	-417.7
<i>Panel B. With Wages < 1st EITC Kink</i>						
Coeff	0.0460** (0.00272)	0.00939** (0.00121)	0.0128** (0.00197)	509.0** (30.10)	646.5** (43.90)	448.5** (15.40)
N	349240	349240	349240	349240	349240	349240
<i>DV Mean Level, Jan Births</i>	0.108	0.00804	0.0503	1403.1	2319.4	-404.7
<i>Panel C. With Wages $\geq$ 1st EITC Kink</i>						
Coeff	0.00240 (0.00128)	0.000229 (0.000152)	0.000858 (0.00108)	3.108 (11.59)	-12.85 (21.44)	-30.45** (11.37)
N	1033500	1033500	1033500	1033500	1033500	1033500
<i>DV Mean Level, Jan Births</i>	0.0597	0	0.0374	670.6	1585.3	-422.1

Notes: Table displays estimates from the baseline regression discontinuity design specification in Equation (4) on third-party reported earnings. The sample is all individuals with births in the last fifteen days of December of each tax year t in 2011-2018 or the first fifteen days of January immediately following tax year t , omitting births within three days of the start of the new year. "Wages < 1st kink" subsample includes all individuals in tax units (self plus spouse if filing a 1040 jointly) with year t wages in the EITC phase-in region for households with one child in that year (irrespective of whether their birth actually occurred in December or January); the complementary subsample includes all other individuals. Outcomes are from year t or are changes from year t relative to the prior year $t - 1$, as specified. "Sharp bunchers" are individuals with earning income within \$500 of the level where the first EITC kink occurs. "Net Tax Benefit from Reporting SE" is the increase in net taxes one would pay if both they and their spouse (if present on a 1040) did *not* report their self-employment earnings to the IRS. We report mean year- t levels of each dependent variable for individuals with first births in January of $t + 1$ in each subsample. Robust standard errors are displayed in parentheses.

Table 3: RDD Estimates: Interactions with Knowledge Spread

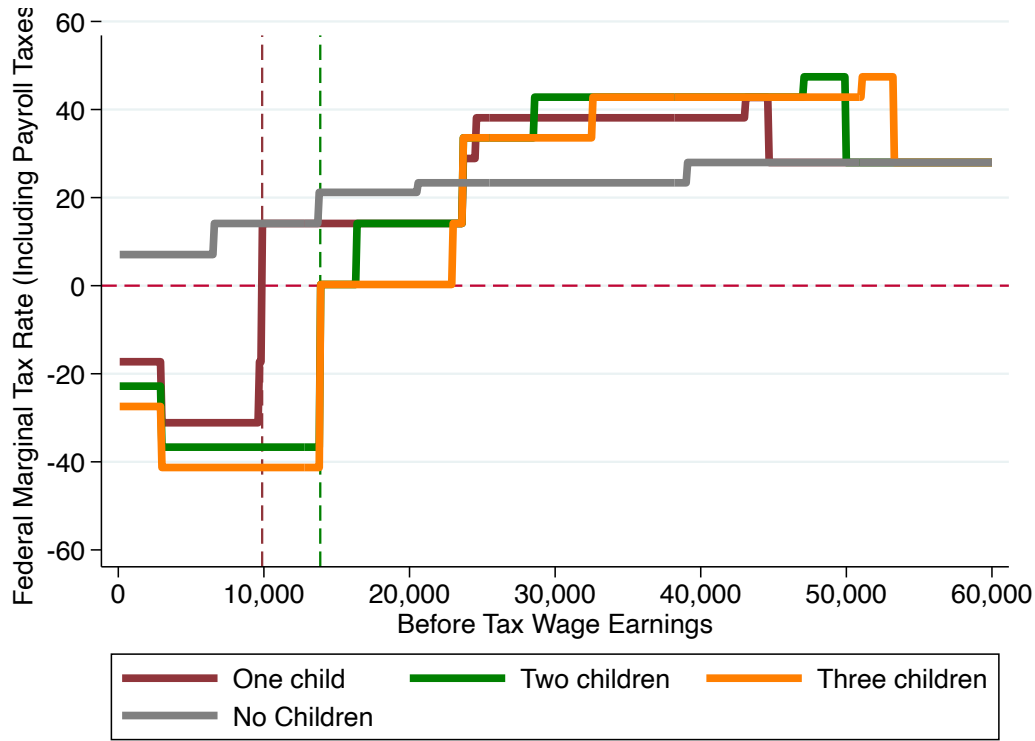
	Δ Any 1040-SE Income		Δ Any 1040-SE & Sharp Buncher		Δ Any 1099-NEC		Simulated W2-Only One Child EITC	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<i>Panel A. All Parents</i>								
December Birth	0.00250*	-	-0.000716*	-	0.000982	-	-0.162	-
	(0.00126)	-	(0.000323)	-	(0.00138)	-	(6.197)	-
December Birth × ZIP Bunching	0.337**	0.303**	0.0998**	0.0831**	0.0120	-0.0250	-8.682	44.61
	(0.0453)	(0.0597)	(0.0146)	(0.0183)	(0.0477)	(0.0633)	(198.0)	(266.4)
N	3094399	3094387	3094399	3094387	3094399	3094387	3094399	3094387
<i>Panel B. With Wages < 1st EITC Kink</i>								
December Birth	0.0134**	-	-0.000445	-	0.00335	-	8.664	-
	(0.00336)	-	(0.00130)	-	(0.00340)	-	(11.71)	-
December Birth × ZIP Bunching	0.999**	0.868**	0.311**	0.254**	0.0788	-0.0616	-221.5	344.4
	(0.123)	(0.143)	(0.0526)	(0.0665)	(0.112)	(0.144)	(339.5)	(467.4)
N	665093	665074	665093	665074	665093	665074	665093	665074
<i>Panel C. With Wages ≥ 1st EITC Kink</i>								
December Birth	0.00151	-	-0.000294	-	0.000666	-	-1.938	-
	(0.00132)	-	(0.000151)	-	(0.00149)	-	(7.200)	-
December Birth × ZIP Bunching	0.0557	0.0506	0.0156*	0.0135	-0.0199	-0.0191	65.15	-108.6
	(0.0457)	(0.0611)	(0.00654)	(0.00800)	(0.0514)	(0.0694)	(238.5)	(320.1)
N	2429306	2429295	2429306	2429295	2429306	2429295	2429306	2429295
Zip × Month FE		✓		✓		✓		✓
Cohort × Month FE		✓		✓		✓		✓

Notes: Table displays estimates from the regression discontinuity specification in Equation (5) in the text including interactions with local sharp-bunching rates at the 3-Digit ZIP code level for each year. Bunching rates are calculated omitting individuals in the RDD sample following Chetty, Friedman, and Saez (2013). The sample is all individuals with births in the December of each tax year t in 2001-2018 or the January immediately following tax year t with a valid ZIP code reported on an information return or tax filing in year t . The bunching variable is interacted with the discontinuity and the running-variable slope terms. Columns 2, 4, 6, and 8 include year t (cohort) by birth calendar month and ZIP3 by birth calendar month fixed effects so that all bunching variation in the interaction term comes from within-ZIP changes over time; these FEs absorb the RDD main effect. Robust standard errors are displayed in parentheses. See notes to 2 for additional details.

Appendix (For Online Publication)

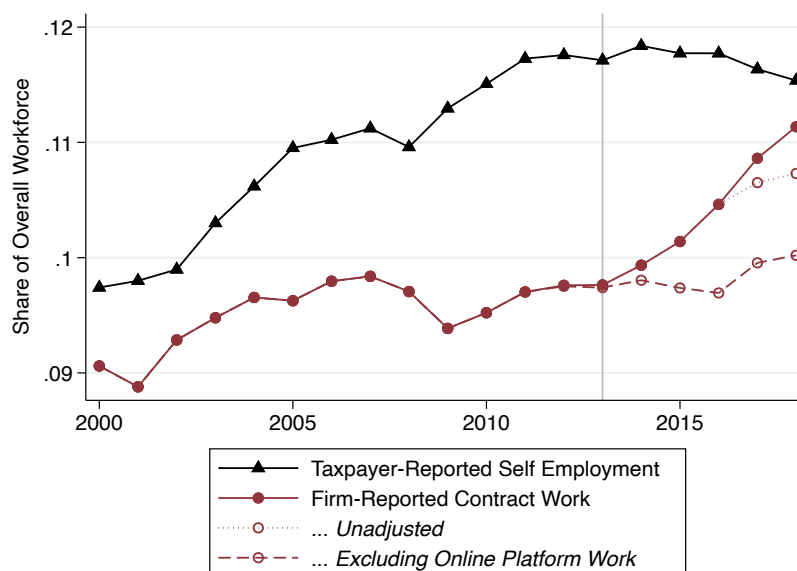
A Appendix Figures

Figure A.1: Effective Federal Marginal Tax Rate For Reporting Additional Dollar of Self-Employment Income



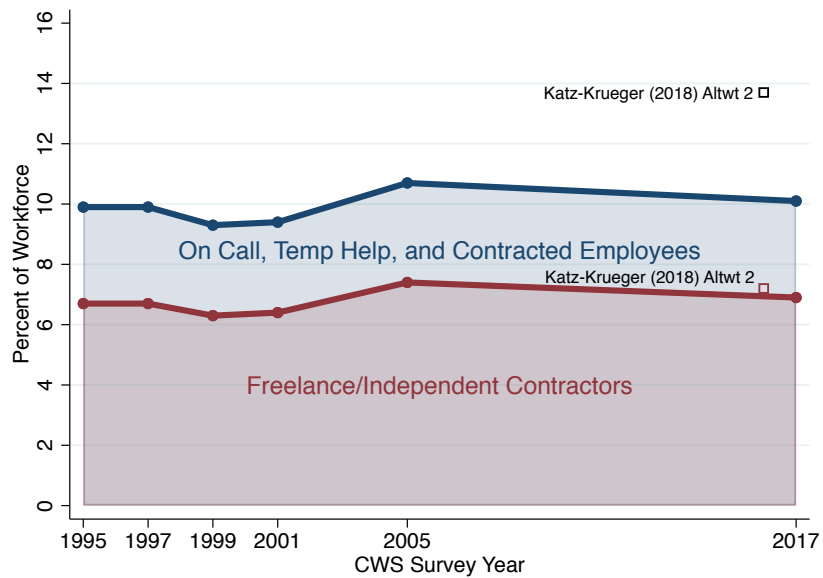
Notes: Figure shows the effective marginal tax rate for reporting an additional dollar of self-employment income, for a given level of before-tax wage earnings. Calculation takes into account the full tax schedule in tax year 2015 with no other credits/deductions except the EITC, CTC and standard deductions, and assumes Schedule SE payroll taxes are paid on the self-employment income and taxpayers deduct the employer-share of the payroll tax on self-employment income. Calculation assumes married filing jointly, however the marginal tax rates below the first kink point are identical for married and single parents who claim children. The area to the left of the vertical lines indicate the first kink-point of the EITC schedule, for households with 1 (maroon line) and 2 or more (green line) children.

Figure A.2: Firm-Reported Self-Employment Earnings, Raw Data and 1099-K Imputation



Note: After 2016, several online platform companies reported payments to gig workers on form 1099-K, adhering to the higher \$20,000 reporting threshold for that form. In the solid red line, the 2017 and 2018 levels of total 1099-reported work reflect an imputation of how trends in online platform work would have evolved if reporting thresholds had remained constant described in Garin, Jackson, and Koustas (2022) and Garin, Jackson, Koustas, and Miller (2023); this series reproduces the corresponding series in Figure 4 of Garin, Jackson, and Koustas (2022). This imputation is based on trends in state-level 1099-K data in Massachusetts and Vermont with a lower \$600 threshold in those states; see Garin, Jackson, Koustas, and Miller (2023) for details on the construction of the imputed series. While the imputed series approximates the number of workers engages in such work, it overstates the number of these workers for whom a 1099 return was actually filed. These raw data are shown in the “Unadjusted” series. See additional notes for Figure 2.

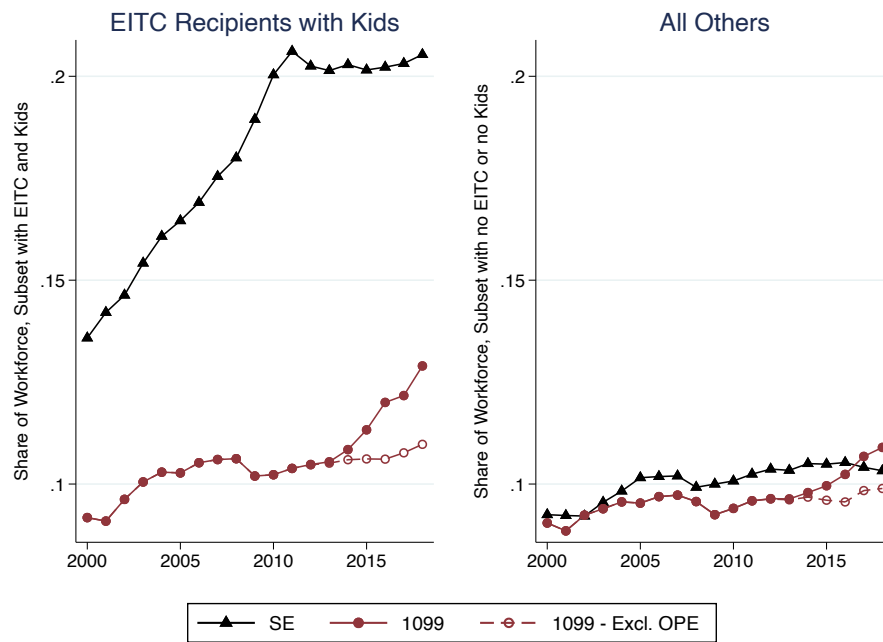
Figure A.3: Survey-Based Measures of Independent Contracting



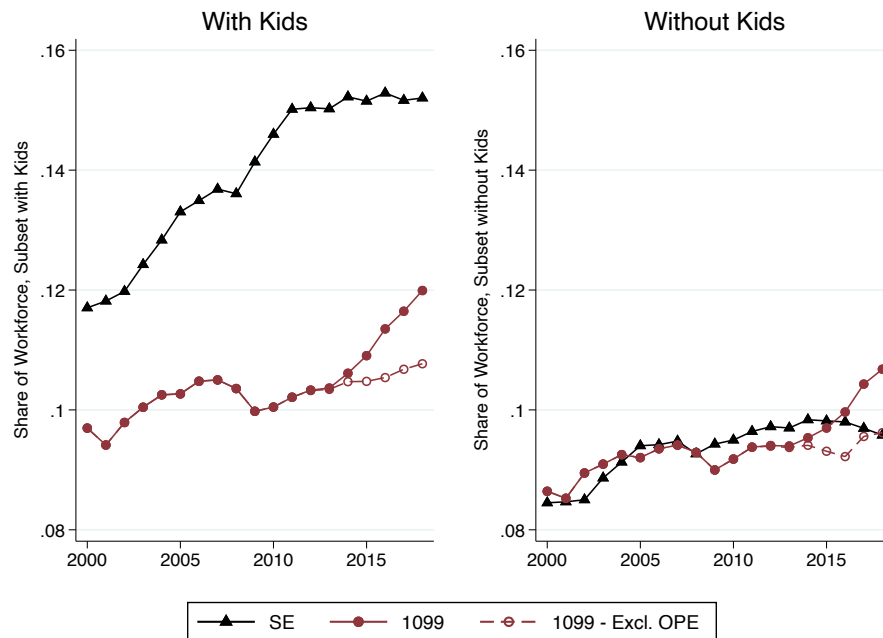
Source: Contingent Worker Supplement (1995, 1997, 1999, 2001, 2005, 2017), and Katz and Krueger (2019b). We report Katz and Krueger (2019b)’s “Alttwt. 2” estimates, an alternative weighting of the trends originally presented in Katz and Krueger (2019a) that downweight multiple job holders to match the rate of multi-job holding in the October 2015 CPS.

Figure A.4: Share of Workforce with Self-Employment and 1099 Information Returns

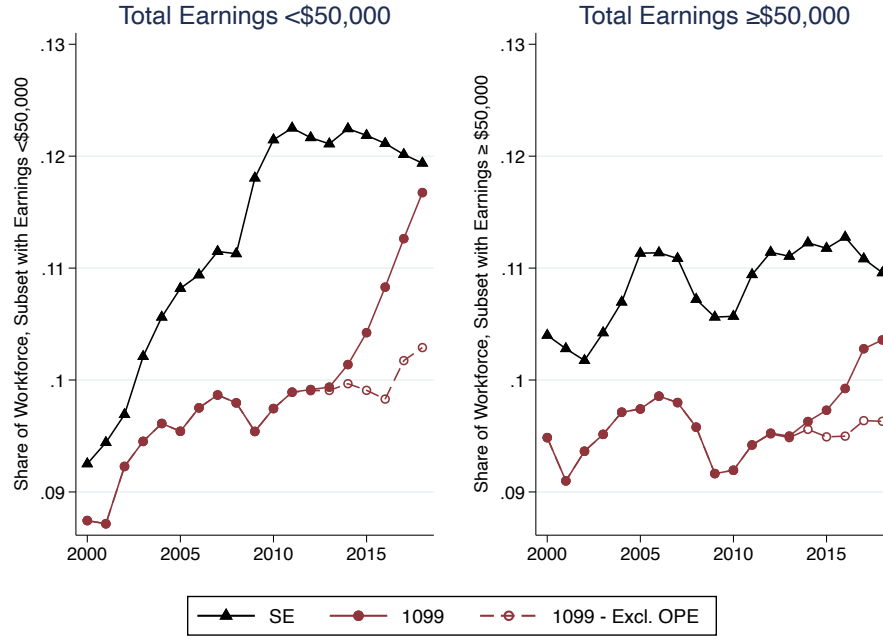
(a) By EITC and Kids



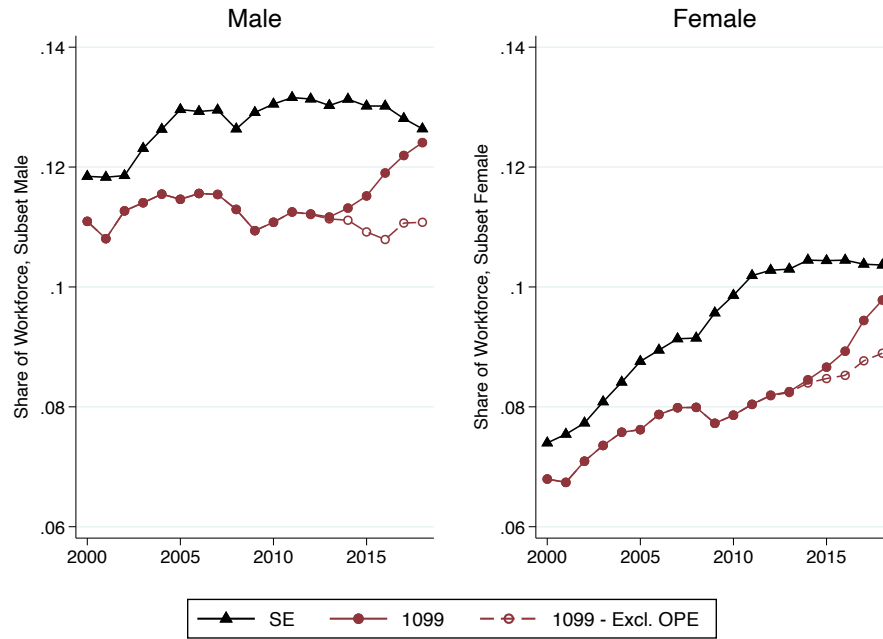
(b) By Presence of Children



(c) By Total Earnings

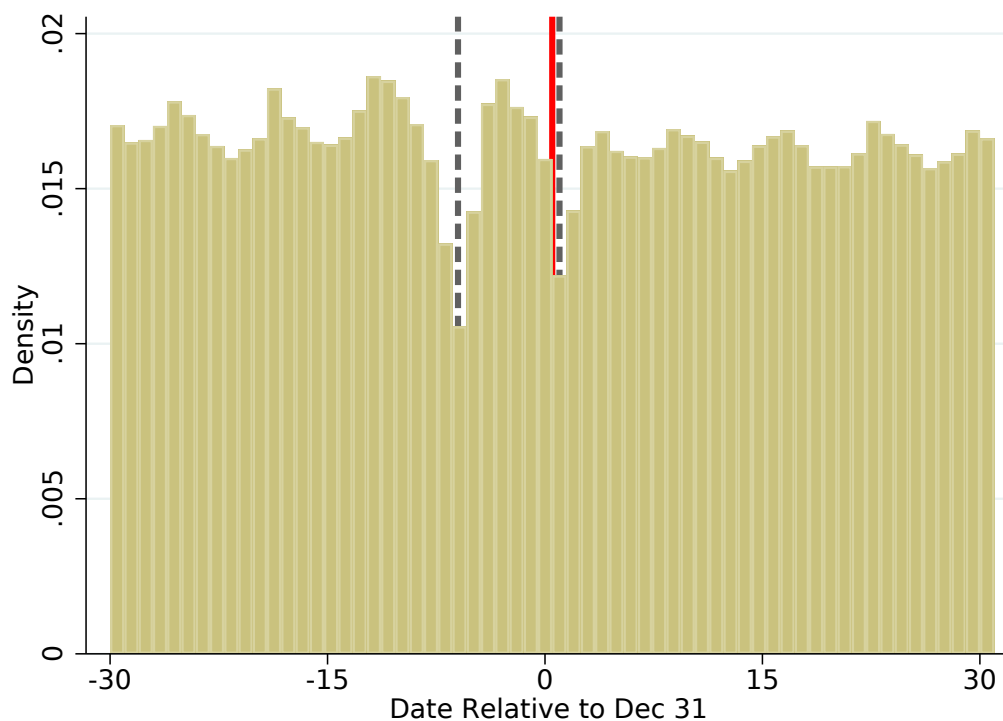


(d) By Gender



Notes: Figure shows the share of the overall tax workforce by tax year with any SE income as filed on Schedule SE (black line) and individuals who receive a 1099 Information Return (maroon line). After the entry of OPE, we additionally distinguish the receipt of 1099 Information Returns including and excluding those received from OPE firms (dashed maroon line). In Panel (a), the workforce definition is split on EITC recipients with kids claimed on their 1040. In panel (b), we split by presence of kids on their 1040, In panel (c), total earnings refers to the sum of wage and self-employment income by a primary tax filer and their spouse as reported on a 1040. In Panel (d), we split by gender.

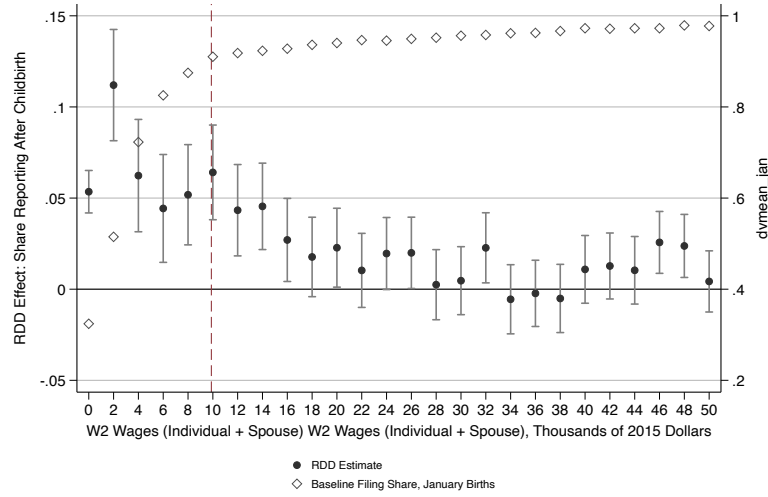
Figure A.5: Distribution of First Births Around End of Tax Years 2011-2018



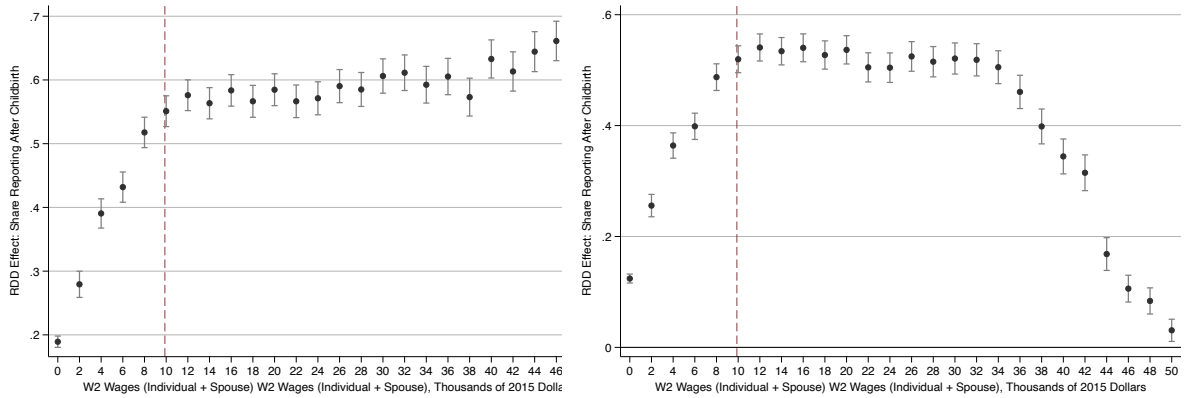
Notes: Histogram reports distribution of all first births in December of each tax year 2011-2018 or the following January in our SSA sample (corresponding to the sample in our baseline analysis). The solid red line denotes the end of tax year t and the dashed grey lines correspond to the Federal holidays on Christmas day (December 25) and New Year's day (January 1).

Figure A.6: RDD Filing Effects by Tax Unit W2 Wage Earnings

(a) Change: Files 1040

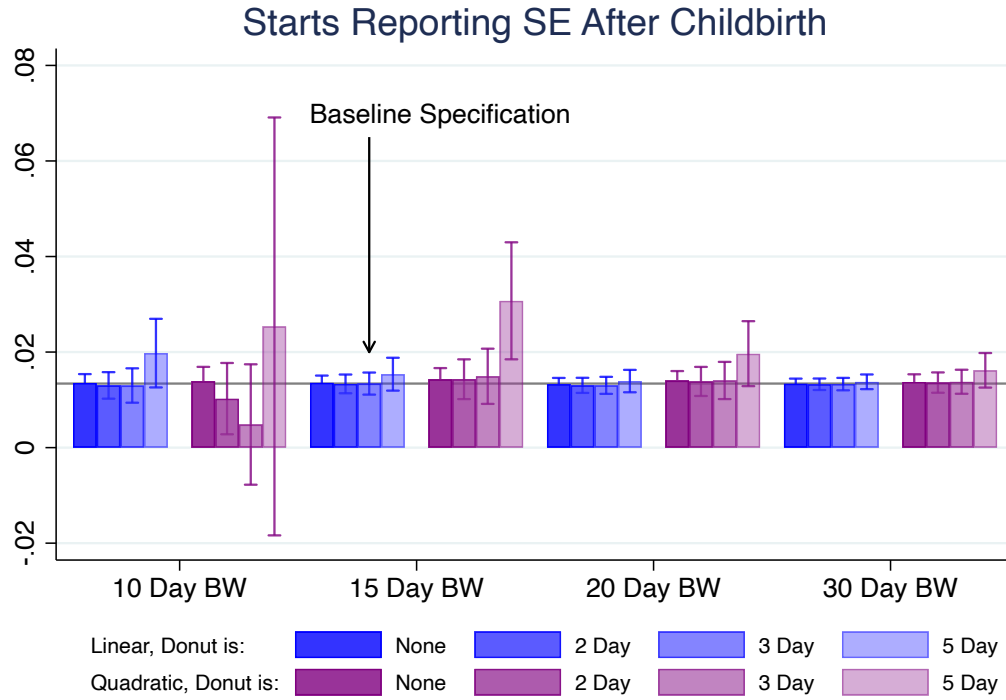


(b) Change: Any Children Claimed on 1040 EITC
(c) Change: Any Deps. Claimed on Sch.



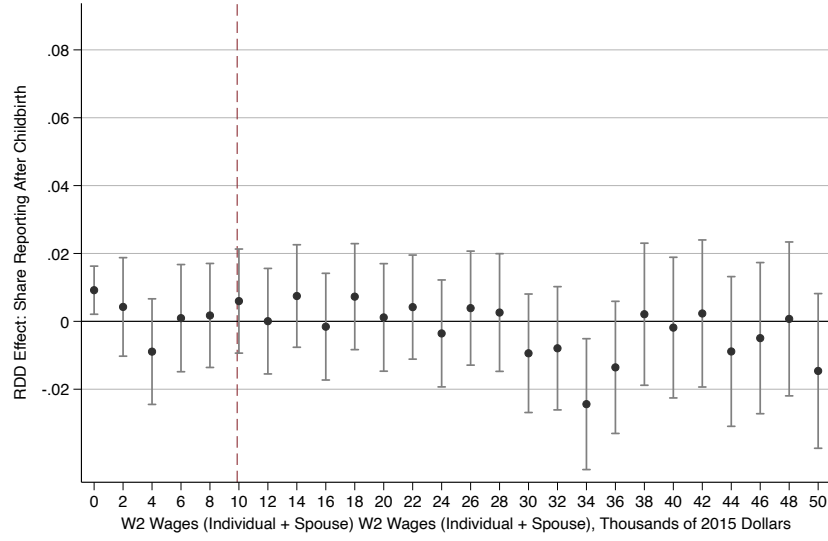
Notes: Figure presents results from the baseline regression discontinuity design specification in Equation (4) pooling births in each December 2011-2018 and each subsequent January, estimated separately for individuals within \$2000 bins of year- t tax unit (self plus spouse) W2 wages, measured in constant 2015 Dollars. The dashed maroon line is the earnings amount where the first EITC kink occurs for families with one child based on the 2015 schedule.

Figure A.7: Robustness of Main RDD Effects



Notes: Figure displays main regression discontinuity effects on the change in whether the one reports any Schedule SE earnings in tax year t relative to the prior year $t - 1$ from Column 1 in Table 2 under alternative specifications. Donut hole widths are bandwidths omitted from the regression sample. Quadratic specifications allow slopes to differ across the threshold. The horizontal black line corresponds to the size of the benchmark estimate in Table 2.

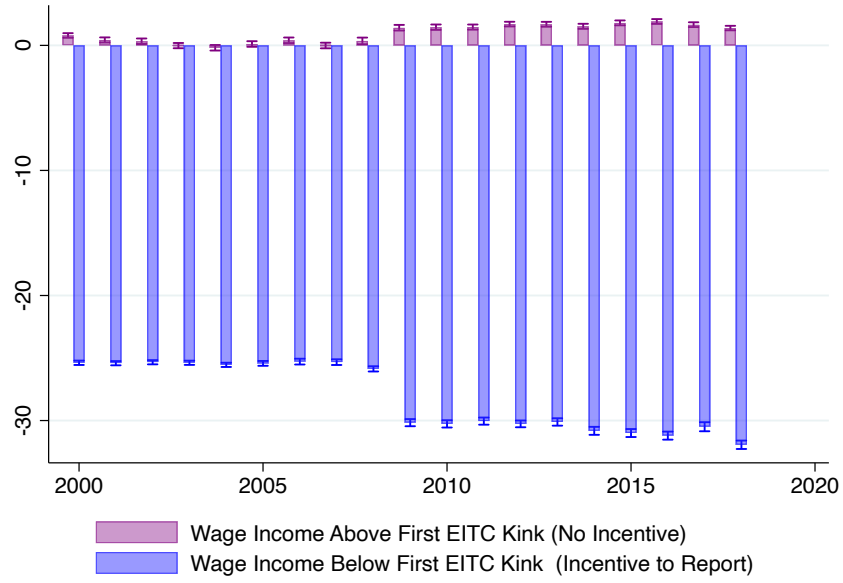
Figure A.8: RDD Effects on 1099 Earnings by Tax Unit W-2 Wages



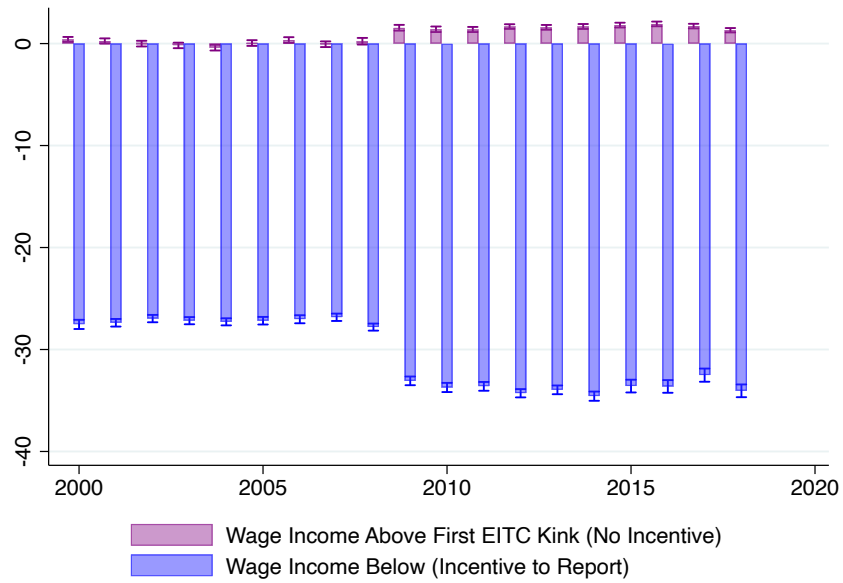
Notes: Figure replicates Figure 5 using the change in having non-employee income reported on a 1099-MISC in tax year t relative to the prior year $t - 1$ as the outcome. The dashed maroon line is the amount where the first EITC kink occurs for families with one child based on the 2015 schedule.

Figure A.9: RDD Effects on MTRs After Wages by Year

(a) Federal Taxes Only (Including SECA)

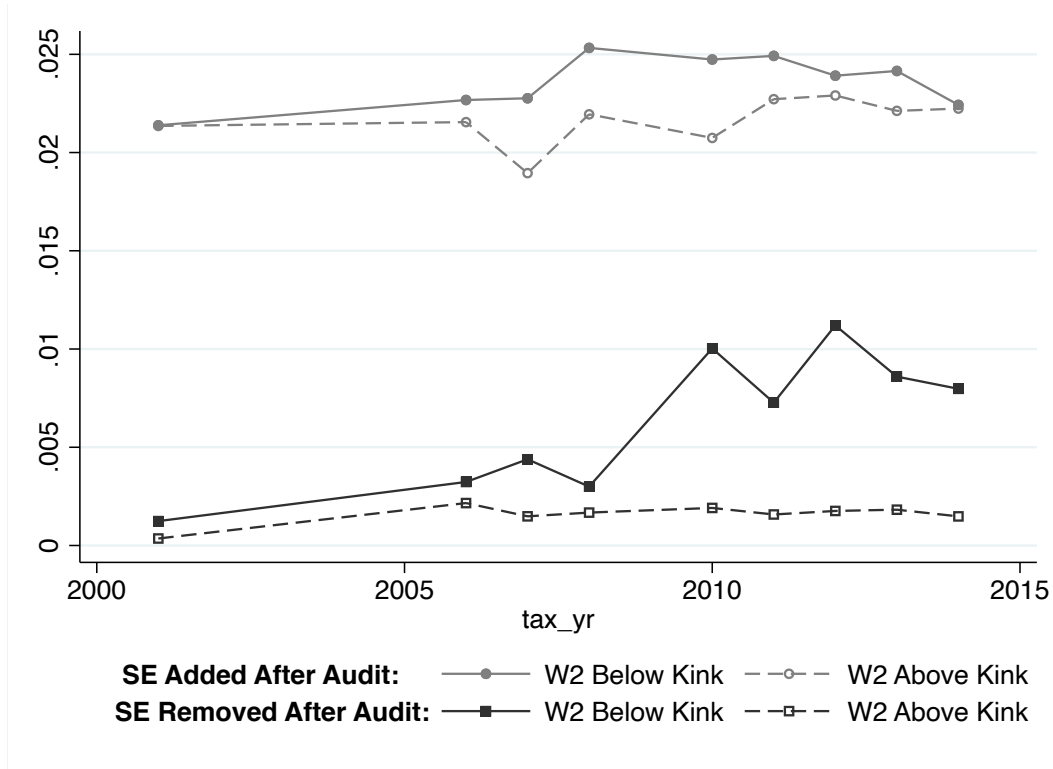


(b) Federal + State Taxes



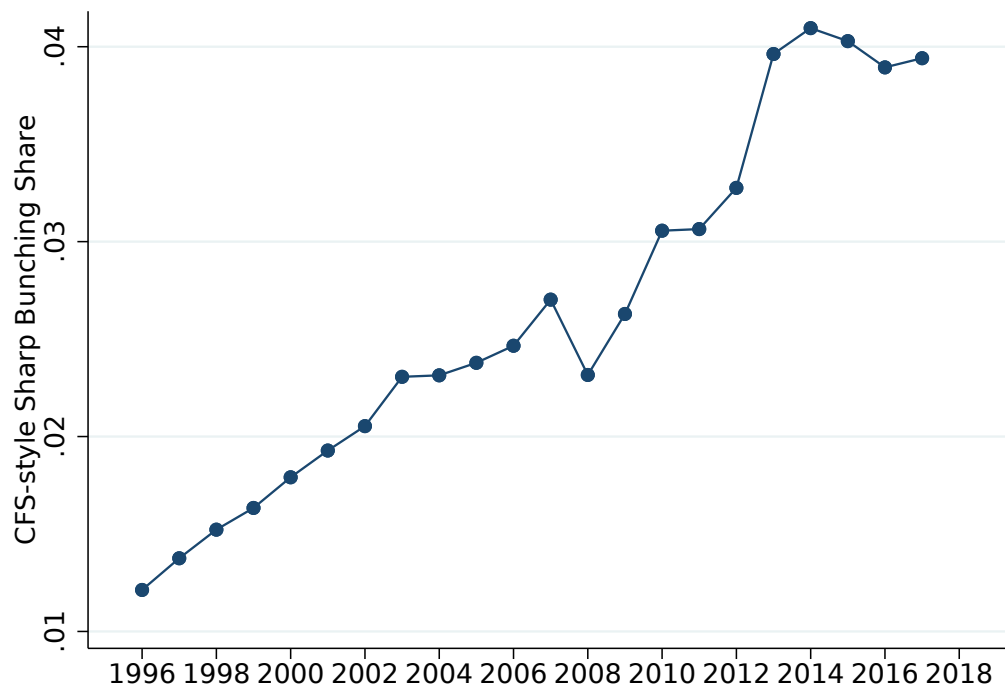
Notes: Figure reports our baseline RDD estimates from estimating Equation 4 in the text within individual cohorts. Years correspond to the tax year t , at the end of which the births occur in the corresponding December or January. Outcomes are marginal tax rates on a first dollar of self employment earnings, conditional on own and spouse's W2 wage earnings, calculated using TAXSIM.

Figure A.10: Changes in Self-Employment Status after NRP Audits



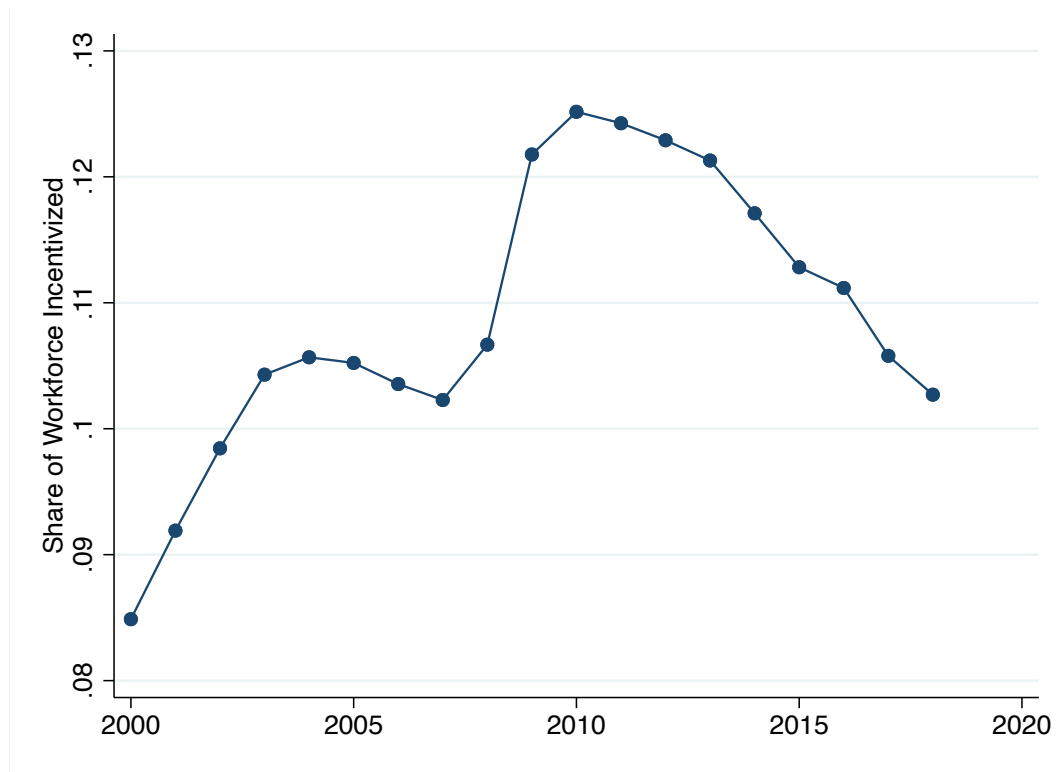
Notes: Figure displays results of audits of a representative stratified random sample of 1040 filers conducted in tax years 2001 and 2006–2014 as part of the IRS’s National Research Program (NRP) Individual Income Tax Reporting Compliance Studies. Using sampling weights for representativeness, the figure plots the share of *individuals* with 1040 returns who are found to have incorrectly not reported self-employment income on Schedule SE when they should have, and the share of individuals found to have reported positive self-employment income on Schedule SE when they actually should have reported none. Each propensity is calculated separately for individuals with and for individuals without an incentive to report self-employment. Individuals are classified based on their firm-reported W2 income and the number of eligible children determined by the audit.

Figure A.11: CFS-style Sharp Bunching Share Among Eligible Taxpayers with Children, 1996-2017



Notes: Figure plots the average of the share of tax payers who are sharp bunchers, following the methodology of Chetty, Friedman, and Saez (2013).

Figure A.12: Share of Workforce with Incentive to Report SE



Notes: The share incentivized in each year represents the number of individuals with children and wages below the corresponding EITC kink point as a share of the tax workforce.

Appendix Tables

Table A.1: RDD Estimates: Effects on Filing Status and Reporting Incentives

	MTR After Wages (1)	Has Neg MTR After Wages (2)	Δ Any 1040 (3)	Δ Any Children (4)	Δ Any EITC Dependents (5)
<i>Panel A. All Parents</i>					
Coeff	-6.534** (0.0826)	0.251** (0.00138)	0.0221** (0.00153)	0.598** (0.00180)	0.225** (0.00162)
N	1382740	1382740	1382740	1382740	1382740
<i>DV Mean Level, Jan Births</i>	22.72	0	0.860	0.131	0.0758
<i>Panel B. With Wages < 1st EITC Kink</i>					
Coeff	-30.69** (0.0565)	0.991** (0.000634)	0.0614** (0.00472)	0.299** (0.00366)	0.252** (0.00354)
N	349240	349240	349240	349240	349240
<i>DV Mean Level, Jan Births</i>	8.370	0	0.528	0.105	0.0912
<i>Panel C. With Wages $\geq$ 1st EITC Kink</i>					
Coeff	1.619** (0.0352)	0.000393** (0.0000662)	0.00875** (0.00129)	0.699** (0.00192)	0.215** (0.00181)
N	1033500	1033500	1033500	1033500	1033500
<i>DV Mean Level, Jan Births</i>	27.56	0	0.973	0.140	0.0707

Notes: Table displays estimates from the baseline regression discontinuity design specification in Equation (4) on third-party reported earnings. The sample is all individuals with births in the last fifteen days of December of each tax year t in 2011-2018 or the first fifteen days of January immediately following tax year t , omitting births within three days of the start of the new year. “Wages < 1st kink” subsample includes all individuals in tax units (self plus spouse if filing a 1040 jointly) with year t wages in the EITC phase-in region for households with one child in that year (irrespective of whether their birth actually occurred in December or January); the complementary subsample includes all other individuals. Outcomes are from year t or are changes from year t relative to the prior year $t - 1$, as specified. Marginal tax rates (MTRs) after wages are calculated as the federal marginal tax rate on the first dollar of self-employment earnings (including SECA taxes) beyond one’s W-2 reported wage/salary earnings and those of any spouse reported on a 1040, given the year their child was *actually* born. We report mean year t levels of each dependent variable for individuals with first births in January of $t + 1$ in each subsample. Robust standard errors are displayed in parentheses.

Table A.2: RDD Estimates: Lead and Lag Effects

	Any SE Earnings			Any 1099 Earnings		
	Tax Year t-1	Tax Year t	Tax Year t+1	Tax Year t-1	Tax Year t	Tax Year t+1
	(1)	(2)	(3)	(4)	(5)	(6)
<i>Panel A. All Parents</i>						
Coeff	-0.00101 (0.00120)	0.0124** (0.00127)	0.00395** (0.00135)	-0.000625 (0.00124)	-0.000592 (0.00126)	-0.000792 (0.00124)
N	1382740	1382740	1382740	1382740	1382740	1382740
<i>DV Mean Level, Jan Births</i>	0.0683	0.0719	0.0878	0.0734	0.0761	0.0738
<i>Panel B. With Wages < 1st EITC Kink</i>						
Coeff	-0.00139 (0.00270)	0.0446** (0.00315)	0.0146** (0.00335)	-0.00429 (0.00258)	0.000536 (0.00281)	-0.000281 (0.00272)
N	349240	349240	349240	349240	349240	349240
<i>DV Mean Level, Jan Births</i>	0.0900	0.108	0.143	0.0813	0.0973	0.0908
<i>Panel C. With Wages ≥ 1st EITC Kink</i>						
Coeff	-0.000849 (0.00131)	0.00155 (0.00131)	0.000433 (0.00140)	0.000626 (0.00141)	-0.000940 (0.00139)	-0.000937 (0.00139)
N	1033500	1033500	1033500	1033500	1033500	1033500
<i>DV Mean Level, Jan Births</i>	0.0610	0.0597	0.0692	0.0707	0.0689	0.0680

Notes: Table displays estimates from the baseline regression discontinuity design specification in Equation (4) on third-party reported earnings. The sample is all individuals with births in the last fifteen days of December of each tax year t in 2011-2018 or the first fifteen days of January immediately following tax year t , omitting births within three days of the start of the new year. “Wages < 1st kink” subsample includes all individuals in tax units (self plus spouse if filing a 1040 jointly) with year t wages in the EITC phase-in region for households with one child in that year (irrespective of whether their birth actually occurred in December or January); the complementary subsample includes all other individuals. Outcomes are from years $t - 1$, t , and $t + 1$, as specified. We report mean levels of each dependent variable in years $t - 1$, t , and $t + 1$ for individuals with first births in January of $t + 1$ in each subsample. Robust standard errors are displayed in parentheses.

Table A.3: Panel Relationship Between ZIP Bunching and SE Reporting

	Individuals With Children & Wages Below Kink	Individuals With Children & Wages Above Kink	Individuals Without Children
	(1)	(2)	(3)
Outcome: Workforce Share with SE			
ZIP Bunching Share	3.636** (0.290)	0.123** (0.0267)	0.113** (0.0325)
N	15709	15744	15782
Outcome: Workforce Share with 1099 NEC			
ZIP Bunching Share	-0.00210 (0.148)	0.0674** (0.0179)	0.0797** (0.0228)
N	15709	15744	15782
Zip FE	✓	✓	✓
Year FE	✓	✓	✓

Notes: Panels display estimates of panel regressions of self-employment rates and non-employee compensation reported on 1099-MISC within each specified workforce segments on the year-by-ZIP3 bunching measures calculated as in Chetty, Friedman, and Saez (2013). Sample is all individuals in the tax workforce 2000–2018, collapsed to the ZIP-year-subgroup level. Regressions are weighted by the workforce population in each cell. Standard errors are clustered by year and ZIP3.

Table A.4: Workforce Share Reporting Self-Employment on Schedule SE in Counterfactual Scenarios

	<u>Baseline</u>	<u>Scenario 1</u> RDD Adjusted	<u>Scenario 2</u> Incentivized = Unincentivized
2000	0.0974	0.0974	0.0974
2001	0.0980	0.0973	0.0976
2002	0.0990	0.0975	0.0982
2003	0.1030	0.1019	0.1019
2004	0.1062	0.1038	0.1045
2005	0.1095	0.1093	0.1074
2006	0.1102	0.1086	0.1076
2007	0.1112	0.1078	0.1080
2008	0.1096	0.1069	0.1055
2009	0.1129	0.1096	0.1076
2010	0.1151	0.1105	0.1083
2011	0.1173	0.1133	0.1092
2012	0.1176	0.1133	0.1097
2013	0.1171	0.1126	0.1091
2014	0.1184	0.1137	0.1100
2015	0.1177	0.1133	0.1096
2016	0.1177	0.1128	0.1096
2017	0.1164	0.1132	0.1084
2018	0.1154	0.1108	0.1074

Notes: Table reports baseline (unadjusted) share of workforce with self-employment earnings alongside counterfactual series adjusted for shifts in reporting behavior and demographic change “Scenario (1)” examines how self employment would have evolved in the absence of any reporting incentives captured in our RDD estimates; specifically, it reports the counterfactual replacing our RDD estimates reported in Figure 6 with zero in all years. The adjustment in “Scenario (2)” replaces SE rates for individuals with incentives to report SE with the rates among comparable individuals without this incentive in each year. See text for further details.

Table A.5: Workforce Share with 1099-Reported Non-Employee Compensation in Counterfactual Scenarios

	<u>Baseline</u>	<u>Scenario 1</u> RDD Adjusted	<u>Scenario 2</u> Incentivized = Unincentivized
2000	0.0906	0.0906	0.0906
2001	0.0888	0.0892	0.0894
2002	0.0929	0.0927	0.0937
2003	0.0948	0.0951	0.0950
2004	0.0965	0.0967	0.0964
2005	0.0963	0.0966	0.0959
2006	0.0980	0.0989	0.0976
2007	0.0984	0.0988	0.0982
2008	0.0971	0.0973	0.0968
2009	0.0939	0.0944	0.0940
2010	0.0952	0.0953	0.0955
2011	0.0970	0.0977	0.0971
2012	0.0976	0.0979	0.0973
	[0.0975]	[0.0978]	[0.0972]
2013	0.0976	0.0984	0.0970
	[0.0974]	[0.0982]	[0.0968]
2014	0.0993	0.0996	0.0984
	[0.0980]	[0.0983]	[0.0971]
2015	0.1014	0.1021	0.1001
	[0.0974]	[0.0981]	[0.0961]
2016	0.1046	0.1054	0.1030
	[0.0969]	[0.0977]	[0.0954]
2017	0.1086	0.1091	0.1069
	[0.0996]	[0.1000]	[0.0981]
2018	0.1114	0.1121	0.1096
	[0.1002]	[0.1009]	[0.0986]

Notes: Table reports baseline (unadjusted) share of workforce with 1099-non-reported non-employee compensation alongside counterfactual series adjusted for reporting incentives and demographics. Shares in square brackets exclude OPE work. “Scenario (1)” adjusts self-employment downward according using our annual RD estimates reported in Figure 5a. “Scenario (2)” replaces SE rates for individuals who have incentives to report SE with the rates among comparable individuals without this incentive.

B Self-Employment Reporting and Changing EITC Incentives: Event Study Around Childbirth

As discussed in the main text, only households with children face a negative marginal tax rate for reporting self-employment income. To further test the hypothesis that self-employment growth is tied to EITC incentives, we follow Chetty, Friedman, and Saez (2013) and examine how self-employment reporting changes around a person’s first childbirth, when they become eligible for a generous credit. We expand upon Chetty, Friedman, and Saez (2013) in two main ways. First, to investigate the extent behavior is changing *over time*, we examine the change across different time periods. Second, we separate the rise in self-employment around childbirth into 1099-reported self-employment and self-reported work. An increase in 1099-reported work may suggest changing worker needs around childbirth draw workers into self-employment for the first time. We begin with a simple exercise, examining the raw change in self-employment at childbirth, before formalizing our analysis in an event-study framework.

We start by examining the simple raw change in self-employment in the year of childbirth. We take childbirths for all parents reported in the SSA database whether or not the child is claimed as a dependent on tax filings by that parent. Figure B.1a reports the change in self-employment filing in the year of childbirth from the year before, for every cohort of first births from 1997-2018. The figure shows that the extent to which individuals begin reporting self-reported self-employment exactly when it becomes advantageous to do so has increased over this period by 0.9 percentage points, from a level of 0.9 percentage points in 1997 to 1.8 percentage points by 2014. 1099-reported work—which individuals have no discretion over reporting—differs in two key ways. First, on average, there is *no* increase in 1099-reported work in the year of childbirth. Second, there is no underlying trend in the rate of doing 1099-reported work in the year of childbirth. Appendix Figure B.2 further breaks down the trends by gender of the parent. We find that all of this increase comes from mothers: the change in self-employment in the year of childbirth among mothers has gone from 0.4 percentage points in 1997 to 2 percentage points by 2014. In contrast, 1099-reported work *decreases* in the year of childbirth for mothers; the decrease is actually slightly greater in magnitude today than in the past.

We next proceed to formalize this analysis and examine additional periods after childbirth using an event-study specification that will control for aging and business cycle effects. Our

event study specification is standard and given as follows:

$$y_{it} = \sum_{k \in K} \beta_k^p I\{\text{FirstChildbirth}_i = t + k\} + \gamma_{a(i) \times g(i)}^p + \gamma_{t \times g(i)}^p + e_{it}^p \quad (8)$$

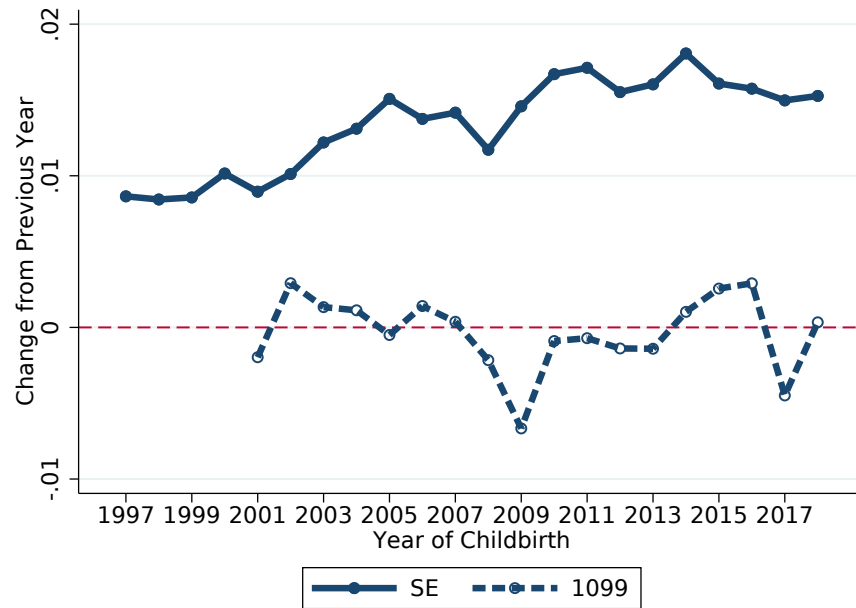
where i indexes parent, t indexes year, $a(i)$ gives the age of i . FirstChildbirth_i is i 's year of first birth. $g(i)$ is the parent's gender, thus allowing for time and age effects to differ by parental gender.⁴⁷ We examine two key outcomes: having any contract/freelance work, and being an S.E. taxpayer with no contract/freelance work. We run separate regressions for different 3-year rolling windows, $p \in \{2003 - 2005, 2004 - 2006, \dots, 2012 - 2014\}$. We exclude an indicator for the period one year prior to first birth, so that the event-time coefficients are all relative to period -1, and examine an event window of 4 years pre and post event ($k \in \{-4, \dots, 4\} \setminus -1$). Standard errors are clustered at the individual level.

Figure B.1b plots the full set of event study coefficients we estimate for two cohorts of births: 2003-2005 births and 2011-2013 births. As in the raw means, we find that 1099-reported self-employment is flat around childbirth for both cohorts. But self-reported self-employment is a different story—the propensity to self-report self-employment income increases sharply in the year of birth and by about 0.75 percentage points in subsequent years. Moreover, the magnitude of this time 0 response has grown over time: while self-reported self-employment rates grew by 0.5 percentage point after childbirth in 2003–2005, the corresponding increase was around 1.25 percentage points in 2011–2013. This contrasts with firm-reported contract work, which did not become more common after childbirth in either time period. Appendix Figure B.3 reports estimates separately by gender of the parent. As we found earlier, these changes over time are largely driven by mothers.

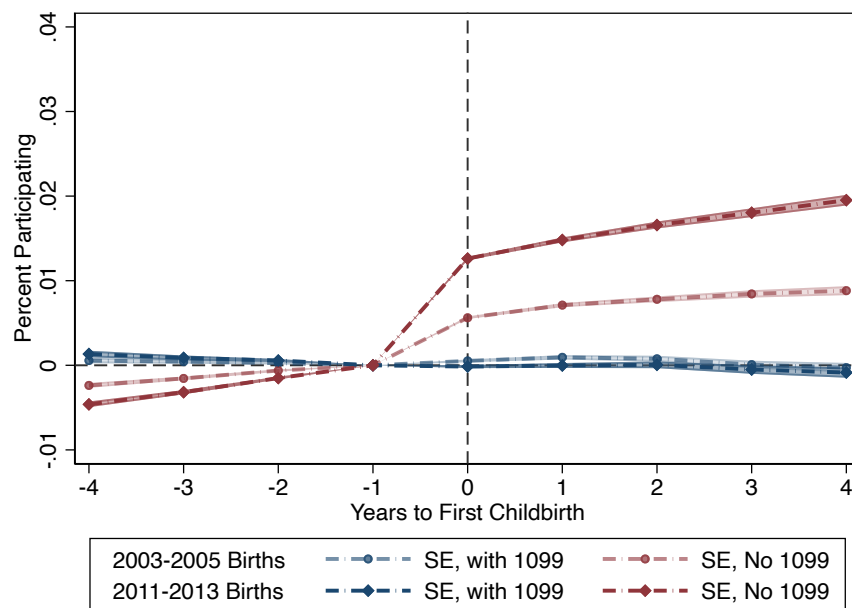
⁴⁷ Accordingly, the event-study coefficients are the average of coefficients run separately for men and women, which we report in Appendix Figure B.3.

Figure B.1: Change in Self-Employment Around First Childbirth, 1997-2018 Births

(a) Changes by year of First Birth: 1997-2018 Births



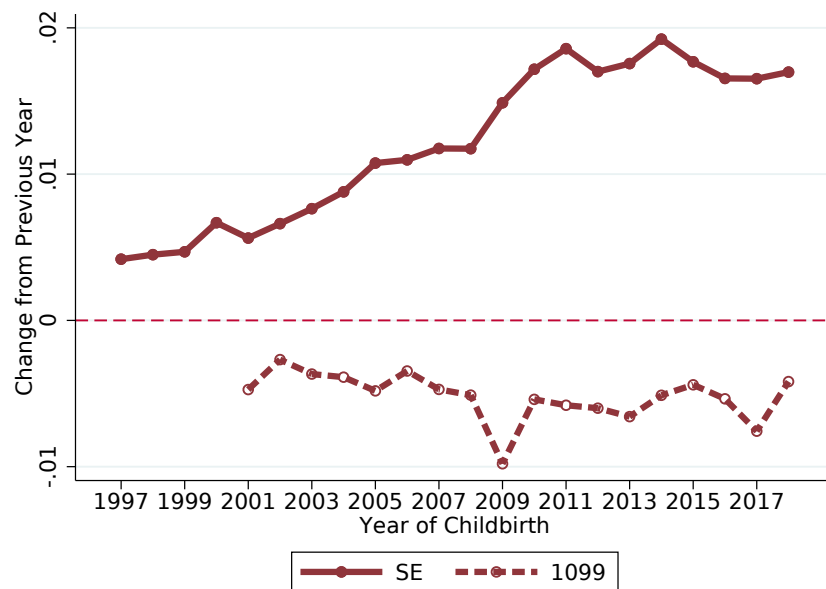
(b) Childbirth Event Study Estimates, 2003-2005 Births Versus 2011-2013 Births



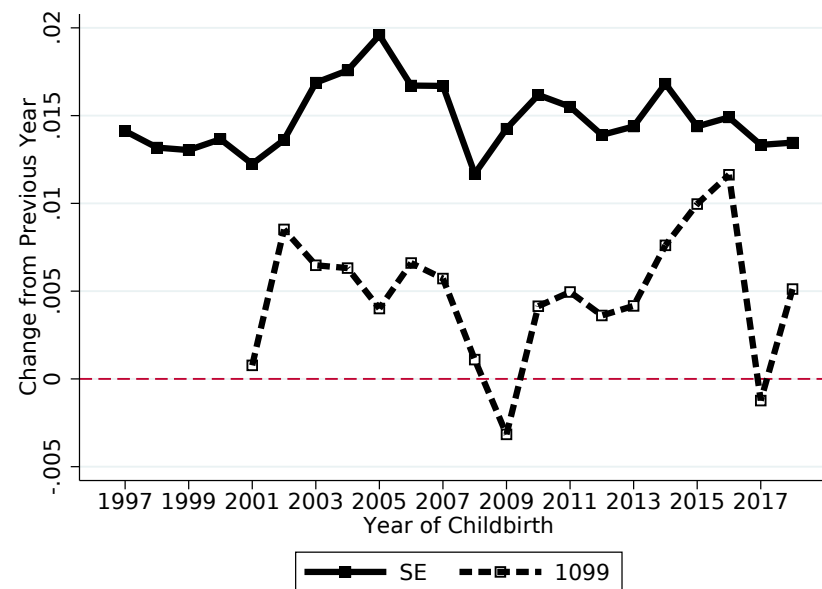
Notes: Panel A shows the average change in propensity to file SE (solid line) or receive a 1099 Information Return (dashed line), in the year of first childbirth reported on the x-axis. Panel B plots event study coefficients for separate regressions run on the indicated time-period and for the indicated outcome. See text for more details.

Figure B.2: Change in Self-Employment Around First Childbirth, 1997-2018 Births, By Gender of Parent

(a) New Mothers



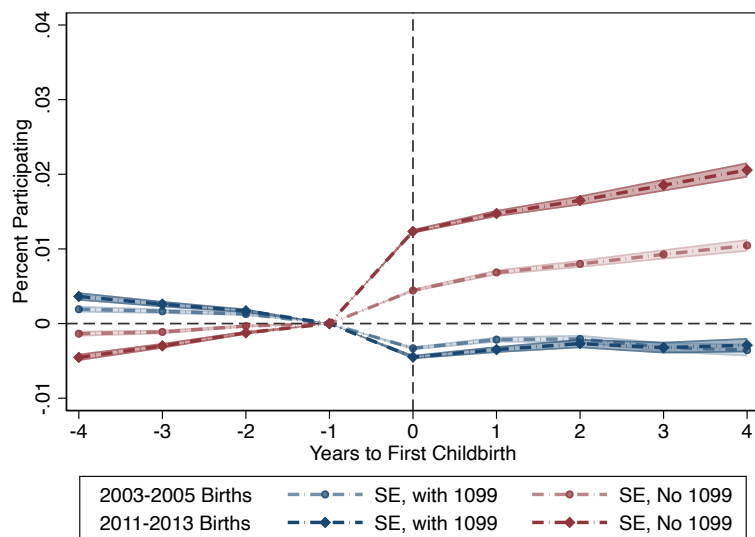
(b) New Fathers



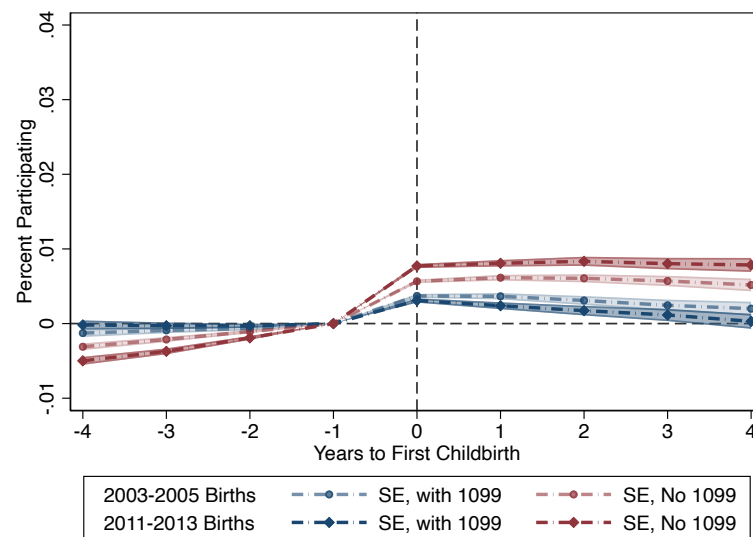
See notes for Figure B.1a.

Figure B.3: Childbirth Event Study Estimates, Additional Estimates

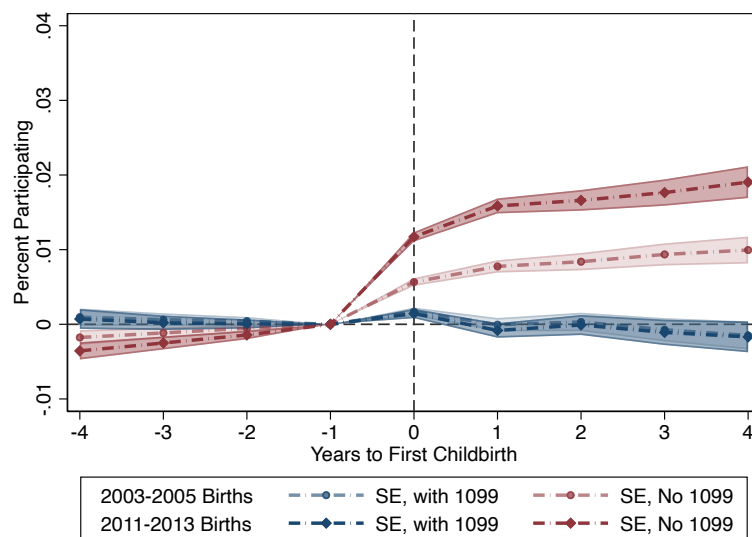
(a) Mothers Only



(b) Fathers Only



(c) All Parents, December Births Only



Notes: See notes for Figure B.1b.

C Data Appendix

This appendix describes the technical details of our data construction where we combine data from a variety of different tax forms.

The core of our analysis draws on de-identified, or “masked”, W2, 1099-MISC, and 1099-K information returns along with 1040 individual tax returns and associated schedules (e.g. Schedule SE). We begin with the population of individuals who appear as primary or secondary filers on a 1040 in each year. We create a record of all de-identified individuals, using masked Taxpayer Identification Numbers (TINs) appearing on these forms, attributed to either the primary filer or the attached spouse.

For all years, we merge in self-employment information for individuals and their spouses from Schedule SE. On Schedule SE (a schedule of Form 1040), individuals report all self-employment income subject to SECA taxation, so long as the total exceeds \$400. This includes active income from wholly-owned businesses on Schedule C, income from partnerships on Schedule K1, and farm income on Schedule F. Importantly, SECA taxes are assessed on individuals, not income tax filing units, so Schedule SE is always identified at the individual level.

We next turn to cleaning and processing the information returns. For Form W-2, we pull all W-2s with TINs that have been validated by the IRS. We eliminate duplicate or amended returns, and we drop a small number of invalid TINs (approximately 50,000 in 2016) and TINs considered “unmatchable” (approximately 5.2 million). Both of these are small compared to the overall number of W-2s, which exceeded 240 million in 2016. We use the recipient TINs to match W-2s to our main file of individuals. Since a large number of individuals with low W-2 earnings are not required to file 1040 returns, we add all cases with valid W-2s but no 1040 to our population file.

We then merge on information from Form 1099-MISC. We pull everyone with non-zero non-employee compensation reported in Box 7. To identify the online platform economy, we use the list of roughly 50 large labor platforms from Collins, Garin, Jackson, Koustas, and Payne (2019) that are mentioned in public databases than can be identified in the tax data (along with the corresponding EIN) using the unmasked firm name. Using the corresponding masked EIN, we then identify all 1099-MISCs in our cleaned file coming from these platforms and classify them as OPE income.

Reporting rules for intermediaries have changed over time in important ways that affect our measurement of the OPE. In 2011, a new law went into effect requiring companies that processed credit cards, electronic payments, or other transactions to report each recipient’s

payments on a new information return, “Form 1099-K.”⁴⁸ Starting in 2012, several online intermediaries in the OPE began issuing the new Form 1099-K instead of 1099-MISC for non-employee compensation. The income paid to gig workers on OPE labor platforms is, for all practical purposes, non-employee compensation. However, 1099-Ks are also issued for income from sales that is not non-employee compensation. We therefore also identify and track the 1099-Ks issued by the approximately 50 important online “gig” platforms where self-employed individuals offer labor services to firms or individual clients mentioned above. We then measure the total payments individuals receive from these companies that are reported on either a 1099-K or a 1099-MISC with non-employee compensation. We also explore alternative approaches to identifying OPE work, as some companies cannot be identified by this method.⁴⁹ For example, we use mentions of platform names in taxpayer-reported descriptions of business activity (line A) on Schedule C to identify additional instances of OPE work.

A potentially important limitation to studying the 1099-K is that companies in the labor OPE classifying themselves as third party networks are only required to file this form if the total amount of such transactions exceeds \$20,000 and the aggregate number of such transactions exceeds 200. In practice, this does not appear to impact our analysis through 2016, as we find most of the major platforms have issued 1099-Ks to all platform participants, regardless of the earnings level, in at least some years. However, beginning in 2017, more platforms begin to abide by the reporting thresholds, and so our measure of gig work is underestimated after 2016. In our analysis, we use Box 1 gross receipts to measure payments. We clean these forms using the same methodology described for the 1099-MISCs. We attribute 1099-K OPE payments to individuals, and add this to OPE income. We consider this income to be a part of the “1099 economy” and include it in measures of “1099 recipients” or “1099 income.” So that our definition is more comparable over time, we only classify someone as an OPE worker if they receive a 1099-MISC or have 1099-K earnings of \$600 or more; (Collins, Garin, Jackson, Koustas, and Payne, 2019) provides tabulations that include full counts of 1099-K workers, regardless of amount earned.

Worker characteristics Marital status and claimed dependents are defined for 1040 filers only. Marriage is determined from listing a spouse on a 1040. Dependents are determined from listing dependents (other than the spouse) on the 1040 and from a database of parent-child links maintained by the Social Security Administration. For measures of household earnings, wages and 1099 earnings are merged in for the spouse. Additional characteristics

⁴⁸This measure was included in The Housing and Economic Recovery Act of 2008, but did not take effect until the 2011 tax year.

⁴⁹For some platforms that pay through the payment processor Paypal, the 1099 will be issued by Paypal, and cannot be separately tied to a company in the OPE.

are merged in from other sources. Birth dates and gender are pulled from the DM-1 file, populated by the Social Security Administration.

D Proofs of Propositions

D.1 Proof of Proposition 4.1

If non-complier types are not eligible for or not aware of the credit ($E_i = 0$ or $I_i = 0$, respectively), or if they are eligible and aware but have wage earnings in excess of the top of the phase-in range ($E_i = 1$, $I_i = 1$, and $w_i > \hat{r}$) then they expect all self-employment profits to be taxed at a positive marginal tax rate and report $z_i = 0$. By contrast, if $E_i = 1$, $I_i = 1$, and $w_i < \hat{r}$, then there is a net benefit from reporting positive self-employment earnings up to $z_i = \hat{r} - w_i$.

In principle, it is optimal for all non-complier types with $E_i = 1$, $I_i = 1$, and $w_i < \hat{r}$ to report *exactly* $z_i = \hat{r} - w_i$. One can relax this sharp condition without fully modelling reporting decisions as in Allingham and Sandmo (1972) by considering that the government likely knows this and suspects individuals who report exactly $z_i = \hat{r} - w_i$ are reporting fraudulently. One might then suppose that individuals, knowing the government to behave this way, choose some amount $z_i \in (0, \hat{r} - w_i]$ with the amount depending on personal attitudes towards detection risk.

D.2 Proof of Proposition 4.2

Individuals without children or with wage earnings above the refund-maximizing amount ($BTK_i = 0$) report self-employment honestly with probability $1 - \theta$ and report zero self employment earnings with probability θ , such that

$$\rho^0 = (1 - \theta)\sigma^0 + \theta \times 0 = (1 - \theta)\sigma^0$$

which is strictly below the true unemployment rate.

For individuals with $BTK_i = 1$, their behavior depends both on their compliance type N_i and their information type I_i . All honest types report honestly. Non-complier types perceive a strictly positive marginal tax rate and therefore report zero self-employment earnings when they are unaware of the credit ($I_i=0$). Non-complier types with $BTK_i = 1$ who are aware of the credit *always* report positive self-employment earnings. Accordingly

$$\rho^1 = (1 - \theta)\sigma^1 + \theta\lambda \times 1 + \theta(1 - \lambda) \times 0 = (1 - \theta)\sigma^1 + \theta\lambda$$

In a low information environment, the self-reported self-employment rate among these individuals will be below the true self-employment rate. However, with high degrees of information—specifically, if the share of non-compliant filers who know about the credit exceeds the share with actual self-employment profits—the reported self-employment rate can exceed the true rate within this group.

Since the overall shares can be expressed as $\sigma^{tot} = \sigma^0(1-\kappa) + \sigma^1\kappa$ and $\rho^{tot} = \rho^0(1-\kappa) + \rho^1\kappa$, the above results imply that

$$\begin{aligned}\rho^{tot} &= (1-\kappa)\rho^0 + \kappa\rho^1 \\ &= (1-\kappa)[(1-\theta)\sigma^0] + \kappa[(1-\theta)\sigma^1 + \theta\lambda] \\ &= (1-\theta)\sigma^{tot} + \kappa\theta\lambda\end{aligned}$$

thereby completing the proof.