

BABY STEPS IN TAX POLICY: IMPACTS OF THE FEDERAL ADOPTION TAX CREDIT*

Chandler Inman
University of Michigan

November 1, 2025

[Click here for the most recent version.](#)

Abstract

The federal adoption tax credit is among the largest credits claimable by an individual filer in the US and is one of the primary tools used to facilitate child adoption. A temporary reform in 2010 and 2011 made the credit fully refundable, affording the opportunity to trace the heterogeneous adoption response of taxpayers across the distribution of tax liability. The estimates imply increasing access to the credit was effective at encouraging child adoption, inducing an additional 36,188 adoptions among those impacted by reform and that the average increase in fiscal transfers required to encourage an additional adoption ranges between \$15,500 and \$17,500.

*I would like to express my sincerest gratitude to my committee for their constant support and guidance: James Hines, Ashley Craig, Sarah Miller, and Ana Reynoso. I would also like to thank Charlie Brown, for providing consistently useful feedback through our correspondences, as well as the whole of the Michigan Public Finance community, for their high quality comments and willingness to provide a space to think through ideas jointly. Finally, I would like to thank my peers, especially, Katherine Fairley and Peter Choi, for their willingness to discuss ideas throughout all the stages of their development. Email for comments: inmanch@umich.edu.

1 Introduction

While most children are cared for by their biological parents, roughly 2.2 percent of children in the United States are adopted according to the US Census. The costs of adoption can be significant, easily exceeding tens of thousands of dollars, and adopting often involves becoming the parent of an especially vulnerable child who may need additional time and attention.¹

To offset these costs, subsidize the care of these children, and encourage more parents to adopt, the US federal government offers a tax credit worth up to \$17,280 per adoption as of 2025. Despite its high nominal amount, the credit has been completely nonrefundable for most of its history, meaning the credit was only usable for households to the extent that they had positive tax liability to offset.² This implies that households with low levels of tax liability are offered less incentive to adopt relative to households with higher tax liability, and those families with low levels of tax liability who choose to adopt, along with their adoptive children, receive a lower benefit for doing so.

This paper demonstrates how financial incentives can be an effective tool at encouraging child adoptions by studying the impacts of the federal adoption tax credit. Specifically, the responsiveness of adopting behavior to the tax incentive across the distribution of tax liability, the types of adoptions the credit ultimately encourages, and the increase in tax expenditures necessary to encourage an additional adoption are estimated. Each of these considerations are of prime importance for optimal credit design, as they determine both the capacity of the credit to encourage additional adoptions and redistribute resources to adoptive families along with their newly adopted children.

The strategy taken is to compare the adopting behavior of households that gained access to additional credit to those who experienced only an insignificant change due to a reform that temporarily made the credit fully refundable from 2010-2011. To parse the heterogeneous treatment effect of the reform across households, estimation relies on the multi-valued difference in differences design suggested by Callaway et al. (2024).

These comparisons have not previously been feasible, since they require connecting household characteristics to contemporaneous adopting behavior, which has not been explicitly measured in any data set. The present work offers an approach to do just this, by using the demographic characteristics in publicly available Census data to impute whether

¹Foster children made up more than 35 percent of non-step adopted children in 2008, and this percentage has grown over time (Malm and Vandivere, 2009). Despite these barriers, survey evidence suggests that roughly 1/3 of families have considered adoption at some point in their lives (“US Adoption Statistics — Adoption Network”, 2020).

²The credit was first instituted in its current form in 2002 and was worth \$10,000. It was set to automatically adjust for inflation annually.

a child is newly adopted. This enables estimation of the impact of exposure to reform, and hence the impact of gaining access to additional credit, on adopting behavior across household-level characteristics.³

The reform increased adoptions by 58.6 percent for those with zero tax liability and 46.3 percent for those with positive liability, relative to their pre-treatment averages, implying an average partial elasticity of 4 percent per \$1000 of additional credit for the 0 liability group and 5.6 percent per \$1000 of additional credit for the positive liability group. In total, the reform induced the treated groups to adopt 36,188 additional children than they would have in the absence of the refundability reform, and the estimates suggest these tended to be domestic adoptions of children aged less than 2.⁴ The average additional expenditure required to encourage an additional adoption varied across the distribution of tax liability⁵, costing on average \$15,801 for those with zero liability, \$15,533 for those with positive liability, and \$17,425 for those with \$1-\$5,000 of liability.

If a policy maker were seeking to simply maximize adoptions among these groups at a minimum cost, targeting lower and higher liability groups may be more cost effective than targeting those in the center. Of course, the value of these expenditures ought not be conceptualized as merely a cost, given that they are transfers to households. For example, the fact that those with lower income and hence lower tax liability are typically more inclined to adopt from foster care implies an increase in expenditures to them may have a significantly high social marginal value, even if the increase resulted in no increase in adoptions.

This paper is the first to evaluate the effect of a large financial incentive on adopting behavior in the US across households and for all adoptions. Despite studying a broader population than that assessed in the existing literature, the estimates presented here are intuitively commensurate with other studies that have evaluated only the effects on adoptions from foster care (Buckles, 2013; Brehm, 2021; Doyle and Peters, 2007; Hansen, 2007).⁶ Of particular relevance to this paper, Brehm (2021) estimates the impact of the 2010 adoption tax credit reform on adoptions from foster care, estimating that the credit’s imminent expiration induced an increase of roughly 2,400 additional public adoptions in the month of December 2011, constituting a 44% increase, roughly scaling with the estimates presented

³Exposure to reform relies on the imputation of the level of additional credit households gained access to due to the reform using the National Bureau of Economic Research’s tax simulation tool, i.e. TAXSIM. Importantly, the income variables are relative to the year prior, which helps accounts for any endogenous income responses induced by the reform.

⁴These estimates do not suggest older children did not also benefit from reform, only that the measure used will most reliably detect the adoption of younger children.

⁵The average fiscal expenditure is how much tax expenditures have to increase to encourage an additional adoption, given there are households who would have adopted and claimed the credit regardless.

⁶While there is much more data on children adopted from foster care, very little information is recorded on the economic characteristics of the households conducting the adoption.

in this paper.

Additionally, studies that have examined the relationship between household characteristics and other adoption types beyond public adoptions, provide results that are either not nationally-representative, or do not make use of exogenous variation to estimate causal effects. Despite these limitations, these papers broadly corroborate the descriptive evidence found using this paper’s measure of new adoptions. This provides important validation that the measure presented in this paper tracks actual adopting behavior. Baccara et al. (2014) using data from a private adoption facilitator, matching theory, and the fee for facilitation services, estimate a higher willingness to pay for girls and white children relative to boys and black children, which tracks the aggregate finding of a higher proportion of new adoptions being composed of girls and a higher proportion being white at higher income levels. Similar findings are documented by Khun et al. (2020), which utilizes the National Survey of Adoptive Parents⁷ and variation in adoption cost to estimate the rate of substitution of adoption types between 1990-2008 using a logit framework.

While there is a relatively active literature on the economics of foster care, very little is known about the economics of international or private domestic adoption, and while there have been studies on the the adoption of children from foster care, very little is known about the adoptive parents.⁸

This paper is the first to examine how exogenous variation in financial incentives impacts adopting behavior for particular adoptive households, across all adoptions, providing new evidence that financial incentives are an effective means of encouraging child adoption and that taxpayers across the distribution of tax liability are responsive to them.

2 Adoption Costs in the US

The process of adopting and the costs involved are highly heterogeneous and depend on the type of adoption being performed, the jurisdictions under which the adoptive parent resides, and the jurisdictions regulating the welfare of the adoptive child. Adoptions done through a private agency domestically or internationally can have expenses ranging from \$15,000 to \$50,000 (Malm and Vandivere, 2009). These avenues primarily result in the adoption of younger children, with more than 2/3 of these being less than a year old. These children are less likely to be diagnosed with a disability, are whiter, and more likely to be girls relative

⁷This is a nationally representative survey conducted in 2008, that contains a wealth of information on adoptive parents and children.

⁸Even beyond the economics literature, very little is known about the parents placing their children up for adoption outside of foster care. This group has, historically, suffered enormously and has been very understudied. See the stellar work by Ellen Herman (2012), who documents a variety of primary sources that outline the struggles of birth parents, adoptive parents, and adopted children throughout US history.

to the general population (Malm and Vandivere, 2009).

In contrast, adoptions from foster care tend to have nearly zero upfront costs, with states and counties often subsidizing their care, even after the adoption is finalized. These children are much more likely to have a disability, are older, and come from more vulnerable backgrounds. Adoptions of these children tend to have highly beneficial upfront welfare consequences, as spending a longer time in foster care is associated with a variety of adverse outcomes across education, health, and crime, each of which generate large negative spillovers (Bald et al., 2022; Baron and Gross, 2022; Hansen and Hansen, 2006). Despite these children likely needing the highest amount of care, in 2009, 45% of the children adopted from foster care lived in households with income less than twice the poverty threshold (Malm and Vandivere, 2009).

3 The Federal Adoption Tax Credit

The federal adoption tax credit was passed in its current form in 2001 under The Economic Growth and Tax Relief Reconciliation Act. The maximum claimable credit amount per adoption was set at \$10,000, but is recalculated yearly to adjust for inflation. The credit is nonrefundable, meaning only taxpayers with sufficient levels of tax liability are able to claim it, but possesses a carry forward of up to 5 years. In addition, the credit possesses a phase out range, which was set to begin after \$150,000 in MAGI⁹, reducing proportionally to \$0 at \$190,000.¹⁰ Interestingly, while the lower bound is set to adjust for inflation yearly, the statutory upper bound—the MAGI level where the claimable credit becomes 0—is fixed at a nominal distance of \$40,000 above the lower bound¹¹, implying that in real terms, the credit’s phaseout region becomes more steep over time.

The credit is designed to distinguish between different kinds of adoptions. In the case of international and private domestic adoptions, the credit may be used for eligible adoption expenses, which include all costs associated with the process of adopting a child such as legal services, home study fees, and the cost of travel. In the case of “special needs” adoptions, which mostly comprise adoptions from foster care, the full credit may be claimed regardless of adoption expenses.¹² The credit is claimable the year the adoption is finalized or, if the

⁹A filer’s MAGI is their AGI, plus additional sources of income that were excluded from their AGI, such as IRA contributions and student loan interest. For most taxpayers, AGI and MAGI are roughly equivalent. Instructions can be found on form 8839.

¹⁰In 2025, the credit amount is \$16,810 and the lower bound is set at a MAGI of \$252,150 with an upper bound at \$292,150.

¹¹In other words, the point at which 0 credit is claimable is simply the point at which the phaseout region begins, which is adjusted for inflation annually, plus \$40,000.

¹²Each state has their own specific definition of special needs. Nonetheless, the vast majority of adoptable children in foster care are given this designation. In 2013, this included 91 percent of foster children (Brehm,

adoption is private and domestic, the year after the expense is incurred. The carry-forward provision allows the credit to be usable for these adoption expenses so long as a taxpayer has enough liability within that time horizon.¹³

These provisions broadly imply that the credit is usable across adoption types, whether they are public or private¹⁴, and that the primary constraint to its usage when adopting, for the majority of taxpayers, is tax liability.

3.1 The Reform

In 2010, under the Affordable Care Act, the tax credit was made fully refundable and the maximum value was increased by \$1,000. This means that taxpayers with zero tax liability who would have previously not been able to benefit from the credit, now qualified for up to \$13,360 to offset adoption expenses. Then, on January 1st 2012, the credit reverted back to its non-refundable state, and the \$1,000 increase was also allowed to expire.

The reform led to increases in both tax expenditures on the credit and the number of taxpayers who filed for it. Tax expenditures on the credit rose from \$278 million in 2009 to \$1.2 billion in 2010, and the number of filers who claimed the credit increased from 80,676 in 2009 to 140,676 in 2010. While expenditures were also higher in 2011, \$668.1 million relative to the pre-reform period, the number of filers declined to 51,539. Part of the reason for this may be due to the decreased value of the carry-forward provision, given the credit was made refundable. More concretely, under the regime in which the credit is nonrefundable, a fraction of the people filing for the credit in a given year are made up of individuals who incurred adoption expenses within the prior 5-year period, but did not have enough tax liability to make full use of it. When the credit became refundable in 2010, tax liability no longer was a constraint, and so, unlike tax year 2009 and 2010, the filings in 2011 do not represent as many individuals carrying forward credit from the year previous.

While the average filer in every income bracket received additional credit, the largest proportional increases occurred for those making under \$50,000. From 1997 to 2009, 10 percent of tax expenditures for the credit were attributable to those making below \$50,000. In 2010-2011, this expenditure share increased to 46 percent. In the same years, those making less than \$15,000 received 0 percent of tax expenditures, and in 2010 to 2011, the share increased to 13 percent (Brehm, 2021). These statistics suggest the reform substantially in-

2021).

¹³In addition, adoption expenses paid for through qualified adoption assistance provided by an employer may not be covered by the credit and reduces the filer's maximum credit amount. Special needs adoptions provide an exception to this.

¹⁴Given private adoption is so expensive, plenty of expenses can be incurred to make it usable. In the case of public adoptions, the credit is typically not tied to cost, since most of these are special needs.

creased the usability of the credit for those with lower tax liability in addition to providing a sizable cash transfer for those already benefiting from it.

A full accounting of the potential social welfare consequences of this policy requires the evaluation of a number of additional considerations beyond its impact on the number of adoptions. Even if the credit induced no change in behavior¹⁵, whether or not the policy induced a net gain in social welfare hinges on the relative social benefit of the fiscal transfer to these families, the social value of public funds lost due to the increase in tax expenditure, and the resulting associated fiscal externalities, which includes the present value of future fiscal savings or losses it induced (Hendren and Sprung-Keyser, 2020).

The reform disproportionately increased payments to those with very low incomes, increasing the credit’s redistributive effect. If the social benefit of transferring these funds to these families is high enough, even if the policy induced no change in adopting behavior, it could have resulted in a net gain in social welfare regardless. Even beyond the redistributive implications of the policy, there is evidence that cash infusions to those with children and to those with foster children in particular can result in increased expenditures on childcare and improved educational outcomes (Buckles, 2013; Simon et al., 2024), which in turn, could offset the increase in tax expenditures induced by reform through the value of higher receipts in the future. It would not be surprising if the social value of expenditure savings induced by the policy could have been enough to offset the rise in expenditures as current evidence suggests this is often the case for policies targeting children (Hendren and Sprung-Keyser, 2020).¹⁶ The current analysis does not speak to these considerations, but they are vital to evaluating the merits of the credit, whether the reform improved it, and what changes ought to be made to its structure today.¹⁷

¹⁵Note: this is quite different than inducing no change the number of adoptions.

¹⁶There are great number of additional considerations that the present analysis also does not speak to. For example, if adopting behavior truly was totally inelastic with respect to the credit, it could serve as a tag for higher earning potential, given the home study that is often required (Akerlof, 1978).

¹⁷A note on administrative burden: during this reform period, the IRS, and hence filers as well, suffered a large increase in administrative burden. The complexity of the credit’s income limitations, the different state definitions of “special needs”, the large amount of documentation required to validate expenses, and the large sums at stake induced an audit rate of 69 percent for adoption tax credit claims in the 2012 filing season. 55 percent of these resulted in no change in taxes owed or refund due, with the average correspondence audit taking 126 days (Bogadi, 2012). For 2011, only \$11 million of the 668.1 expenditure was disallowed, with the IRS having to pay \$2.1 million in interest for withholding refunds due over 45 days. For any analysis of welfare, the administrative costs and increased burden of reform must be accurately factored in (Keen and Slemrod, 2017). Although presently outside the scope of this analysis, these considerations will also be vital for deciding whether it is optimal to re-implement this reform in the future.

4 Data

This paper employs the Census’s American Community Survey (ACS) 1% samples for the years 2008 to 2022. These data provide a yearly, cross-sectional random sample of the US population, including detailed self-reported information on income, family structure, and demographic characteristics, all of which are vital for imputing federal tax liability and inferring adopting behavior. For all estimates and descriptive statistics, the ACS’s household survey weights are used so that the results are nationally representative, unless stated otherwise.

4.1 Imputing Tax Liability

While the ACS does not directly provide tax data, the tax liability of each household may be imputed by utilizing the wealth of economic variables contained in the ACS paired with the National Bureau of Economic Research’s (NBER) TaxSim program.¹⁸ This program takes as inputs household-level income variables and, utilizing the state and federal income tax schedules, computes a variety of tax-relevant household-level characteristics. In order to calculate each household’s level of exposure to reform, proxied by the increase in credit access each household experiences due to the credit becoming refundable, reasonable estimates of each household’s modified adjusted gross income and their federal tax liability are necessary. One interesting silver lining of the American Community Survey’s sample design is that income variables are reported for the year prior, while demographic variables are contemporaneous. This allows the use of lagged tax liability, which is less likely to be endogenous compared to contemporaneous tax liability.¹⁹

In order to reduce the number of assumptions that must be made in imputation, each household’s federal tax liability before credits assuming the standard deduction is used.²⁰ The reasons for this are twofold. First, and most importantly, the ACS does not record take-up of credits and does not record deductible expenditures for each household. Imputing these values would require overly strong assumptions on household behavior. Secondly, while in theory the credits and deductions each household chooses to utilize could be imputed, using a household’s federal tax liability before credits measures their maximum capacity to make use of the non-refundable version of the credit, which is not contaminated by the endogenous take up response of other credits. This best captures the underlying motivation

¹⁸While the ACS provides their own estimates of tax liability using TaxSim, they do not provide values before 2009. The estimates are roughly equivalent and if used, do not change the results.

¹⁹Of course, panel data would be used in the ideal scenario, as this would allow the use of lagged income just outside the treatment time. This isn’t possible given the ACS’s structure.

²⁰For reference, in 2022, 91 percent of taxpayers took the standard deduction “<https://taxpolicycenter.org/briefing-book/what-standard-deduction>”, 2025

for the definition of treatment: being offered additional credit, abstracting from the take-up responses of other credits that may be available to the household.

There are some additional important simplifications noted here. First, the Census does not report whether households file jointly or separately. Therefore, it is assumed that households file separately if they are separated but not divorced, but jointly if they are married even if the spouse is not currently living in the household. This is likely a reasonable assumption, since in 2010 only 2.31% of married taxpayers filed separately (“SOI tax stats - Individual income tax returns filed and sources of income — Internal Revenue Service”, n.d.). In addition, while the ACS has data on interest and dividend income, it does not distinguish between them despite the fact they are treated differently by the tax code. It is assumed that dividend income is zero and the remainder is interest income. In addition, the ACS has no information on realized long or short terms capital gains or losses. These are set to zero as well.

To provide an assessment of the accuracy of the imputation, Figure A1 compares the imputed mean average level of federal tax liability before credits, relative to the aggregate average reported by the IRS. The top blue line is the average level of imputed federal tax liability before credits per household among all households imputed as having positive liability. The bottom blue line is the average level of imputed federal tax liability before credits for all households. Since households with very low levels of income are not required to file, these measures ought to serve as rough upper and lower bounds for the true mean level of federal tax liability before credits respectively.²¹ Despite the simplifications required for the imputation, the true aggregate hovers between these two measures during the treatment period, lending credence to the validity of the imputation exercise.

4.2 Imputing Adopting Behavior

Since 2008, the American Community Survey has collected data on whether or not a child is adopted.²² This measure has two important limitations. First, while the ACS does allow for a child to be categorized as a step-child, it does not distinguish between step-parent adoptions and other adoptions. As these types of adoptions would not qualify for the credit, ideally, these types of adoptions should be excluded when estimating the credit’s impact. Second, while the age of adopted children is recorded, the date of adoption is not. Given that children can be adopted at any age, using merely the total number of adopted children only

²¹Technically, even individuals who are imputed as possessing a very low level of tax liability may not be required to file if their income is low enough. Hence, the “upper bound” is only approximate.

²²The Census does have information on whether or not a child was adopted in 2000, but the categorization was slightly different. For example, the step-child category was entirely omitted, meaning household heads with step-children may have called them their “adopted child”.

provides information on the stock, rather than the flow of adoptions at the household level. This is crucial in order to connect contemporaneous adopting behavior with the current characteristics of the adopting family.

The primary goal of this imputation procedure is to connect yearly household-level adopting behavior to the household-level characteristics that determine exposure to reform. A graphical representation of the imputation procedure is given in Figure 1 and a detailed description is located in Appendix A. This procedure uses the age of the adopted child, their immigration status, and their migration status relative to other household members to ascertain whether or not they were adopted within the census year they are observed. With respect to step-parent adoptions, to the extent that step-children likely lived with their adoptive parent the year prior to adoption, these children will not be categorized as newly adopted. Nonetheless, to ensure the trends observed in the data are not being driven by trends in step-parent adoption, children are also identified using an additional measure that is more likely to exclude step parent adoptions, using the migration status of the spouse of the household head.

A potential drawback to this procedure is its potential to misclassify adoptions as not “new” when the child lives with their adoptive family for longer than a year prior to adoption. For private adoptions and for adoptions of very young children, this is unlikely to be a problem.²³ However, for adoptions of certain types, such as grandparent adoptions or adoptions of older children from foster care, this procedure is more likely to misclassify these children.²⁴ Given that the majority of studies analyzing the impact of financial incentives on adoption tend to focus on adoptions from foster care, the results presented here should be seen as complementary with existing results, best able to track the adopting of younger children, but understating the adoption of older, publicly adopted children.²⁵

This measure of new adoptions implies that 73,671 children were newly adopted across the US in 2014. The only publicly available statistic on the total number of non-relative

²³The National Survey of Adoptive Parents records that most private adoptions are young children and that most internationally adopted children were not living with their adoptive family the year prior.

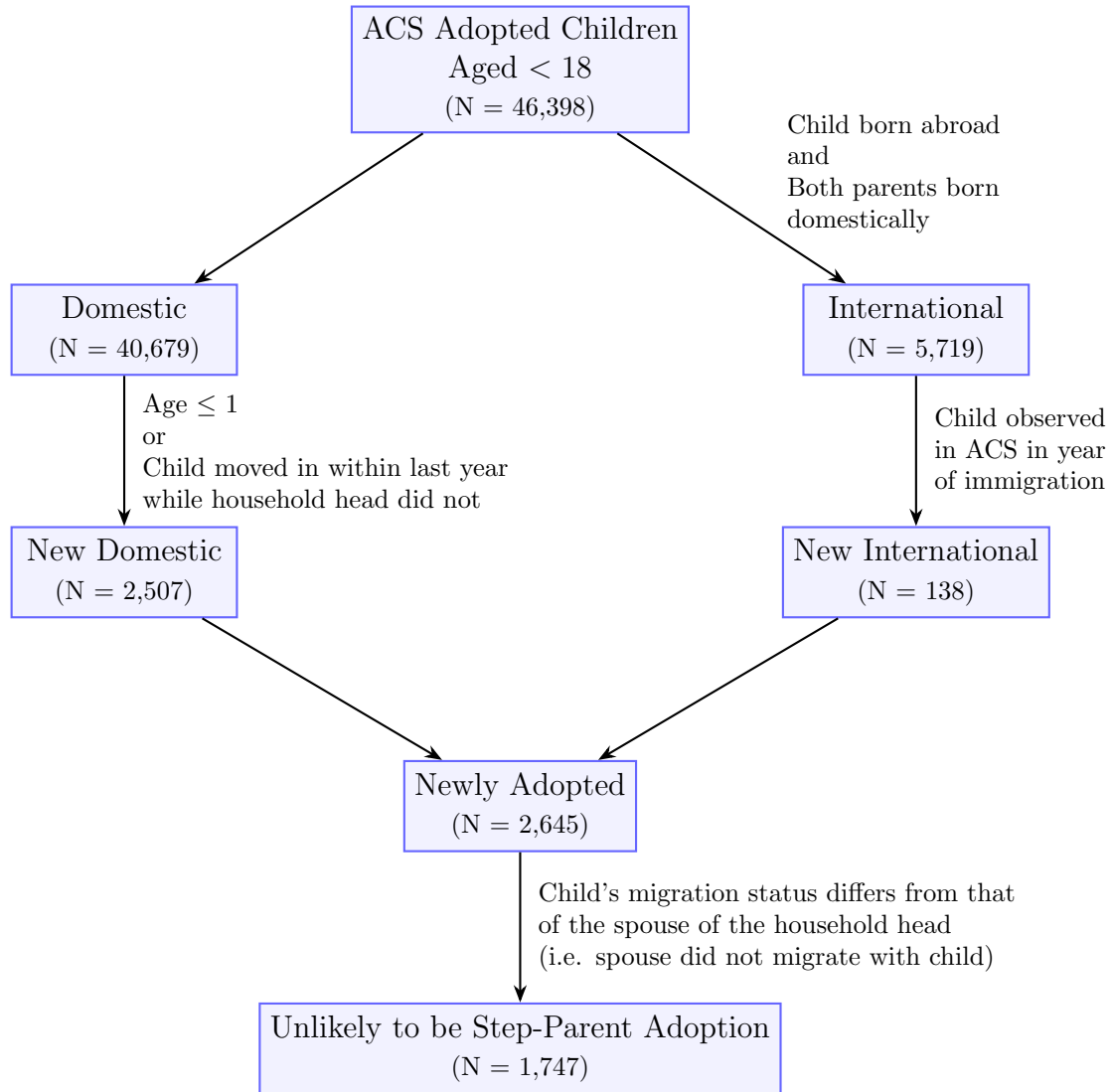
²⁴The average length of time to adoption for an adoptable child in foster care is highly heterogeneous. The average time is around two years (Bald et al., 2022), but notably, this is not the length of time the child is with their adoptive family before adoption. This information is not recorded in the national census of children in foster care. However, according to the National Survey of Adoptive Parents, 44 percent of publicly adopted children lived with a foster family other than their adoptive family immediately prior to their adoptive placement, and 11 percent lived in congregate care.

²⁵This also may not be too much of a concern if policy makers are more interested in determining the credit’s impact on the capacity of adoptive children to find their permanent home. If a child were already living with what will be their permanent family, and the credit merely induced them financially to complete the formal process, this “new adoption” is clearly distinct to the case of a child with no permanent home gaining access to one.

adoptions in the US during the sample period was measured in 2014 at 75,337²⁶, implying that the measure presented in this paper accurately captures new non-relative adoptions.

²⁶This statistic was compiled by Johnston (2022).

Figure 1: Adoption Imputation Tree



Note: N denotes the unweighted number of observations from the 2009-2011 American Community Survey 1% samples.

4.3 Characteristics of Newly Adopted Children

Table A3 show the demographic characteristics of children flagged as newly adopted across the distribution of tax liability. As anticipated, these children are more similar to the average privately adopted child than the average publicly adopted child as reported in the National Survey of Adoptive Parents (Malm and Vandivere, 2009). They are quite young, averaging less than two years of age, more likely to be classified as white, and more likely to be female (Malm and Vandivere, 2009). In addition, and consistent with aggregate data (Khun and Lahiri, 2017), the share of internationally adopted children is declining. This may be why the average age of adopted children and the fraction who are white is rising over time, as internationally adopted children tend to be younger and are less likely to be white. The average age of adopted children tends to hover between 1.5 and 2.5. While it is also generally true that the age of the adoptive child is decreasing in tax liability, the group with the youngest newly adopted children is the zero liability group. Given that publicly adoptable children tend to be adopted by lower income households and that this measure may be missing some older, publicly adopted children, this is not necessarily surprising.²⁷ This implies that the adoption response estimated is best seen as capturing the response of the adoption of younger children, rather than older publicly adoptable children.

4.4 Characteristics of New Adopters vs Non Adopters

The descriptive statistics implied by these measures tend to corroborates other descriptive findings on who adopts (Khun et al., 2020; Khun and Lahiri, 2017; Baccara et al., 2014). Table A1 reports three of the most salient—income, age, and same-sex status—across what will constitute the control and treatment groups, over time. Table A2 reports additional demographic and economic statistics.

The columns of Table A1 Panel A reveal new adopters tend to be higher income within tax liability groups and that this difference has persisted over time. This makes sense given the cost of adoption and that many forms of adoption require the household to complete a home-study, the passage of which is likely associated household income.

The columns of Table A1 Panel B report the mean age of household heads over time. New Adopters tend to be younger, and this difference in age is most striking among those

²⁷Another possibility is that lower-income households may be more likely to adopt the younger children of other family or community members. Since these households have less economic resources, the pooling together of informal resources, such as providing child care, through relationships of mutual reciprocity may be an important means through which these households weather their own idiosyncratic shocks. In other words, a household may adopt a neighbor's or family member's unexpected child to help them weather a hard time, which furnishes the expectation that when they face an idiosyncratic shock that neighbor or family member will help them. The tax credit would apply to expenses incurred in performing such adoptions and so the reform may have helped subsidize this type of informal community support.

households with zero tax liability. Zero tax liability households are more likely to be retirees relative to positive liability households, so for non-adopters, which are a good enough stand-in for the general population, this makes sense. The fact that this trend reverses among new adopters may mean that retirees, who likely already have children of their own, are simply less likely to adopt *given* they have zero liability and hence less resources.²⁸

The columns of Table A1 Panel C report the number of same-sex couples per 1000 households.²⁹ New adopters are around 5-6 times more likely to live in same-sex households relative to non-adopters, a finding also hinted at by Baccara et al. (2014) when examining a private adoption agency, and this difference has only grown over time, likely due to the growing acceptance of same-sex marriage and same-sex families during this period.³⁰

5 Estimating the Adoption Response to Reform

The reform granted households access to substantially different levels of additional credit, depending on their levels of tax liability and income. The core of the design makes use of this variation by comparing the adopting behavior of those who gained access to different levels of credit. The control group includes those households who experienced no or only a small increase in credit access. This intuition is implemented concretely through the discrete-multi-valued treatment difference-in-differences specification proposed by Callaway et. al. (2024).³¹ This design enables the estimation of group-level³² treatment effects, under the identifying assumption that the average change in adopting behavior of the control group is a good proxy for the average change in adopting behavior of each treated group, had the adoption tax credit reform never taken place.

More formally, following Callaway et al. (2024), let Y_{itg} be the adoption outcome of

²⁸This fact likely also implies some “new” grandparent adoptions are not being counted. If a child were to live with a grandparent for more than 2 years before being adopted, the child would not be categorized as newly adopted. This may not be a substantial concern, if the primary adoptions policy makers care about are non-relative adoptions. However, as will be discussed below, this may mean certain types of public adoptions are being missed using this measure.

²⁹While the ACS provides its own measure for whether a couple is a same-sex couple, it only started doing so in 2013. Hence, I impute whether a couple is a same-sex couple by simply looking at whether the household head and their partner are the same sex.

³⁰During the reform period, it is interesting to note the share of same-sex couples who were new adopters fell for the control but tended to weakly rise among most of the treated groups. This may imply treatment tended to be especially beneficial to same-sex couples of particular socioeconomic classes, but not monotonically. Given the fall in the share of same-sex couples among the 1-5000 liability group category, this benefit was likely not monotonically increasing in income. This may be due to differences in the types of children being adopted by same sex couples across the distribution of tax liability.

³¹Callaway et. al. (2024) does not deal with staggered timing of treatment adoption, but rather with variation in treatment intensity for a single treatment.

³²The groups will be defined by their level of federal tax liability before credits. For example, one set of groups are those with 0 tax liability and those with positive liability who are not in the control.

interest for a household i at time t in treatment group g . Let D_{it} be a multi-valued treatment indicator, taking on a discrete set of values that are group specific given by d_g . All estimates are relative to the control group who received only a marginal increase³³ in credit access, given by d_0 . The main specification, given by equation (1) below, is estimated separately through OLS for each treatment year, 2010 and 2011, relative to 2009.

$$Y_{itg} = \alpha_g + \delta_t + \sum_{g=0}^G 1(D_{it} = d_g) \beta_g + \epsilon_{itg} \quad (1)$$

The parameter β_g is interpreted as the average treatment effect on group g among those receiving d_g , defined in the subsequent section, under the assumption that the mean change in adopting behavior of the control group between 2009 and the treatment year of interest, is equal to the mean change in adopting behavior of treated group g , had the reform never taken place.³⁴

5.1 Calculating the Average Relative Level of Treatment Received by Groups

The estimates of β_g capture a non-parametric average adoption response of each group due to reform. To measure the average level of treatment actually received by each group, d_g , the relative level of additional credit households gained access to must be estimated. To do this, the level of credit each household would qualify for under the new regime in a given treatment year is calculated, based on their imputed federal tax liability before credits and their imputed MAGI. Then, the level of credit the household would qualify for under the counterfactual regime where the credit is non-refundable is calculated. The difference between these two quantities is the imputed household-level increase in available credit.³⁵ Within each group, this quantity is averaged over households. To make the average group-level treatment relative to the control group, the control group's average increase in available credit is subtracted.³⁶

³³These include households who benefited from the 1000 dollars increase, but did not benefit from the credit being made refundable due to their high level of tax liability, as well as households with high enough incomes to not qualify for any of the credit.

³⁴In addition, given the data is cross sectional, it must be assumed that changes in the unobservable and observable characteristics that are relevant to adopting behavior would have been stationary in the absence of treatment. If, for example, the group of households with zero tax liability was markedly different between 2009 and 2011 due to some factor unrelated to treatment, changes in outcomes due to this compositional change would influence the estimated treatment effects.

³⁵Again, with the caveat that their liability is computed using income variable reported for the prior year

³⁶As an example, consider a group of households with \$5000 of tax liability with sufficiently low MAGI to not enter the phaseout range. In 2009, these households that year would have access to \$5000 of credit. However, after the reform, households in 2011 with \$5000 of tax liability had access to \$13,360 of credit. This constitutes an increase of \$7360 dollars of additional credit after reform. Since the control group gained, on average, roughly \$700 dollars of credit, this translates into a group-level average relative credit increase of \$6660.

This average relative increase in credit access can be combined with the estimated average treatment effects to estimate average treatment effects per dollar of increased available credit. It is important to note that this is distinct from estimating the marginal treatment effect of increasing credit access by an additional dollar. In order to estimate this quantity, a much stronger version of parallel trends would have to be imposed (Callaway et al., 2024). Namely, that the change in adopting behavior of the group receiving d_g is a valid proxy for all other groups, g' , had they received d_g instead of $d_{g'}$.

5.2 Choosing Treatment and Control Groups

Treatment groups are defined by their level of federal tax liability before credits assuming the standard deduction, as this tracks the level of maximum credit they could qualify for within a given year and allows for a clean identification of household groups over time. The control group is defined as those households with tax liability above \$13,360, the maximum credit available in 2011. These households experienced a modest increase in credit access of \$700 on average, owing to the \$1000 increase in the credit due to reform.³⁷

Figure A2 shows the average imputed increase in credit across \$1000 tax liability bins. The reason why tax liability does not perfectly track exposure to additional credit is because the MAGI income limitations reduce the maximum credit available to households.

Grouping taxpayers by their level of tax liability requires balancing several tradeoffs. By choosing finer bins, households are more similar within a group and receive a more similar level of treatment, allowing for a better characterization of the treatment effect heterogeneity induced by the policy. However, since finer bins also contain fewer households, their mean adopting behavior over-time is less stable, making the identifying assumption of parallel trends less credible. Finer bins may also make estimates more susceptible to bias induced by compositional change unrelated to treatment over time.³⁸

Given these considerations, two sets of treatment groups are defined. The first is composed of the group of households with zero liability and the group of households with positive liability who are not in the control. These are termed the coarse liability bins. The second set is composed of four treatment groups: those with zero liability, those with \$1-\$5000

³⁷A natural alternative to this group would be the set of households with MAGI above the maximum threshold, as these households never qualified for the credit. However, this alternative group exhibited trends that were non-parallel after repeal, implying the parallel trends assumption necessary to use them would not be credible.

³⁸Another alternative to using bins would be to attempt to estimate treatment effects using a flexible function of the imputed increased level of credit access, as proposed by Callaway et al. (2024). This approach was not taken for two reasons. First, it requires *strong parallel trends*, which is unlikely to be valid in this setting. Second, evaluating the plausibility of parallel trends as well as whether or not enough statistical power is available to form valid counterfactuals is simply more transparent using the binned approach.

of liability, those with \$5001-\$10000 of liability, and those with \$10001-\$13360 of liability. These are termed the \$5000 liability bins.

5.3 Comparing Adopting Behavior Over Time

Figure 2 shows the mean evolution of adopting behavior for two treated groups as well as the control. These treated groups are those with 0 tax liability, and those with positive tax liability below the maximum credit amount in 2011. The dashed lines represent when the treatment occurred and when the credit reverted back to its previous nonrefundable state.³⁹

While the most common practice to assess the plausibility of parallel trends is the examination of pre-trends, as mentioned previously, the ACS does not record whether or not a child is adopted before 2008. Nonetheless, it is apparent that between 2008 and 2009 both the treated and control groups experienced parallel sharp declines in adopting behavior. Given there are not additional pre-treatment periods over which to observe pre-treatment trends, to garner further support that the treated and control groups would have evolved in parallel in the absence of treatment, their behavior after the reform expired is analyzed.

The danger in using this evidence, relative to using pre-trends, is that treatment may have induced the treated group to evolve in parallel to the control group when, in the absence of treatment, the treated groups would have evolved in a non-parallel fashion. In this sense, pre-trends provide less “contaminated” evidence that the control and treated group would have actually evolved in parallel in the absence of treatment, compared to using post-trends after a repeal. Nonetheless, examining the trend in behavior after repeal still lends credibility to the parallel trends assumption for two reasons. First, as can be seen in Figure 2, the treated and control group exhibit qualitatively quite similar behavior many years after treatment. None of the deviations in trend year-over-year, which will be measured in a variety of different ways, are statistically significant. Second, while treatment may have impacted long-run adopting behavior both directly through its 5 year carry forward provision⁴⁰ or possibly indirectly, through other mechanisms (i.e. long run-network effects⁴¹, a now larger adoption industry due to the subsidy, etc.), this does not imply that it induced what would have been a non-parallel relationship to be parallel. For example, if the credit induced a level increase in adopting behavior among the treated groups, post-repeal trends would still provide a similar type of evidence as pre-trends.

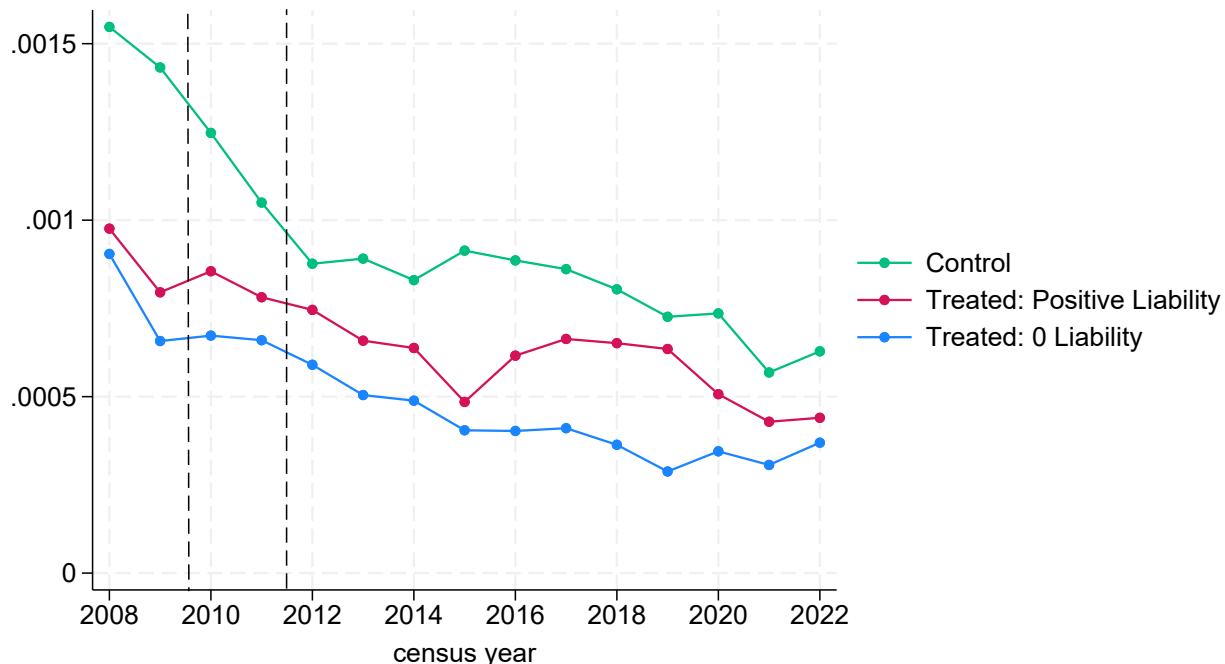
³⁹The expiry of the refundability regime occurred in January 2012, when the credit again became non-refundable and the \$1000 increase was reversed.

⁴⁰Expenditures on the credit declined to their pre-repeal state, suggesting payments of carried forward credit were not substantial during this time (Brehm, 2021).

⁴¹For example, if the credit encouraged people who never would have adopted to adopt, they could influence friends and family members to adopt through networking. The National Survey of Adoptive Parents reveals that many who choose to adopt previously knew and were influenced by other adopters.

All this is to say, it is impossible to directly evaluate parallel trends. Using pre-trends provides intuitive support because, the argument goes, why would, in the absence of treatment, the treated and control have evolved in a non-parallel fashion when they were evolving in parallel before?⁴² To the extent that this argument is convincing, a parallel argument exists in this context. Why, after the major provisions of the reform are repealed, would the treated and control group exhibit statistically similar trends, but not have done so in the past in the absence of those major provisions?

Figure 2: Mean Number of New Adoptions per Household



Note: This figure reports the mean number of new adoptions per household by treatment group from the 2008-2022 American Community Survey (ACS). The procedure for imputing which children were newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had non-zero tax liability, and the second treated group includes those with zero tax liability. The vertical dashed lines represent when the reform was enacted (starting in 2010) and when it was repealed (starting in 2012).

5.4 Estimating Year-Over-Year Placebo ATTs

In order to quantify whether the treated and control groups evolve in parallel both in the pre-period and during the post-repeal periods, a variety of placebo tests are performed to verify whether or not these trends are statistically distinguishable.

The first set of tests involve estimating equation (1), but acting as if treatment occurred in different years. During the pre-period, a pseudo-ATT is estimated by assuming treatment

⁴²Post hoc ergo propter hoc?

occurred in 2009, relative to 2008. During the post-repeal period, the same is done for each year, estimating a treatment effect for 2012 relative to 2011, a treatment effect in 2013 relative to 2012, and so on up to 2022. A statistically significant pseudo-ATT would be equivalent to rejecting the hypothesis that the trend in the outcome between the treated groups and the control group are the same (Callaway and Sant’Anna, 2021).

The next set of tests involves estimating pseudo-ATTs over 2-year time horizons. In this case, equation (1) is estimated to create pseudo-ATTs but is estimated over two year intervals. Again, a statistically significant treatment effect in these contexts is equivalent to rejecting the hypothesis that the treated and control group are evolving in parallel over that two year time horizon.

Finally, to obtain more conservative estimated placebo-ATTs using larger samples, as well as to better evaluate the parallel trends assumption underlining the treatment effect estimate that pools 2010 and 2011 together, equation (1) is estimated but with observations over two year sets being pooled together. Given the larger number of observations, these bi-annual pseudo-ATTs ought to have more statistical power, and so are more able to reject the hypothesis that the treated and control groups are exhibiting non-parallel trends in the post-reform period.

6 Results

6.1 Impact on the Mean Number of Adoptions Across Coarse Liability Bins

Table 1 column 1 shows the point estimates from specification (1) for β_g for years 2010 and 2011. All regressions utilize the ACS’s survey weights, making the sample and treatment effects nationally representative.⁴³

The other columns of Table 1 provide regression coefficients estimated including state fixed effects, state by year fixed effects, as well as a set of carefully chosen demographic characteristics that would not be impacted by treatment.⁴⁴ Estimates without controls are typically preferable, as including these would require both conditional parallel trends, as well as an implicit homogeneity and linearity assumption in how the covariates impact adopting behavior (Karim and Webb, 2024; Deb et al., 2024).

The point estimates in Table 1 represent the average treatment effect of d_g among those

⁴³Point estimates are of comparable size when making no use of the survey weights. Survey weights allow for the proper correction of heteroskedasticity in standard error calculations. See Solon et. al. (2013) for more details and “IPUMS USA”, n.d. for how the weights are constructed to incorporate sampling methodology.

⁴⁴These include whether or not the household head has been married for 2 years, whether or not there are any non-adopted children 2 years old or greater in the household, and race

in group g . Figure A7(a) plots these ATTs, along with the year-over-year pseudo-ATTs. Consistent with the findings in Brehm (2021), only significant effects in 2011 are detected. This is consistent with the view that it may have taken time for individuals to respond to the increased incentive to adopt. Table A6 presents specification (1) when pooling 2010 and 2011 together. These results partially validate this explanation, as point estimates for each liability group are reduced, but remain significant and are of the same order of magnitude. The treatment effects do not change substantially with the iterative addition of various controls, which speaks to their robustness under alternative conditional parallel trends assumptions. Neither the pre-period nor any of the post repeal pseudo-ATTs are statistically significant, indicating that during the period when the reform expired, the treated and control groups do not exhibit statistically different year-over-year trends.

The point estimate of 3.854 implies an average treatment effect of 3.854 additional adoptions per 10,000 households in the 0 liability group in 2011. Similarly, the point estimate of 3.69 in the positive liability group implies a treatment effect of 3.69 additional adoptions per 10,000 households in 2011. These increases constitute a 58.6 percent increase and a 46.3 percent increase in the average number of new adoptions per household, relative to their pre-treatment 2009 average, respectively.

The approximate increase in credit access, relative to the control, among these two groups in 2011 was \$12,660 and \$8,164, respectively. This implies a partial elasticity of 4 percent per \$1000 of additional credit for the 0 liability group and 5.6 percent per \$1000 of additional credit for the positive liability group. Scaling these estimates to their relative population sizes implies treatment increased the total number of adoptions in 2011 by 13,643 for the 0 liability group and 22,545 for the positive liability group. These effect sizes are large⁴⁵ and heterogeneous. The lower per dollar responsiveness of the 0 liability group relative to the positive liability group makes sense in light of the often high upfront costs associated with adoption, above and beyond the credit amount. For example, if it costs, say \$20,000 to complete a particular adoption, additional resources beyond the credit would be needed to carry it out. Given their level of resources, this adoption would likely not be feasible among those in the 0 liability group but would likely be possible for those in the positive liability group.

⁴⁵The measure for new adoptions implies there were 100,000 total adoption in 2009. As mentioned previously, this is likely an under count due to the nature of public adoptions.

Table 1: Treatment Effects for Coarse Liability Bins: Number of Newly Adopted Children

	(1)	(2)	(3)	(4)	(5)
ATT 0 Liab 2010	2.009 (1.731)	1.992 (1.732)	1.644 (1.730)	1.631 (1.730)	1.444 (1.746)
ATT Pos Liab 2010	2.451 (1.588)	2.450 (1.588)	2.401 (1.587)	2.403 (1.587)	2.205 (1.600)
ATT 0 Liab 2011	3.854** (1.640)	3.846** (1.640)	3.540** (1.639)	3.530** (1.639)	3.590** (1.662)
ATT Pos Liab 2011	3.690** (1.527)	3.674** (1.527)	3.687** (1.527)	3.671** (1.527)	3.723** (1.545)
Demographic Controls	No	No	Yes	Yes	Yes
State FE	No	Yes	No	Yes	Yes
State x Year FE	No	No	No	No	Yes
Observations	3,600,610	3,600,610	3,599,510	3,599,510	3,599,510

Note: The table displays the dynamic results for the regression design with coarse bins, where the outcome of interest is the reform's impact on the mean number of newly adopted children per household. The coefficients are transformed to represent the treatment effect per 10,000 households. The demographic controls include the household head's age and age squared, race, whether they were married for 2 or more years, and whether or not the household head has any non-adopted children over 2. Under unconditional parallel trends, the coefficients in model (1) may be interpreted as the average treatment effect on the treated, where treatment is conceptualized as the quantity of additional credit the group gained access to, though not necessarily actually used. Models (2)-(5) may be interpreted similarly, though under separate conditional parallel trends assumptions. The number of observations includes those in the sample between 2009-2011. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6.2 Impact on the Mean Number of Adoptions Across \$5000 Liability Bins

In order to better parse the potentially heterogeneous response to treatment across tax liability, the treatment effects across \$5000 liability bins are estimated. These results, along with the relevant year-over-year placebo ATTs are displayed in Table A7. The treatment effect in 2010 continues to remain insignificant across liability bins. However, the treatment effect for 2011 is significant for those with \$1-\$5,000 of tax liability and those with \$10,001-\$13,360 of tax liability, although the latter group fails its pre-period placebo test. Interestingly, the treatment effect for the \$5,001-\$10,000 group is insignificant in 2011. While this may simply be driven by lower power, this may be indicative of a smaller adoption response given the lower level of treatment.

A treatment effect of 3.67 for the \$1-\$5,000 liability group constitutes a 51 percent increase relative to their pre-treatment average. Since on average this group is exposed to \$10,567 of additional credit relative to the control, this suggests an average responsiveness of 4.8 percent per \$1000 of additional credit. This is a smaller response than that of the positive liability group, which had an average per \$1000 responsiveness of 5.6 percent, which is consistent with the hypothesis that those with more means may have a greater adoption

response due to the high fixed costs associated with adoption.⁴⁶ Scaling the \$1-\$5,000 liability treatment effects by their population suggests this group adopted 13,909 additional children than they would have in the absence of treatment.

6.3 Estimating Average Expenditure Increase to Encourage an Additional Adoption

When the government modified the credit, additional tax expenditures flowed both to individuals who would have claimed the credit in the absence of reform and to individuals induced to perform a new adoption by the reform.⁴⁷ Of particular interest to policymakers is the average fiscal cost of encouraging an additional adoption, given they are not able to distinguish between these two groups. Aggregate tax expenditures on the adoption tax credit across income groups are reported by the IRS yearly. However, these are not reported across tax liability. In order to translate this expenditure data into this context, the expenditures on households across liability bins need to be imputed. To do so, first, consider all households that performed a new adoption within a given year. Each household is then assigned to an income group, the groups being the level at which the aggregate expenditures are reported. Naturally, within an income group, there are households that belong to different tax liability bins. The fractions of households that are within each liability bin are computed for each income group. That fraction of total expenditure is then assigned to each liability group.⁴⁸ This gives a measure of the level of tax expenditures flowing to each liability bin. These results are presented in Table A4.

Only the treatment effect in 2011 is incorporated; however, the average change in expenditures, both in 2010 and 2011 are used, as the change in expenditures in 2010 may have been expenses for an adoption that was finalized in 2011. Assuming the total change in expenditure on each liability group are attributable to reform, which seems to be a valid assumption given expenditures on the credit were relatively level before reform (Brehm, 2021), the average fiscal cost per new adoption was approximately \$15,801 for the zero liability group, \$15,533 for the positive liability group, and \$17,425 for the \$1-\$5,000 liability group. These results imply that it took more funds flowing to the \$1-\$5,000 group to induce

⁴⁶While I am reticent to interpret the \$10,001-\$13,360 group, due to it failing the pre-period placebo test, its somewhat unstable mean, and high treatment effect, the large response I observe is also consistent with this hypothesis.

⁴⁷To use the standard terminology, these are the always takers and the compliers (Imbens and Angrist, 1994).

⁴⁸As an example, suppose the IRS reports \$10,000 of tax expenditure was spent on those households making \$20,000-30,000. If 50 percent of these households have 0 liability and the rest have positive liability, 50 percent of the expenditures is allocated to the 0 liability group and 50 percent on the positive liability group. It's important to remember this is among newly adoptive households, not all households.

the average new adoption, relative to the zero liability group and the higher liability groups. This suggests a non-monotonic per dollar cost per adoption across the distribution of tax liability.

The degree of this heterogeneity is an important parameter for optimal tax policy. If, for example, policymakers wish to maximize the quantity of adoptions they can induce per dollar of expenditure, then this information allows appropriate targeting of incentives across the tax liability distribution. Additionally, given the heterogeneity in the types of adoptions performed by households of different incomes, if policymakers wish to increase transfers to families undertaking particular adoption types (e.g. public adoptions), then this measure provides some information regarding the government’s ability to increase the amount of transfer per child.

6.4 Traditional TWFE and the Dynamic Response

The estimation strategy proposed by Callaway et al. (2024) can be implemented by estimating a series of regressions, decomposing heterogeneous treatment effects across taxpayers using equation (1). However, with additional parametric assumptions, an intuitive event-study design using a single regression can be implemented, where treatment is taken to be every year after the implementation of the policy, operationalized by interacting years after treatment with group indicators. This is analogous to a standard two-way fixed effects approach. Figures A13 and A14 plot these results. The point estimates for 2010 and 2011 do not differ from those estimated using the previous approach, and there is evidence to suggest that treatment positively affected adopting behavior, even after the reform was repealed. There are many potential explanations for this. One suggestive statistic from the National Survey of Adoptive Parents reports that 42% of adoptive families were asked and agreed to help other families with performing an adoption.⁴⁹ This could imply the existence of network effects, meaning that a larger pool of adoptive parents could in itself increase the number of new families who adopt. Another potential mechanism is a larger adoption facilitation industry induced by reform. If some of the benefit from the credit was passed on to private adoption facilitators through increased use or higher fees, this additional revenue could be used to invest in a larger workforce or more advertising, potentially increasing the adoption facilitation industry in aggregate.⁵⁰ Importantly, there isn’t evidence that the treatment effects in the post period are of a form that would invalidate the use of these trends to assess the plausibility of parallel trends, as discussed in section 5.3.

⁴⁹One example of help provided was mentorship of another adoptive family.

⁵⁰The welfare consequences of this are quite uncertain. Far more work needs to be done on the use of these facilitators and the consequence of their operations.

6.5 Impacts on Other Outcomes

In addition to investigating the reform’s impact on the number of new adoptions, other outcomes can also be investigated.

The first additional outcome is the probability that a household conducts a new adoption. Figure A3 plots this average for the control group, the zero liability group, and the positive liability group. Qualitatively, the trends appear roughly parallel in the pre-treatment period and post-repeal periods, with the possible exception of the period between 2011-2012. Figure A7(b) plots both the estimated treatment effects and the various pseudo-ATTs, all of the latter not being statistically significant. The first model of Table A5 reports the treatment effect point estimates for the coarse liability bins. The treatment effect of 0.000297 for the zero liability group is roughly 50 percent of their 2009 pre treatment probability of adoption, implying the reform had a large extensive margin effect.⁵¹ The point estimate for the positive liability group of 0.000248, implies a 33 percent increase relative to their pre-treatment average probability of adoption. Table A8 further decomposes these treatment effects by 5000 liability bins. None of the positive liability bins remain significant except for the \$10,000-\$13,360 liability group, but this fails the pre-period placebo test.

Given the average age of newly adopted children using this measure is quite young⁵², the impact of treatment on the number of adopted children aged less than 2 is calculated. Figure A4 plots this outcome over time for the 0 liability and positive liability treated groups as well as the control group. Here, the trends between 2013 and 2015 appear somewhat divergent. Nonetheless, the year over year trends are not statistically distinguishable, and the break between treated and control groups is quite pronounced during the treatment period. The treatment effects and pseudo ATTs are plotted in Figure A8(a), while the second model of Table A5 displays the point estimates. Again, the treatment effects are positive and significant for both liability groups in 2011. The point estimate of 3.40 for the 0 liability group constitutes a 59 percent increase relative to their 2009 level. For the positive liability group, a point estimate of 3.36 constitutes a 53 percent increase relative to their 2009 level. Scaling both of these estimates suggests the zero liability group adopted 12,036 more babies and the positive liability group adopted 20,496 more babies than they would have in the absence of treatment. This suggests the majority of the measured treatment effect, about 90 percent, is being driven by the adoption of children aged less than 2. This treatment effect, again, may be indicative of my measure of new adoptions primarily picking up non-foster care adoptions. As mentioned previously, 2/3 of private adoptions are of children less than

⁵¹These incredibly small numbers are one of the reasons why large, nationally representative data sets are so important when investigating aggregate policy geared to influencing child adoption.

⁵²As noted previously, this may be indicative of my measure missing some types of public adoptions.

1 year old (Malm and Vandivere, 2009). Table A9 reports treatment effect estimates for the \$5,000 bins as well as their placebos. Here, effects remain significant for all bins in 2011. Absolute treatment effects are declining, but the average responsiveness per dollar of credit exposure rises in tax liability.

7 Robustness Checks

7.1 Trends in International Adoptions

International adoptions have been declining precipitously since 2008 (Khun and Lahiri, 2017)⁵³, and given that richer households are more likely to perform an international adoption, this presents the risk of my control group being differentially exposed to an adoption shock during my treatment period. To account for this, treatment effects on the number of new adoptions flagged as domestic are calculated. Figure A5 shows this plot for the coarser liability bins, as well as the control group. The decline in the treated and control groups pre-treatment is still readily apparent, and each of the groups still qualitatively evolved in parallel after repeal, perhaps even a bit more so. The 3rd model in Table A5 shows the point estimates for these treatment effects, while Figure A8(b) plots these treatment effects and the relevant placebo tests. The estimated treatment effects remain positive and statistically significant for 2011, and are of similar magnitude to the primary treatment effect estimates, indicating that the results are being driven primarily by domestic adoptions, and not the decline in international adoptions. Table A10 reports these estimates across the \$5000 bins. These results are robust to a broader definition of new international adoptions, where adopted children are classified as international if they were born abroad, irrespective of the birthplace of their parents.

7.2 A Measure Even More Robust to Step-Parent Adoptions

Finally, to ensure that the trends in the data are being driven by non-step-parent adoptions, estimates are calculated again using only new adoptions where the partner of the household head is also not living in another household 1 year prior. This definition is used because when undertaking a step-parent adoption, if the child moved in within the last year, it is likely that their biological parent did as well. The ACS records the adoption status of children relative to the household head, and not to both members of the household. If the

⁵³This decline is due to a variety of reasons, including greater restrictions placed by two of the most common countries parents adopted from at the time in 2007. These included China, who put in place additional regulations on who may adopt from them internationally and Guatemala, the adoptions from which peaked in 2007 and then declined after an anti-corruption campaign to reduce the potential for human trafficking.

partner of the household head and the adopted child are both living outside the household 1 year prior, this may suggest the partner is the biological parent of the child and the household head performed a step-parent adoption after their partner and the child moved in. Figure A6 plots the mean evolution of outcomes, no longer counting these children as newly adopted. Figure A8(c) plots the treatment effects excluding these children as well. While some power is lost by dropping these observations, making the estimates less precise, it is clear the trends that compose the identifying assumption are not being driven by these observations and hence, it is unlikely trends in step-parent adoptions are driving the results.

7.3 A Discussion of Substitution Effects

One concern is that some of the treatment effects detected are partially the product of those in the control group losing access to children they themselves would have adopted anyway. Indeed, these substitution patterns could occur between the control group and treated groups and between the treated groups themselves. If the entire decline in the control group were being driven by this substitution, then the control group would not be a valid counterfactual for how the treated groups would have evolved in the absence of treatment. While this possibility can't be ruled out, as it would require knowledge of the counterfactual, there is suggestive evidence that the decline is not being primarily driven by substitution. First, adoptions are declining in the control group at a similar rate before treatment is enacted, and this is not explained by the aggregate decline in international adoptions, as shown in Figure A5. Second, as shown in Table A3, since each of these liability groups adopt different kinds of children, if the trend in the control group were being driven entirely by substitution, this ought to show up in the characteristics of newly adopted children. However, the characteristics of adopted children in the control do not change in a statistically significant way, save for the secular decline in international adoptions.

Other forms of substitution are also possible, but their presence is less of a concern. For example, substitution effects occurring within a treated group will be correctly incorporated into the treatment effects estimated. Substitution effects between liability bins could lead to over or under estimation of their group specific treatment effects, depending on the substitution patterns involved, but these effects would tend to counteract one another.⁵⁴ Of course, when evaluating the welfare and efficiency consequences of this policy, determining whether there is shifting going on is quite important.

⁵⁴This provides another justification for adopting a coarser bin strategy

8 Conclusion

This paper has shed light on the behavioral response of child adoption to financial incentives across the distribution of tax liability, using a reform that took place in 2010-2011 that differentially impacted households across the tax liability distribution. These estimates are particularly relevant today as the credit was recently reformed. Starting in tax year 2025, the first \$5000 of credit will be fully refundable. The estimates presented here imply the reform will result in roughly 4,112 additional adoptions among those with 0 tax liability and roughly 2,183 additional for those with \$1-\$5,000 of tax liability.⁵⁵

There are several other additional considerations beyond the scope of this paper that are important to take note of when evaluating whether the adoption tax credit is a desirable policy and which particular features ought to be kept or modified.

Allowing the credit to be refundable has a large redistributive effect, benefiting poorer adoptive households, as evidenced by the studied reform. Estimates in the optimal tax literature suggest surpluses accruing to the poor ought to be valued at a rate of 1.5-2 times the surplus accruing to the rich, implying the redistributive change during the reform period likely made the policy more socially efficient (Hendren, 2020). Estimates in optimal tax policy targeting families suggests policies that provide transfers to lower income households may still be optimal, even if the government has no redistributive motive through its effect on children (Bart and Horton, 2024). These spillovers have already been documented in the context of foster care, implying increasing transfers to adoptive foster families likely improves outcomes for their adoptive children (Simon et al., 2024).

Along other dimensions, Miller et al. (2023) finds causal evidence that being denied an abortion has a number of long lasting, negative consequences. If some portion of these consequences are driven by the inability to find another home for the child, increasing access to adoption may lead to welfare gains for these birth parents.

Coupled with its effectiveness at actually encouraging adoptions, the weight of the evidence suggests the adoption tax credit is a powerful tool, both enabling families to give children homes and giving them the resources to care for them.

⁵⁵My ATT estimate for these tax payers in response to \$12,660 of relative credit exposure is 0.00038. \$12,660 in 2011 is worth roughly \$18,480.12 today. Scaling the treatment effect downward by the real decline in treatment exposure implies an average treatment effect of 1.028 per 10,000 households. Multiplying this quantity across the size of the 0 liability population gives 4,112. For the \$1-\$5,000 liability group, assume an average exposure of \$2,500. Scaling the treatment effect of 0.000367 similarly by the real value of exposure, \$10,567 in 2011 is roughly \$15,545.40 today, implies a treatment effect of 0.5901 per 10,000 households. Scaling this by the population implies 2,183 additional adoptions.

References

- Akerlof, G. A. (1978). The economics of "tagging" as applied to the optimal income tax, welfare programs, and manpower planning. *The American Economic Review*, 68(1), 8–19. Retrieved July 23, 2025, from <http://www.jstor.org/stable/1809683>
- Baccara, M., Collard-Wexler, A., Felli, L., & Yariv, L. (2014). Child-Adoption Matching: Preferences for Gender and Race. *American Economic Journal: Applied Economics*, 6(3), 133–158. <https://doi.org/10.1257/app.6.3.133>
- Bald, A., Doyle, J. J., Gross, M., & Jacob, B. A. (2022). Economics of foster care. *Journal of Economic Perspectives*, 36(2), 223–246. <https://doi.org/10.1257/jep.36.2.223>
- Baron, E. J., & Gross, M. (2022, April). *Is There a Foster Care-To-Prison Pipeline? Evidence from Quasi-Randomly Assigned Investigators* (tech. rep. No. w29922). National Bureau of Economic Research. Cambridge, MA. <https://doi.org/10.3386/w29922>
- Bart, L., & Horton, E. (2024). Optimal transfers with children's utility. *wp*. https://ljbart.github.io/lea-bart.com/Optimal_Transfers.pdf
- Bogadi, P. (2012). Has Significantly and Unnecessarily Harmed Vulnerable Taxpayers, the IRS's Compliance Strategy for the Expanded Adoption Credit Has Increased Costs for the IRS, and Does Not Bode Well for Future Credit Administration. *National Taxpayer Advocate*, 2, 111–133.
- Brehm, M. E. (2021). Taxes and adoptions from foster care: Evidence from the federal adoption tax credit. *Journal of Human Resources*, 56(4), 1031–1072. <https://doi.org/10.3368/jhr.56.4.0618-9539R1>
- Buckles, K. (2013). Adoption subsidies and placement outcomes for children in foster care. *The Journal of Human Resources*, 48(3), 596–627.
- Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024, February). *Difference-in-differences with a continuous treatment* (w32117). National Bureau of Economic Research. Cambridge, MA. <https://doi.org/10.3386/w32117>
- Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230. <https://doi.org/10.1016/j.jeconom.2020.12.001>
- Deb, P., Norton, E., Wooldridge, J., & Zabel, J. (2024, October). *A flexible, heterogeneous treatment effects difference-in-differences estimator for repeated cross-sections* (w33026). National Bureau of Economic Research. Cambridge, MA. <https://doi.org/10.3386/w33026>
- Doyle, J. J., & Peters, H. E. (2007). The market for foster care: An empirical study of the impact of foster care subsidies. *Review of Economics of the Household*, 5(4), 329. <https://doi.org/10.1007/s11150-007-9018-x>
- Feenberg, D. R., & Coutts, E. (2022). *Internet TAXSIM version 35 - MyNBER*. Retrieved September 6, 2025, from <https://taxsim.nber.org/taxsim35/>
- Hansen, M. E. (2007). Using subsidies to promote the adoption of children from foster care. *Journal of Family and Economic Issues*, 28(3), 377–393. <https://doi.org/10.1007/s10834-007-9067-6>
- Hansen, M. E., & Hansen, B. A. (2006). The economics of adoption of children from foster care. *Child Welfare*, 85(3), 559–83.

- Hawkins, A., Hollrah, C., Miller, S., Wherry, L., Aldana, G., & Wong, M. (2023, September). *The long-term effects of income for at-risk infants: Evidence from supplemental security income* (w31746). National Bureau of Economic Research. Cambridge, MA. <https://doi.org/10.3386/w31746>
- Hendren, N. (2020). Measuring economic efficiency using inverse-optimum weights. *Journal of Public Economics*, 187, 104198. <https://doi.org/10.1016/j.jpubeco.2020.104198>
- Hendren, N., & Sprung-Keyser, B. (2020). A unified welfare analysis of government policies*. *The Quarterly Journal of Economics*, 135(3), 1209–1318. <https://doi.org/10.1093/qje/qjaa006>
- Herman, E. (2012). *The adoption history project*. Retrieved October 9, 2025, from <https://pages.uoregon.edu/adoption/>
- <https://taxpolicycenter.org/briefing-book/what-standard-deduction> [Tax policy center]. (2025). <https://taxpolicycenter.org/briefing-book/what-standard-deduction>
- Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2), 467. <https://doi.org/10.2307/2951620>
- IPUMS USA. (n.d.). Retrieved September 4, 2025, from <https://usa.ipums.org/usa/voliii/ACSSamp.shtml>
- Johnston, W. R. (2022, November 12). *Historical statistics on adoption in the united states, plus statistics on child population and welfare* [Johnstonsarchive]. <https://www.johnstonsarchive.net/policy/adoptionstats.html>
- Karim, S., & Webb, M. D. (2024). Good controls gone bad: Difference-in-differences with covariates. <https://doi.org/10.48550/ARXIV.2412.14447>
- Keen, M., & Slemrod, J. (2017). Optimal tax administration. *Journal of Public Economics*, 152, 133–142. <https://doi.org/10.1016/j.jpubeco.2017.04.006>
- Khun, C., & Lahiri, S. (2017). The economics of international child adoption: An analysis of adoptions by u.s. parents. *The Quarterly Review of Economics and Finance*, 64, 22–31. <https://doi.org/10.1016/j.qref.2016.07.001>
- Khun, C., Lahiri, S., & Lim, S. (2020). WHY DO u.s. PARENTS PREFER PRIVATE TO FOSTER CARE ADOPTIONS? THE ROLE OF ADOPTION SUBSIDIES, GENDER, RACE, AND SPECIAL NEEDS. *Economic Inquiry*. https://collected.jcu.edu/fac_bib.2020/47
- Kurnaz, M. (2018). Optimal taxation of families: Mirrlees meets becker. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3287232>
- Malm, K., & Vandivere, S. (2009, October 31). *Adoption USA. a chartbook based on the 2007 national survey of adoptive parents* [ASPE]. Retrieved July 6, 2025, from <http://aspe.hhs.gov/reports/adoption-usa-chartbook-based-2007-national-survey-adoptive-parents-0>
- Miller, S., Wherry, L. R., & Foster, D. G. (2023). The economic consequences of being denied an abortion. *American Economic Journal: Economic Policy*, 15(1), 394–437. <https://doi.org/10.1257/pol.20210159>
- Moriguchi, C. (2012, June). *The evolution of child adoption in the united states, 1950-2010: An economic analysis of historical trends* (Discussion Paper Series No. 572). Institute of Economic Research, Hitotsubashi University. <https://ideas.repec.org/p/hit/hituec/572.html>

- Negi, A., & Negi, D. S. (2025). Difference-in-differences with a misclassified treatment. *Journal of Applied Econometrics*, 40(4), 411–423. <https://doi.org/10.1002/jae.3116>
- Potter, M. H., & Font, S. A. (2021). State contexts and foster care adoption rates. *Children and Youth Services Review*, 126, 106049. <https://doi.org/10.1016/j.childyouth.2021.106049>
- Saez, E. (2010). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy*, 2(3), 180–212. <https://doi.org/10.1257/pol.2.3.180>
- Simon, D., Sojourner, A., Pedersen, J., & Skallet, H. O. (2024, June). *Financial incentives for adoption and kin guardianship improve achievement for foster children* (w32560). National Bureau of Economic Research. Cambridge, MA. <https://doi.org/10.3386/w32560>
- Sisson, G. (2022). Estimating the annual domestic adoption rate and lifetime incidence of infant relinquishment in the united states, *Contraception*, 105, 14–18. <https://doi.org/10.1016/j.contraception.2021.08.008>
- SOI tax stats - individual income tax returns filed and sources of income — internal revenue service.* (n.d.). Retrieved July 6, 2025, from <https://www.irs.gov/statistics/soi-tax-stats-individual-statistical-tables-by-filing-status>
- SOI tax stats - individual statistical tables by size of adjusted gross income — internal revenue service.* (n.d.). Retrieved July 6, 2025, from <https://www.irs.gov/statistics/soi-tax-stats-individual-statistical-tables-by-size-of-adjusted-gross-income>
- Solon, G., Haider, S. J., & Wooldridge, J. (2013, February). What are we weighting for? <https://doi.org/10.3386/w18859>
- US Adoption Statistics — Adoption Network. (2020, October). Retrieved February 17, 2025, from <https://adoptionnetwork.com/adoption-myths-facts/domestic-us-statistics/>

A Adoption Imputation Procedure

Consider all children flagged as adopted by the ACS who are less than age 18. A child is flagged as an international adoption if both parents are born domestically, and their child is born internationally. The rest are flagged as domestic adoptions.

Next, to classify whether an adoption is “new”, two procedures are used depending on if the child is flagged as international or domestic. If the child is international, a child is flagged as newly adopted if their year of immigration is the same as the year they are observed.⁵⁶ If the child is domestic, a combination of their age and migration status relative to the head of household is used. If the child is young enough less than or equal to age 1 in the ACS, they are flagged as newly adopted.⁵⁷ For all other children less than this age, the MIGRATE variable is used. If an adopted child’s MIGRATE variable indicates that they lived outside the household the year prior and the household head did not move, they are flagged as newly adopted.

Finally, in order to create a measure more robust to the inclusion of step-parent adoptions, children are grouped into the category “unlikely to be a step-parent adoption”. If both the child and the household head’s spouse lived in a different household the year prior, they are classified as a possible step-parent adoption and not included in this category. Although many step-parent adoptions probably occur while the child is living with the household head for a longer period of time, this provides a means of assessing whether changes in step-parent adoptions are driving the adoption trends in the data.

⁵⁶An alternative measure using the migration status and relative immigration status of the child alone produces similar results.

⁵⁷As explained in detail by Baccara et. al. (2014), it is possible for a parent to be the expected adopted parent before the child is born. Of course, the adoption itself can only be finalized after birth, and depending on the state, the birth mother may be able to change her mind.

B Descriptive Tables

Table A1: Main Summary Statistics

PANEL A: Mean Annual Household Income (\$1,000s)					
		<i>2008-2009</i>	<i>2010-2011</i>	<i>2012-2013</i>	<i>2014-2015</i>
<i>New Adopters</i>	Control	188.23 (4.73)	189.04 (5.05)	204.25 (7.40)	221.29 (8.48)
	0 Liab	24.31 (1.52)	24.21 (1.23)	23.53 (1.30)	26.24 (4.01)
	1-5000 Liab	52.51 (1.33)	54.68 (1.43)	56.15 (1.49)	54.83 (1.53)
	5001-10,000	81.55 (1.03)	87.13 (1.53)	90.42 (3.25)	86.35 (1.22)
	Positive Liab	65.60 (1.09)	71.48 (1.28)	71.62 (1.67)	74.25 (2.29)
<i>Non-Adopters</i>	Control	171.92 (0.18)	182.26 (0.19)	186.94 (0.20)	198.21 (0.21)
	0 Liab	22.36 (0.03)	23.27 (0.03)	23.89 (0.03)	25.25 (0.04)
	1-5000 Liab	44.44 (0.03)	45.58 (0.03)	46.71 (0.04)	48.60 (0.04)
	5001-10,000	75.61 (0.04)	77.57 (0.05)	79.41 (0.05)	81.84 (0.05)
	Positive Liab	56.81 (0.03)	59.79 (0.03)	60.82 (0.04)	63.89 (0.04)
PANEL B: Mean Age of Household Head					
		<i>2008-2009</i>	<i>2010-2011</i>	<i>2012-2013</i>	<i>2014-2015</i>
<i>New Adopters</i>	Control	40.633 (0.326)	41.215 (0.456)	40.874 (0.471)	41.114 (0.541)
	0 Liab	33.255 (0.644)	33.879 (0.609)	34.228 (0.743)	35.739 (0.922)
	1-5000 Liab	36.417 (0.424)	36.268 (0.697)	35.719 (0.51)	35.632 (0.567)
	5001-10,000	38.607 (0.468)	37.564 (0.499)	38.216 (0.559)	37.121 (0.658)
	Positive Liab	37.409 (0.305)	37.073 (0.429)	36.697 (0.37)	36.647 (0.407)
<i>Non-Adopters</i>	Control	49.400 (0.0205)	50.088 (0.0239)	50.433 (0.0233)	50.801 (0.023)
	0 Liab	56.891 (0.0322)	56.467 (0.0296)	57.059 (0.03)	57.608 (0.0301)
	1-5000 Liab	47.152 (0.0223)	47.868 (0.0232)	48.148 (0.0239)	48.454 (0.0244)
	5001-10,000	47.430 (0.026)	48.193 (0.028)	48.776 (0.0287)	49.052 (0.0289)
	Positive Liab	47.276 (0.0166)	48.008 (0.017)	48.406 (0.0176)	48.732 (0.0177)
PANEL C: Number of Same-Sex Couples per 1,000 Households					
		<i>2008-2009</i>	<i>2010-2011</i>	<i>2012-2013</i>	<i>2014-2015</i>
<i>New Adopters</i>	Control	36.7 (8.59)	26.3 (8.9)	82.4 (15.2)	118 (23.7)
	0 Liab	5.15 (2.98)	12.6 (4.94)	19.6 (6.61)	24.6 (12.2)
	1-5000 Liab	32.5 (8.03)	22.8 (7.01)	46 (12.3)	53.2 (16.1)
	5001-10,000	21.3 (7.48)	55.3 (16.5)	46.8 (13)	33.3 (10.9)
	Positive Liab	29.6 (6.03)	33.2 (7.05)	48.8 (9.43)	40.5 (9.38)
<i>Non-Adopters</i>	Control	6.59 (0.131)	6.57 (0.145)	8.25 (0.158)	11 (0.178)
	0 Liab	3.87 (0.0908)	3.91 (0.089)	4.23 (0.0911)	4.37 (0.0919)
	1-5000 Liab	5.09 (0.0916)	5.36 (0.098)	5.91 (0.105)	6.59 (0.113)
	5001-10,000	4.83 (0.123)	4.96 (0.128)	5.62 (0.138)	7.32 (0.159)
	Positive Liab	5.12 (0.0722)	5.38 (0.0755)	5.97 (0.0811)	6.99 (0.0879)

Note: Taken from the American Community Survey 1% samples. All estimates are weighted. Standard errors of the means are reported in parentheses.

Table A2: Summary Statistics

PANEL A: Share of Household Heads who are White					
		<i>2008-2009</i>	<i>2010-2011</i>	<i>2012-2013</i>	<i>2014-2015</i>
<i>New Adopters</i>	Control	0.873 (0.0171)	0.87 (0.0183)	0.859 (0.0262)	0.876 (0.0243)
	0 Liab	0.562 (0.0328)	0.596 (0.0344)	0.626 (0.0346)	0.691 (0.0369)
	1-5000 Liab	0.772 (0.0224)	0.743 (0.0306)	0.814 (0.0244)	0.756 (0.0332)
	5001-10,000	0.827 (0.0255)	0.854 (0.026)	0.871 (0.0273)	0.911 (0.0236)
	Positive Liab	0.802 (0.0161)	0.796 (0.0199)	0.839 (0.0171)	0.823 (0.0204)
<i>Non-Adopters</i>	Control	0.871 (0.000596)	0.865 (0.000715)	0.86 (0.00069)	0.855 (0.00068)
	0 Liab	0.745 (0.000692)	0.744 (0.000653)	0.74 (0.000668)	0.735 (0.000679)
	1-5000 Liab	0.776 (0.000605)	0.778 (0.00063)	0.774 (0.000647)	0.765 (0.000668)
	5001-10,000	0.84 (0.000745)	0.84 (0.0008)	0.836 (0.000812)	0.83 (0.000815)
	Positive Liab	0.8 (0.000456)	0.803 (0.000469)	0.798 (0.000484)	0.792 (0.00049)
PANEL B: Share of Household Heads who are Married					
		<i>2008-2009</i>	<i>2010-2011</i>	<i>2012-2013</i>	<i>2014-2015</i>
<i>New Adopters</i>	Control	0.901 (0.0134)	0.925 (0.0141)	0.901 (0.0239)	0.914 (0.0247)
	0 Liab	0.482 (0.0339)	0.462 (0.0343)	0.429 (0.0345)	0.438 (0.0401)
	1-5000 Liab	0.696 (0.0244)	0.662 (0.0315)	0.734 (0.0308)	0.684 (0.0341)
	5001-10,000	0.893 (0.0215)	0.879 (0.0234)	0.871 (0.0242)	0.914 (0.0214)
	Positive Liab	0.774 (0.0171)	0.757 (0.0209)	0.794 (0.0209)	0.787 (0.0208)
<i>Non-Adopters</i>	Control	0.76 (0.000748)	0.769 (0.000853)	0.758 (0.000831)	0.76 (0.000803)
	0 Liab	0.338 (0.000708)	0.345 (0.000682)	0.34 (0.000688)	0.336 (0.000688)
	1-5000 Liab	0.43 (0.000679)	0.433 (0.000716)	0.421 (0.000724)	0.411 (0.00073)
	5001-10,000	0.658 (0.000927)	0.661 (0.00099)	0.655 (0.000999)	0.654 (0.000984)
	Positive Liab	0.511 (0.000543)	0.518 (0.000565)	0.509 (0.000575)	0.505 (0.000571)
PANEL C: Mean Number of Children per Household					
		<i>2008-2009</i>	<i>2010-2011</i>	<i>2012-2013</i>	<i>2014-2015</i>
<i>New Adopters</i>	Control	1.970 (0.0601)	2.034 (0.0753)	1.939 (0.0777)	1.925 (0.069)
	0 Liab	2.760 (0.09)	2.819 (0.105)	2.731 (0.122)	2.783 (0.121)
	1-5000 Liab	2.308 (0.0674)	2.287 (0.0758)	2.197 (0.0808)	2.361 (0.0974)
	5001-10,000	1.987 (0.0711)	2.224 (0.101)	2.034 (0.0926)	1.970 (0.099)
	Positive Liab	2.185 (0.0489)	2.233 (0.0568)	2.156 (0.0591)	2.175 (0.065)
<i>Non-Adopters</i>	Control	0.816 (0.00179)	0.825 (0.00212)	0.812 (0.00201)	0.806 (0.00195)
	0 Liab	0.725 (0.00193)	0.738 (0.00183)	0.726 (0.00184)	0.704 (0.00184)
	1-5000 Liab	0.724 (0.00152)	0.714 (0.00159)	0.701 (0.00162)	0.695 (0.00163)
	5001-10,000	0.771 (0.00206)	0.766 (0.00223)	0.751 (0.00223)	0.748 (0.0022)
	Positive Liab	0.74 (0.00118)	0.728 (0.00122)	0.718 (0.00124)	0.709 (0.00123)

Note: Taken from the American Community Survey 1% samples. All estimates are weighted. Standard errors of the means are reported in parentheses.

Table A3: Characteristics of Newly Adopted Children Before and During Reform

	2008-2009	2010-2011	2012-2013	2014-2015
Average Age				
Total	2.199 (0.109)	2.381 (0.134)	2.365 (0.148)	2.411 (0.15)
Control	2.060 (0.165)	2.588 (0.281)	2.372 (0.292)	2.276 (0.278)
0 Liab	1.685 (0.21)	1.996 (0.251)	1.927 (0.228)	2.723 (0.36)
1-5000	2.484 (0.242)	2.818 (0.288)	2.394 (0.29)	2.333 (0.294)
5001-10,000	2.526 (0.262)	1.789 (0.213)	3.101 (0.422)	2.348 (0.315)
Positive Liab	2.518 (0.173)	2.495 (0.191)	2.575 (0.224)	2.346 (0.201)
Fraction International				
Total	0.113 (0.00794)	0.0721 (0.00818)	0.0575 (0.00725)	0.0441 (0.00677)
Control	0.217 (0.0183)	0.153 (0.023)	0.0964 (0.0191)	0.0897 (0.0172)
0 Liab	0.016 (0.007)	0.0113 (0.0066)	0.0103 (0.00537)	0.0126 (0.00962)
1-5000	0.0555 (0.0115)	0.0516 (0.0142)	0.0451 (0.012)	0.0311 (0.0134)
5001-10,000	0.138 (0.0209)	0.089 (0.0224)	0.0805 (0.0192)	0.0303 (0.0118)
Positive Liab	0.0928 (0.0109)	0.0673 (0.0113)	0.0625 (0.0105)	0.0339 (0.00869)
Fraction White				
Total	0.646 (0.0133)	0.656 (0.0153)	0.707 (0.0159)	0.705 (0.0167)
Control	0.674 (0.0216)	0.671 (0.0284)	0.73 (0.0314)	0.725 (0.0309)
0 Liab	0.541 (0.0346)	0.557 (0.0337)	0.617 (0.0354)	0.667 (0.0381)
1-5000	0.675 (0.0243)	0.669 (0.0299)	0.744 (0.0284)	0.699 (0.0337)
5001-10,000	0.663 (0.0298)	0.753 (0.0294)	0.726 (0.0359)	0.739 (0.035)
Positive Liab	0.676 (0.018)	0.697 (0.0207)	0.743 (0.0211)	0.712 (0.023)
Fraction Black				
Total	0.209 (0.0119)	0.224 (0.0134)	0.227 (0.0153)	0.235 (0.0157)
Control	0.138 (0.0161)	0.153 (0.0192)	0.178 (0.0285)	0.213 (0.0275)
0 Liab	0.328 (0.035)	0.32 (0.0312)	0.285 (0.034)	0.244 (0.0352)
1-5000	0.194 (0.0193)	0.251 (0.0288)	0.192 (0.0253)	0.244 (0.0328)
5001-10,000	0.207 (0.0254)	0.167 (0.0247)	0.244 (0.0354)	0.251 (0.0351)
Positive Liab	0.199 (0.0148)	0.207 (0.0186)	0.216 (0.0208)	0.244 (0.0225)
Fraction Female				
Total	0.537 (0.0131)	0.524 (0.0154)	0.522 (0.0162)	0.504 (0.0178)
Control	0.51 (0.0217)	0.545 (0.0284)	0.543 (0.0304)	0.496 (0.0337)
0 Liab	0.548 (0.0342)	0.527 (0.0332)	0.545 (0.0356)	0.514 (0.0385)
1-5000	0.56 (0.0233)	0.528 (0.0299)	0.502 (0.0307)	0.522 (0.0369)
5001-10,000	0.516 (0.0302)	0.478 (0.0359)	0.536 (0.035)	0.471 (0.0383)
Positive Liab	0.548 (0.0178)	0.515 (0.0217)	0.504 (0.0223)	0.504 (0.0251)

Note: Taken from the American Community Survey 1% samples. All estimates are weighted. Standard errors of the means are reported in parentheses.

Table A4: Imputed Total Expenditures Flowing to Liability Groups Over Time

	2008-2009	2010-2011	Difference
Actual Yearly Average Total Tax Expenditure (\$1000s)	315972	908905	592933
Imputed Yearly Average Tax Expenditure (\$1000s) on:			
Control Group	134565	162184	27619
0 Liability Group	1308	216889	215581
1-5000 Liability Group	69596	311973	242377
5001-10,000 Liability Group	88764	164311	75547
10,001-13,360 Liability Group	21194	53472	32279
Treated Positive Liability Group	179554	529756	350202

Note: For the actual totals I use the Internal Revenue Service, Statistics of Income, Individual Income Tax Returns, Publication 1304, Table 3.3, Tax Years 2008-2011. To impute the share of tax expenditures flowing to each liability group, I use the disaggregated expenditures that are broken down by AGI bins provided by the above publication. I then look at the fraction of new adopters within each liability bin that compose each AGI bin. I then assign each liability bin that share of the aggregate AGI level expenditure.

C Results

Table A5: Treatment Effects for Coarse Liability Bins: Probability of Adoption, Infant Adoption, Domestic Adoption

	(1) Prob New Adoption	(2) Num New Infants	(3) Num New Domestic
ATT 0 Liab 2010	0.000231 (0.000148)	1.55 (1.53)	1.43 (1.69)
ATT Pos Liab 2010	0.000247* (0.000135)	2.13 (1.36)	1.94 (1.54)
ATT 0 Liab 2011	0.000297** (0.000151)	3.40** (1.44)	3.25** (1.60)
ATT Pos Liab 2011	0.000248* (0.000141)	3.36*** (1.28)	3.09** (1.47)
Observations	3,600,610	3,600,610	3,600,610

Note: The table displays the dynamic results for the regression design with coarse bins for several additional outcomes: the probability of performing a new adoption, the number of newly adopted children less than 2, and the number of newly adopted children not flagged as being an international adoptions. The coefficients in columns (2) and (3) are transformed to represent the treatment effect per 10,000 households. Under unconditional parallel trends, these may be interpreted as the average treatment effect on the treated, where treatment is conceptualized as the amount of additional credit the group gained access, though not actually used. The number of observations includes those observed between 2009-2011.

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: Treatment Effects for Coarse Liability Bins Pooled Across Treatment Years 2010-2011

	(1) Num New Adoption	(2) Prob New Adoption	(3) Num New Infants	(4) Num New Domestic
ATT 0 Liab Pooled	2.942** (1.47)	0.000234** (0.000110)	1.81* (1.06)	1.84 (1.18)
ATT Pos Liab Pooled	3.083** (1.346)	0.000253** (0.0000993)	2.47*** (0.932)	2.27** (1.07)
Observations	3,600,610	3,600,610	3,600,610	3,600,610

Note: The table displays the regression results when pooling samples together across treatment years for each of my outcomes of interest. The coefficients in columns (1), (3), and (4) are transformed to represent the treatment effect per 10,000 households. By pooling across time, these estimates provide a summary for the average treatment effect on the treated over the whole reform period. The previous analysis suggest, however, that these effect are likely being driven by the adoption response taking place in 2011. The number of observations includes those in the sample between 2009-2011. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A7: Treatment Effects and Placebos 5000 Bins: Number of Newly Adopted Kids Aged < 18

	(1) Placebo ATT 2009	(2) ATT 2010	(3) ATT 2011	(4) Placebo ATT 2012	(5) Placebo ATT 2013	(6) Placebo ATT 2014
0 Liab	-1.22 (1.77)	2.08 (1.74)	3.76** (1.64)	1.05 (1.49)	-0.972 (1.46)	0.398 (1.45)
1-5000 Liab	-0.962 (1.64)	2.14 (1.65)	3.67** (1.61)	2.18 (1.58)	-1.56 (1.54)	0.420 (1.48)
5001-10000 Liab	0.926 (1.87)	2.65 (1.99)	2.73 (1.92)	-0.143 (1.80)	0.530 (1.70)	0.686 (1.68)
10001-13360 Liab	-5.76* (3.13)	4.13 (2.52)	6.88** (2.89)	1.27 (3.52)	-2.63 (3.19)	-0.764 (2.30)
Observations	2,377,354	2,395,780	2,396,833	2,412,555	2,418,989	2,429,504

Note: These results display the average treatment effect among the treated, given they received their average level of treatment relative to the control. The coefficients are transformed to represent the treatment effect per 10,000 households. The placebo ATTs in 2009 and years after 2011 are year over year estimates, assuming treatment took place in the given year, relative to the year prior. Insignificant results imply I fail to reject the hypothesis that the treated group is experiencing a differential trend relative to the control group in that given year. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Treatment Effects and Placebos 5000 Bins: Probability of Performing New Adoption

	(1) Placebo ATT 2009	(2) ATT 2010	(3) ATT 2011	(4) Placebo ATT 2012	(5) Placebo ATT 2013	(6) Placebo ATT 2014
0 Liab	-0.0000775 (0.000159)	0.000231 (0.000148)	0.000297** (0.000151)	0.000160 (0.000140)	-0.000122 (0.000132)	0.0000413 (0.000134)
1-5000 Liab	-0.00000944 (0.000147)	0.000210 (0.000141)	0.000235 (0.000149)	0.000283* (0.000144)	-0.000196 (0.000135)	0.0000857 (0.000135)
5001-10000 Liab	0.000128 (0.000168)	0.000273 (0.000173)	0.000202 (0.000178)	0.0000282 (0.000166)	0.0000167 (0.000149)	0.0000825 (0.000154)
10001-13360 Liab	-0.000533* (0.000285)	0.000432** (0.000219)	0.000498** (0.000241)	0.000221 (0.000287)	-0.000246 (0.000266)	-0.0000329 (0.000216)
Observations	2,377,354	2,395,780	2,396,833	2,412,555	2,418,989	2,429,504

Note: These results display the average treatment effect among the treated, given they received their average level of treatment relative to the control. The placebo ATTs in 2009 and years after 2011 are year over year estimates, assuming treatment took place in the given year, relative to the year prior. Insignificant results imply I fail to reject the hypothesis that the treated group is experiencing a differential trend relative to the control group in that given year. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Treatment Effects and Placebos 5000 Bins: Number of Newly Adopted Kids Aged < 2

	(1) Placebo ATT 2009	(2) ATT 2010	(3) ATT 2011	(4) Placebo ATT 2012	(5) Placebo ATT 2013	(6) Placebo ATT 2014
0 Liab	-1.47 (1.58)	1.55 (1.53)	3.40** (1.44)	-0.0663 (1.29)	-0.171 (1.27)	-0.975 (1.28)
1-5000 Liab	-0.674 (1.39)	1.62 (1.41)	3.18** (1.37)	1.35 (1.35)	-0.921 (1.32)	-0.114 (1.32)
5001-10000 Liab	0.422 (1.61)	3.15* (1.75)	2.96* (1.62)	-1.81 (1.45)	1.57 (1.35)	0.0450 (1.47)
10001-13360 Liab	-5.02* (2.60)	2.59 (1.92)	6.04** (2.38)	1.09 (3.17)	-2.34 (2.86)	-1.16 (1.98)
Observations	2,377,354	2,395,780	2,396,833	2,412,555	2,418,989	2,429,504

Note: These results display the average treatment effect among the treated, given they received their average level of treatment relative to the control. The coefficients are transformed to represent the treatment effect per 10,000 households. The placebo ATTs in 2009 and years after 2011 are year over year estimates, assuming treatment took place in the given year, relative to the year prior. Insignificant results imply I fail to reject the hypothesis that the treated group is experiencing a differential trend relative to the control group in that given year. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

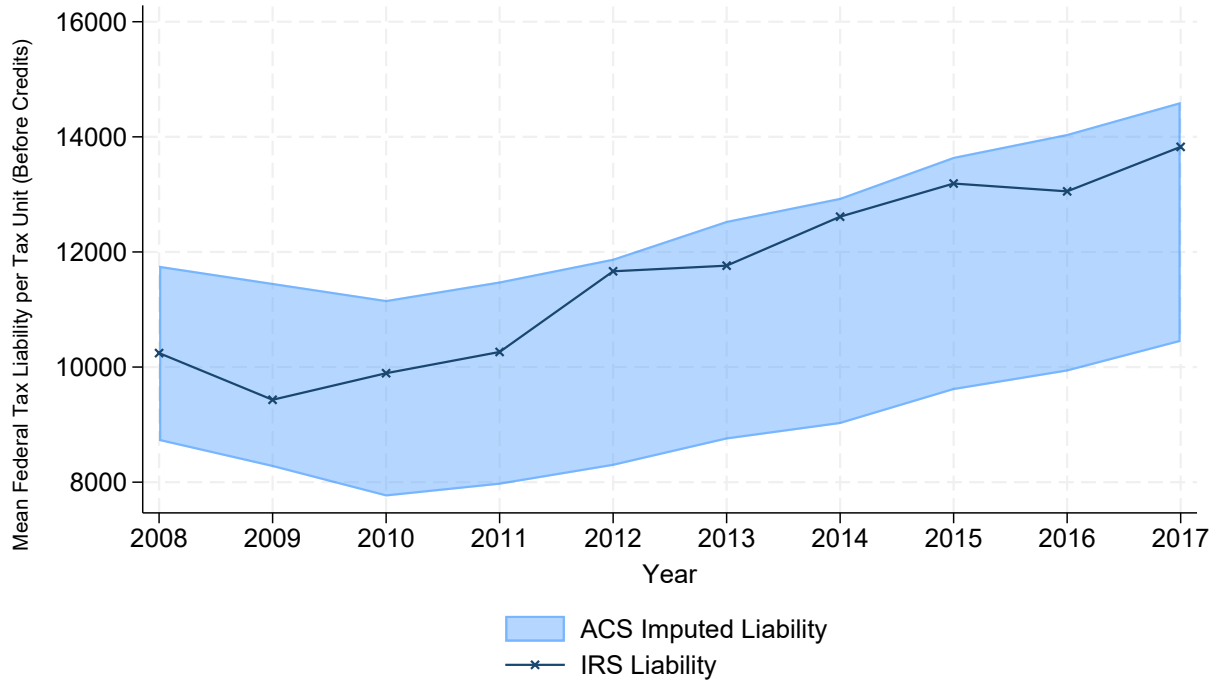
Table A10: Treatment Effects and Placebos 5000 Bins: Number of Domestic Adoptions

	(1) Placebo ATT 2009	(2) ATT 2010	(3) ATT 2011	(4) Placebo ATT 2012	(5) Placebo ATT 2013	(6) Placebo ATT 2014
0 Liab	-1.16 (1.72)	1.43 (1.69)	3.25** (1.60)	0.565 (1.47)	-1.22 (1.46)	0.445 (1.44)
1-5000 Liab	-0.891 (1.58)	1.52 (1.61)	3.05** (1.55)	1.73 (1.56)	-1.73 (1.54)	0.556 (1.47)
5001-10000 Liab	1.19 (1.81)	2.31 (1.94)	2.04 (1.83)	-0.486 (1.74)	0.635 (1.68)	0.806 (1.67)
10001-13360 Liab	-6.13** (2.94)	4.09* (2.33)	6.94** (2.72)	0.849 (3.55)	-3.08 (3.19)	-0.357 (2.24)
Observations	2,377,354	2,395,780	2,396,833	2,412,555	2,418,989	2,429,504

Note: These results display the average treatment effect among the treated, given they received their average level of treatment relative to the control. The coefficients are transformed to represent the treatment effect per 10,000 households. The placebo ATTs in 2009 and years after 2011 are year over year estimates, assuming treatment took place in the given year, relative to the year prior. Insignificant results imply I fail to reject the hypothesis that the treated group is experiencing a differential trend relative to the control group in that given year. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

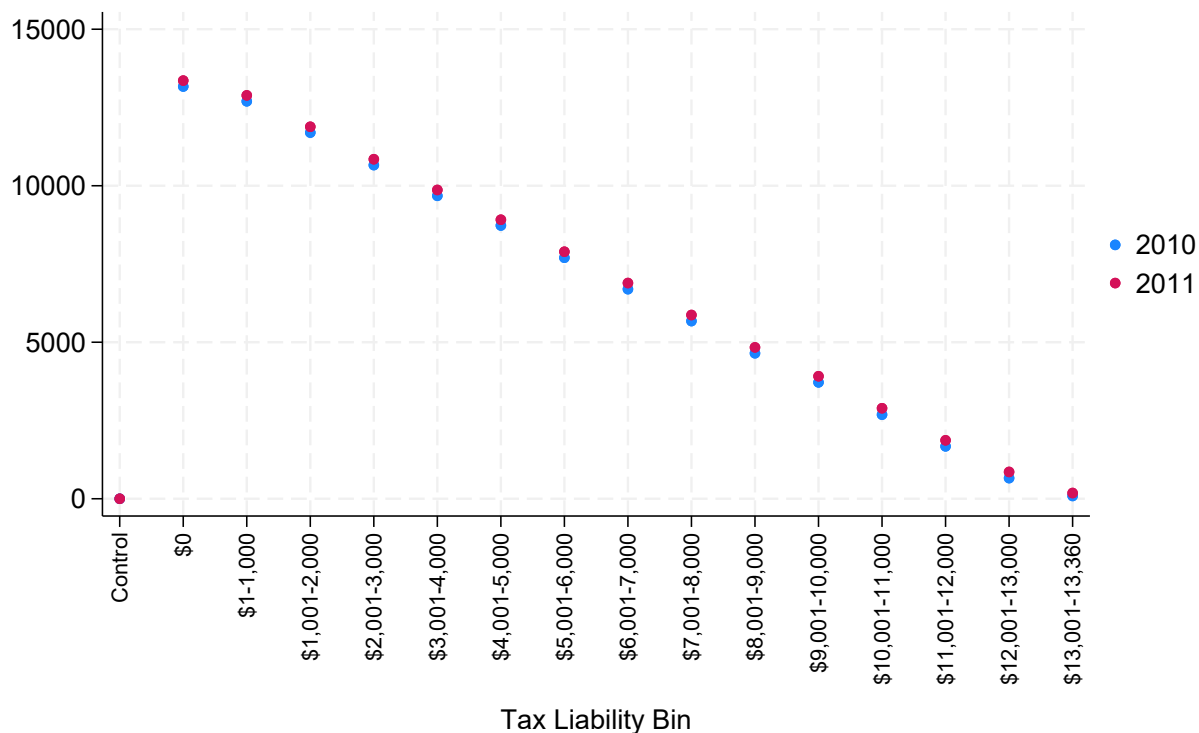
D Figures

Figure A1: Imputed Liability Against Aggregate IRS Liability



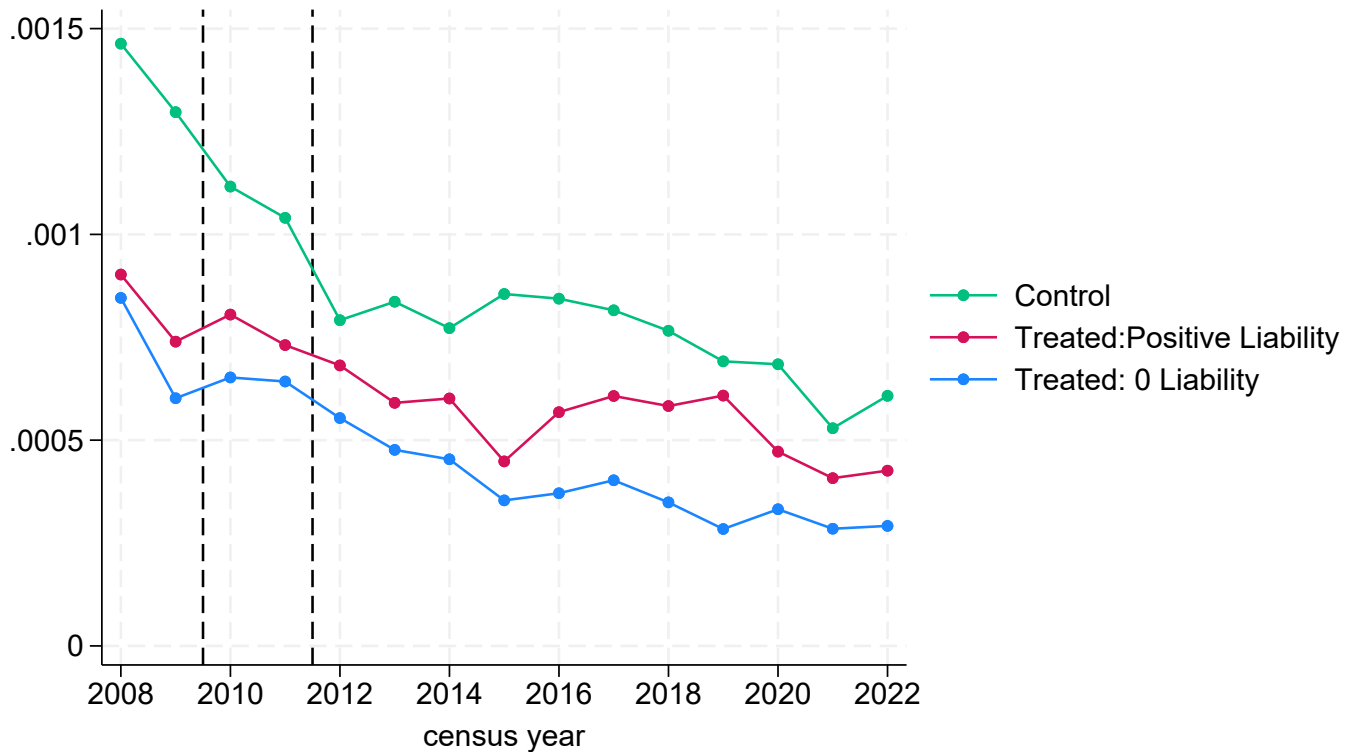
Note: The black line plots the average level of income tax liability before credits and comes from the Internal Revenue Service's Statistics of Income, Individual Income Tax Returns Table 1 for tax years 2008-2017. The top blue line is the average level of imputed federal tax liability before credits per household among all households I impute as having positive liability. The bottom blue line is the average level of imputed federal tax liability before credits for all households. Since households with very low levels of income are not required to file, these measures ought to serve as upper and lower bounds for the true mean level of federal tax liability before credits respectively. Replacing the number of returns with the imputed number of households using the aggregate IRS data does not qualitatively change the graph.

Figure A2: Imputed Mean Increase in Credit Access by Liability Bin Due to the Credit Becoming Refundable



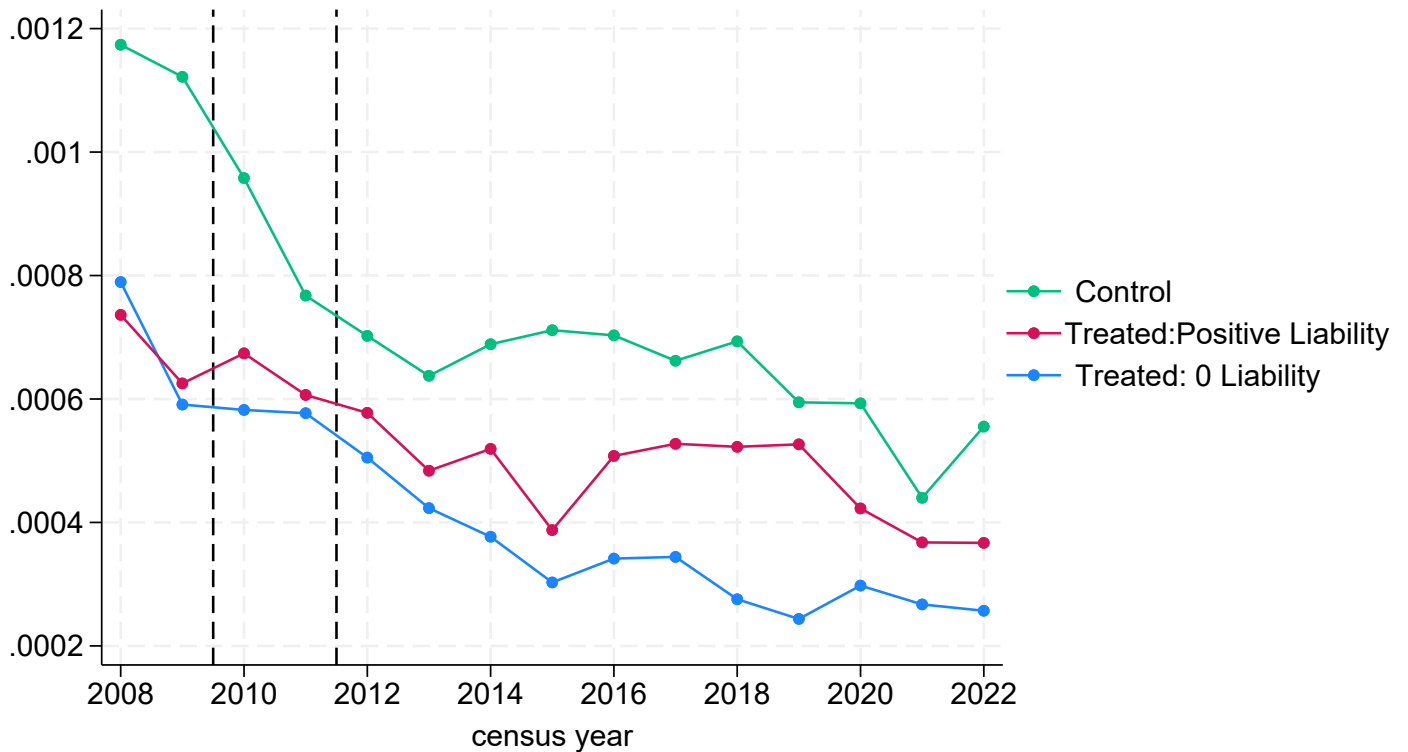
Note: This figure reports the mean increase in credit access due only to the refundability reform across the distribution of tax liability and by treatment year. The control group experienced no increase due to the refundability reform, only experiencing at maximum a \$1,000 increase due to the change in the maximum credit amount, which is not presented in this figure. For treated groups, the amount of treatment decreased in tax liability, with those facing zero tax liability experiencing the largest treatment dose.

Figure A3: Probability of Undertaking a New Adoption



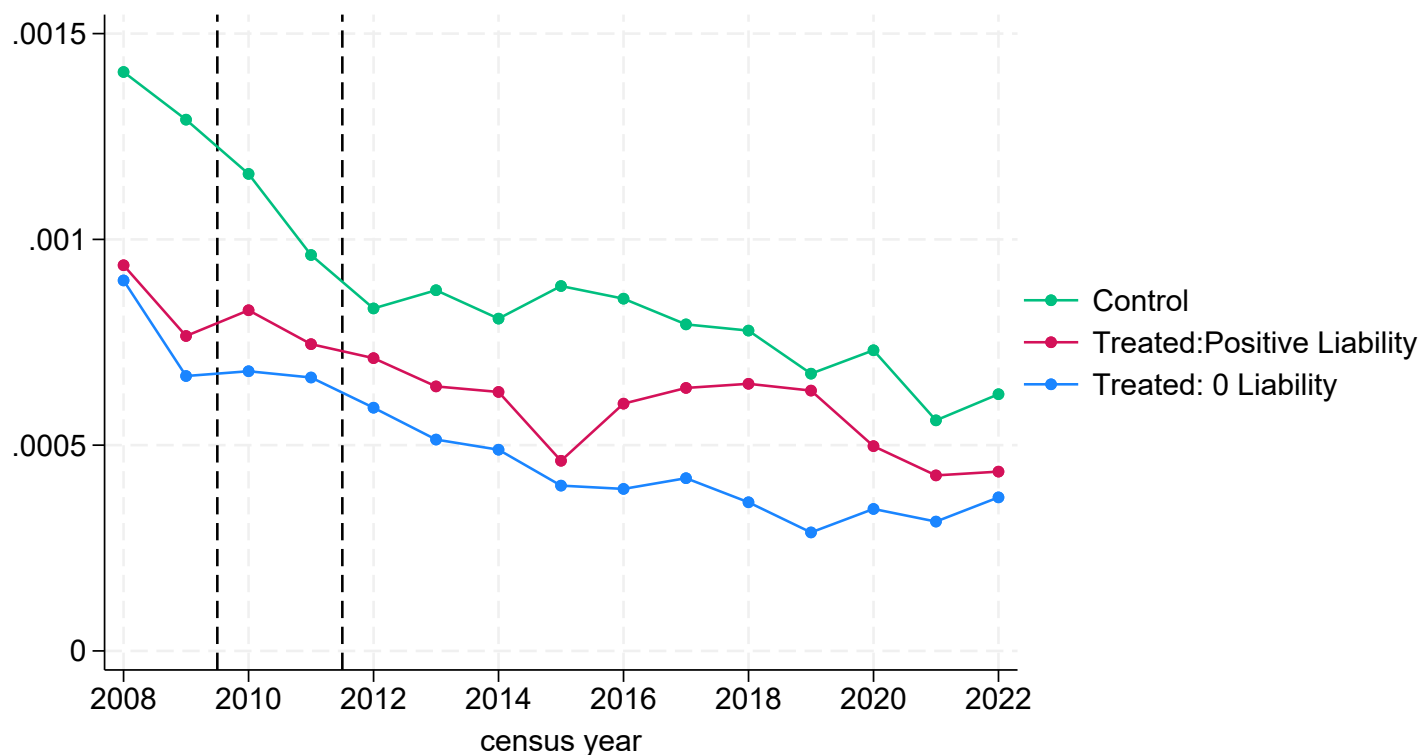
Note: This figure reports the probability that a household undertook a new adoption by treatment group from the 2008-2022 American Community Survey (ACS). The procedure for imputing which children were newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had non-zero tax liability, and the second treated group includes those with zero tax liability. The vertical dashed lines represent when the reform was enacted (starting in 2010) and when it was repealed (starting in 2012).

Figure A4: Mean Number of New Adoptions Aged <2 per Household



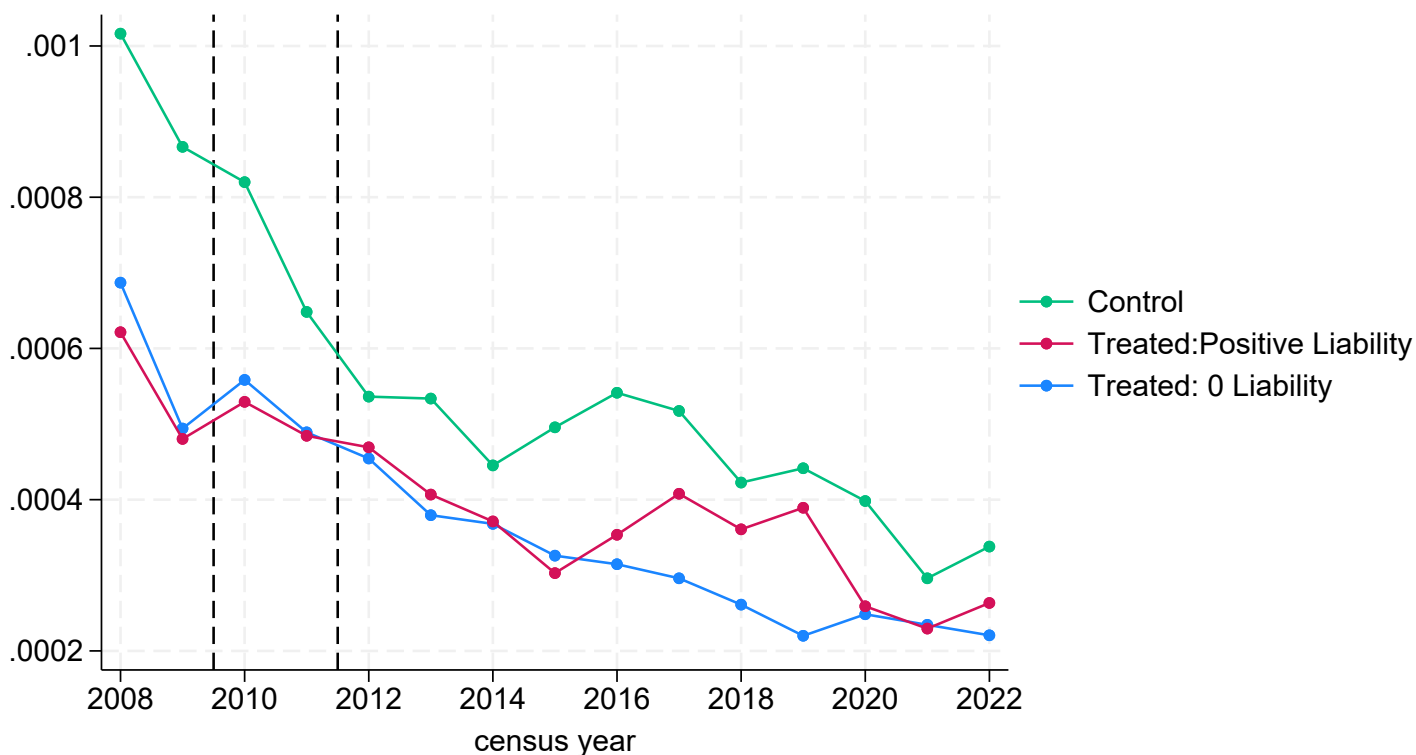
Note: This figure reports the mean number of new adoptions of children aged less than two per household by treatment group from the 2008-2022 American Community Survey (ACS). The procedure for imputing which children were newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had non-zero tax liability, and the second treated group includes those with zero tax liability. The vertical dashed lines represent when the reform was enacted (starting in 2010) and when it was repealed (starting in 2012).

Figure A5: Mean Number of New Domestic Adoptions per Household



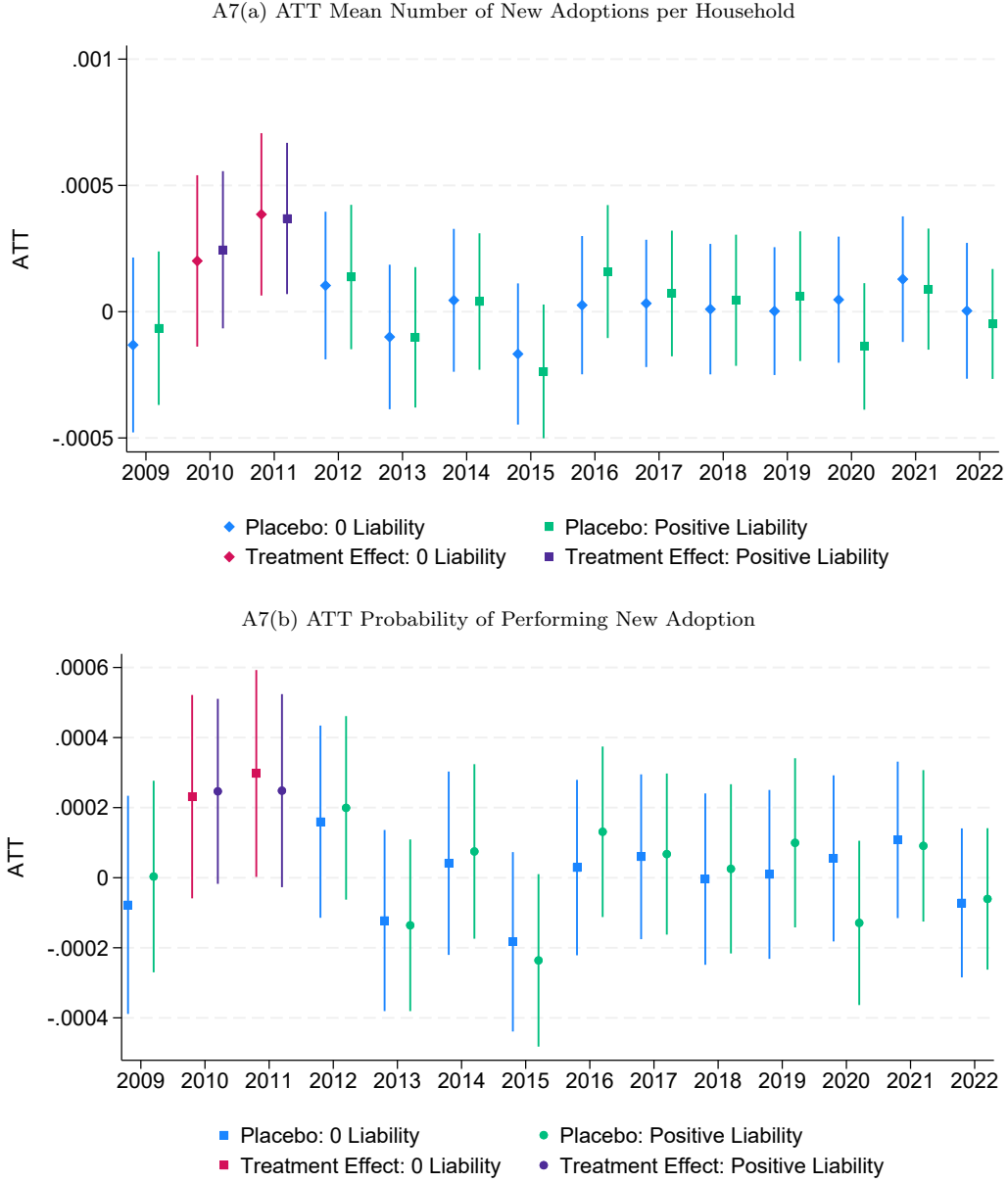
Note: This figure reports the mean number of new domestic adoptions per household by treatment group from the 2008-2022 American Community Survey (ACS). The procedure for imputing which children were classified as domestic adoptions, and which were newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had non-zero tax liability, and the second treated group includes those with zero tax liability. The vertical dashed lines represent when the reform was enacted (starting in 2010) and when it was repealed (starting in 2012).

Figure A6: Mean Number of New Non-Step Adoptions per Household



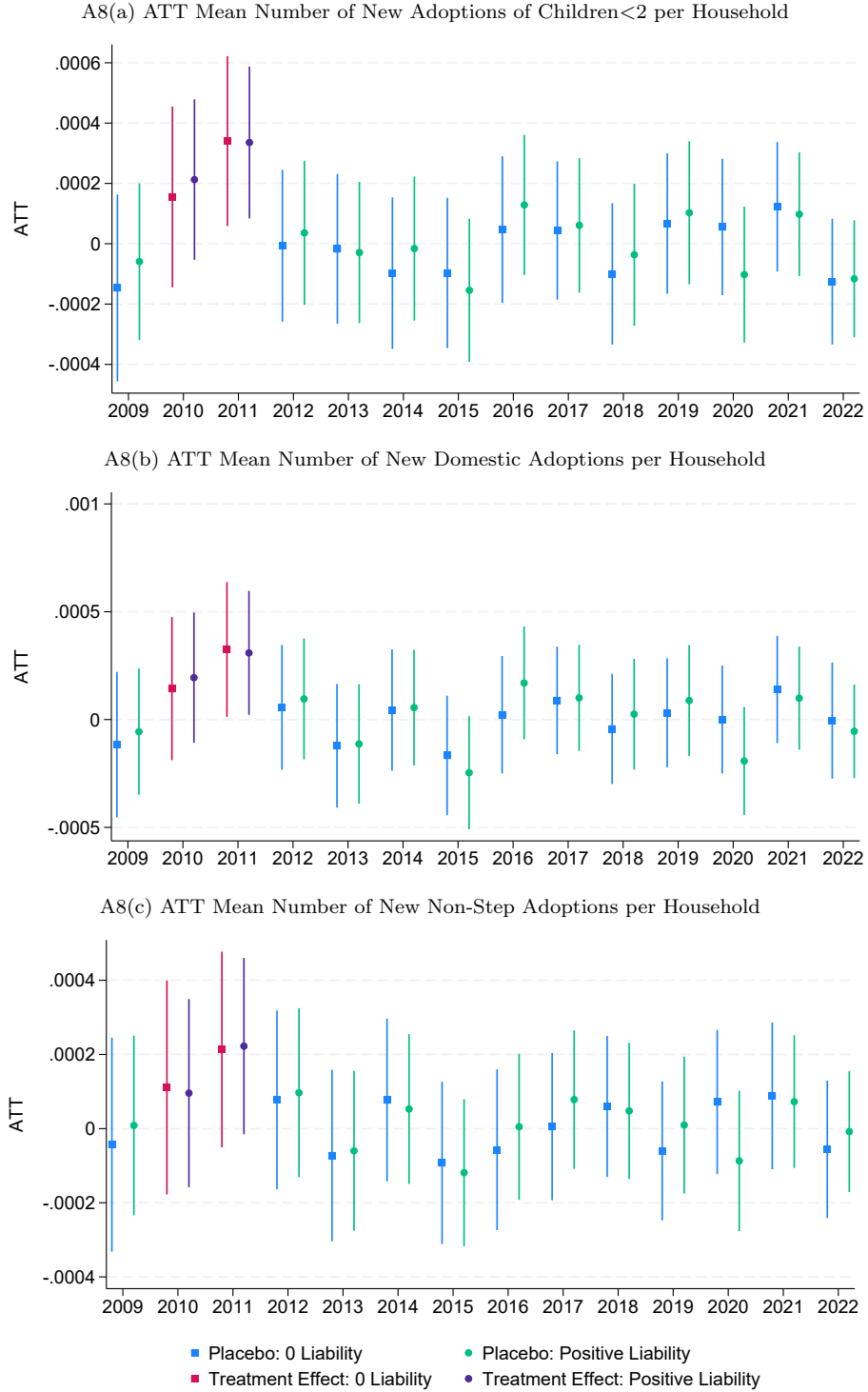
Note: This figure reports the mean number of new non-step-parent adoptions per household by treatment group from the 2008-2022 American Community Survey (ACS). The procedure for imputing which children were classified as step-parent adoptions, and which were newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had non-zero tax liability, and the second treated group includes those with zero tax liability. The vertical dashed lines represent when the reform was enacted (starting in 2010) and when it was repealed (starting in 2012).

Figure A7: ATT Estimates With Placebo Tests for Primary Outcomes



Note: This figure reports the coefficients from the Callaway et al. (2024) specification presented in equation (1) on the mean number of new adoptions per household and on the probability a household performs a new adoption from the 2008-2022 American Community Survey (ACS). The 2010 and 2011 coefficients represent the change in outcomes for treated households relative to control households, as compared to the difference in 2009. For year $Y \in \{2009, 2012, \dots, 2022\}$, the placebo coefficients represent the change in outcomes for treated households relative to control households, as compared to year $Y - 1$. The procedure for imputing which children were classified as newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had zero tax liability, and the second treated group includes those with non-zero tax liability.

Figure A8: ATT Estimates With Placebo Tests for Alternative Outcomes



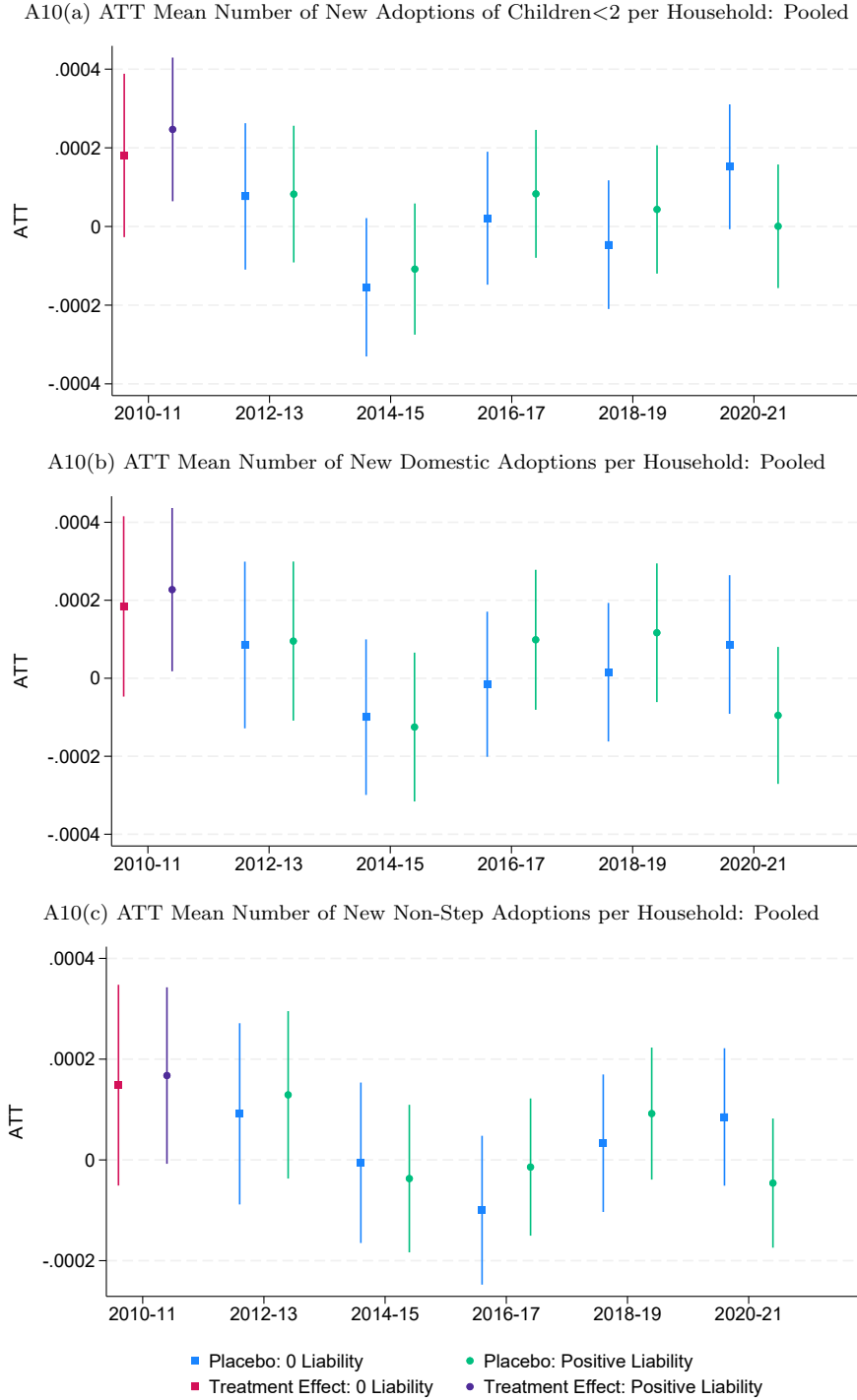
Note: This figure reports the coefficients from the Callaway et al. (2024) specification presented in equation (1) on the mean per-household number of new adoptions: aged <2, classified as domestic adoptions, and classified as non-step-parent adoptions from the 2008-2022 American Community Survey (ACS). The 2010 and 2011 coefficients represent the change in outcomes for treated households relative to control households, as compared to the difference in 2009. For year $Y \in \{2009, 2012, \dots, 2022\}$, the placebo coefficients represent the change in outcomes for treated households relative to control households, as compared to year $Y - 1$.

Figure A9: ATT Estimates With Placebo Tests for Primary Outcomes: Pooled



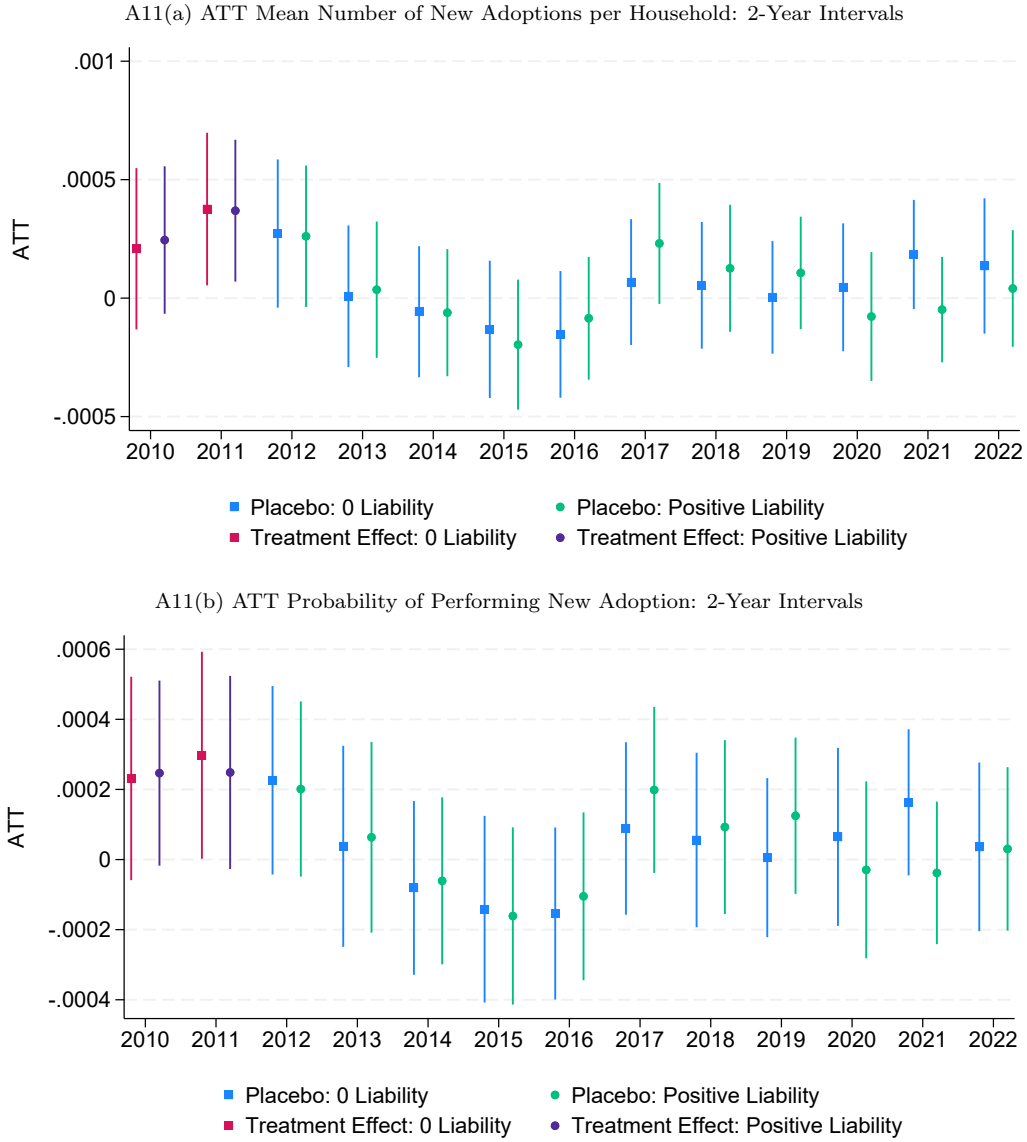
Note: This figure reports the pooled coefficients from the Callaway et al. (2024) specification presented in equation (1) on the mean number of new adoptions per household and on the probability a household performs a new adoption from the 2008-2022 American Community Survey (ACS). The 2010-2011 coefficients represent the change in mean outcomes for treated households relative to control households in the 2010-2011 period, as compared to the difference in the two-year period before reform, 2008-2009. For each subsequent period, the placebo coefficients represent the change in outcomes for treated households relative to control households, as compared to the two-year period preceding it. The procedure for imputing which children were classified as newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had zero tax liability, and the second treated group includes those with non-zero tax liability.

Figure A10: ATT Estimates With Placebo Tests for Alternative Outcomes: Pooled



Note: This figure reports the pooled coefficients from the Callaway et al. (2024) specification presented in equation (1) on the mean per-household number of new adoptions: aged <2, classified as domestic adoptions, and classified as non-step-parent adoptions from the 2008-2022 American Community Survey (ACS). The 2010-2011 coefficients represent the change in mean outcomes for treated households relative to control households in the 2010-2011 period, as compared to the difference in the two-year period before reform, 2008-2009. For each subsequent period, the placebo coefficients represent the change in outcomes for treated households relative to control households, as compared to the two-year period preceding it.

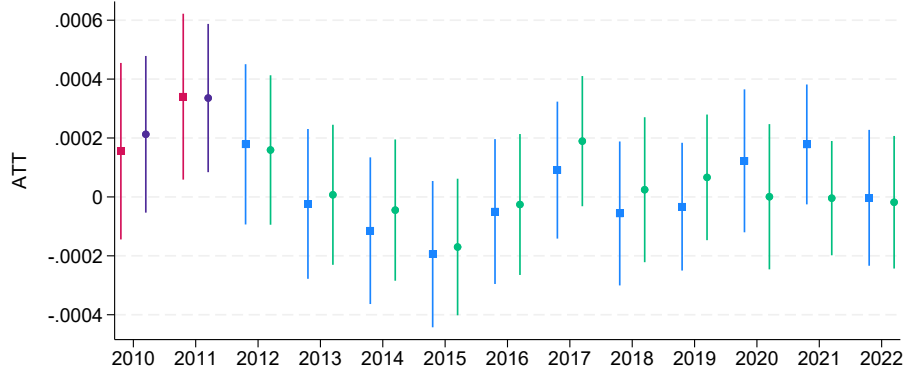
Figure A11: ATT Estimates With Placebo Tests for Primary Outcomes: 2-Years Intervals



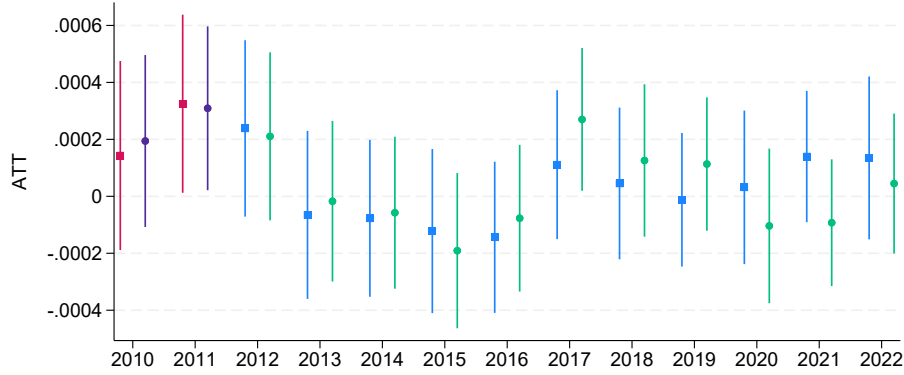
Note: This figure reports the 2-year interval coefficients from the Callaway et al. (2024) specification presented in equation (1) on the mean number of new adoptions per household and on the probability a household performs a new adoption from the 2008-2022 American Community Survey (ACS). The 2010 and 2011 coefficients represent the change in outcomes for treated households relative to control households, as compared to the difference in 2009. There is no estimated coefficient for 2009, since the data only span backward to 2008. For year $Y \in \{2012, \dots, 2022\}$, the placebo coefficients represent the change in outcomes for treated households relative to control households, as compared to year $Y - 2$. The procedure for imputing which children were classified as newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had zero tax liability, and the second treated group includes those with non-zero tax liability.

Figure A12: ATT Estimates With Placebo Tests for Alternative Outcomes: 2-Year Intervals

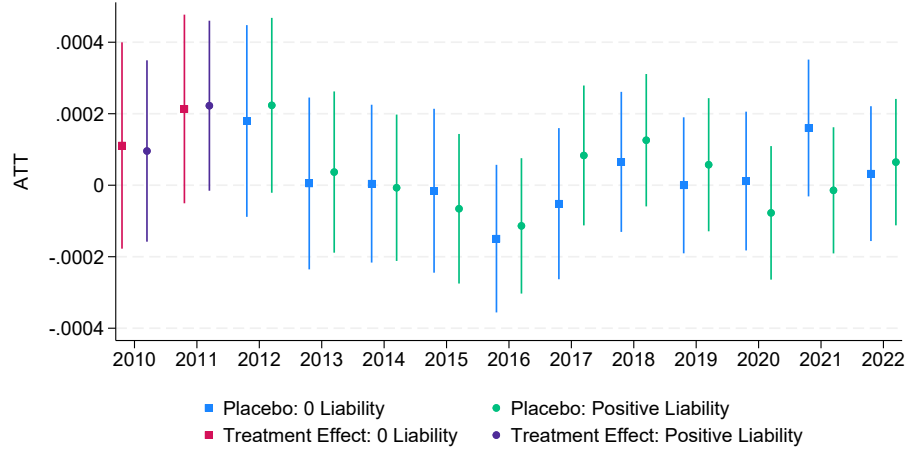
A12(a) ATT Mean Number of New Adoptions of Children <2 per Household: 2-Year Intervals



A12(b) ATT Mean Number of New Domestic Adoptions per Household: 2-Year Intervals

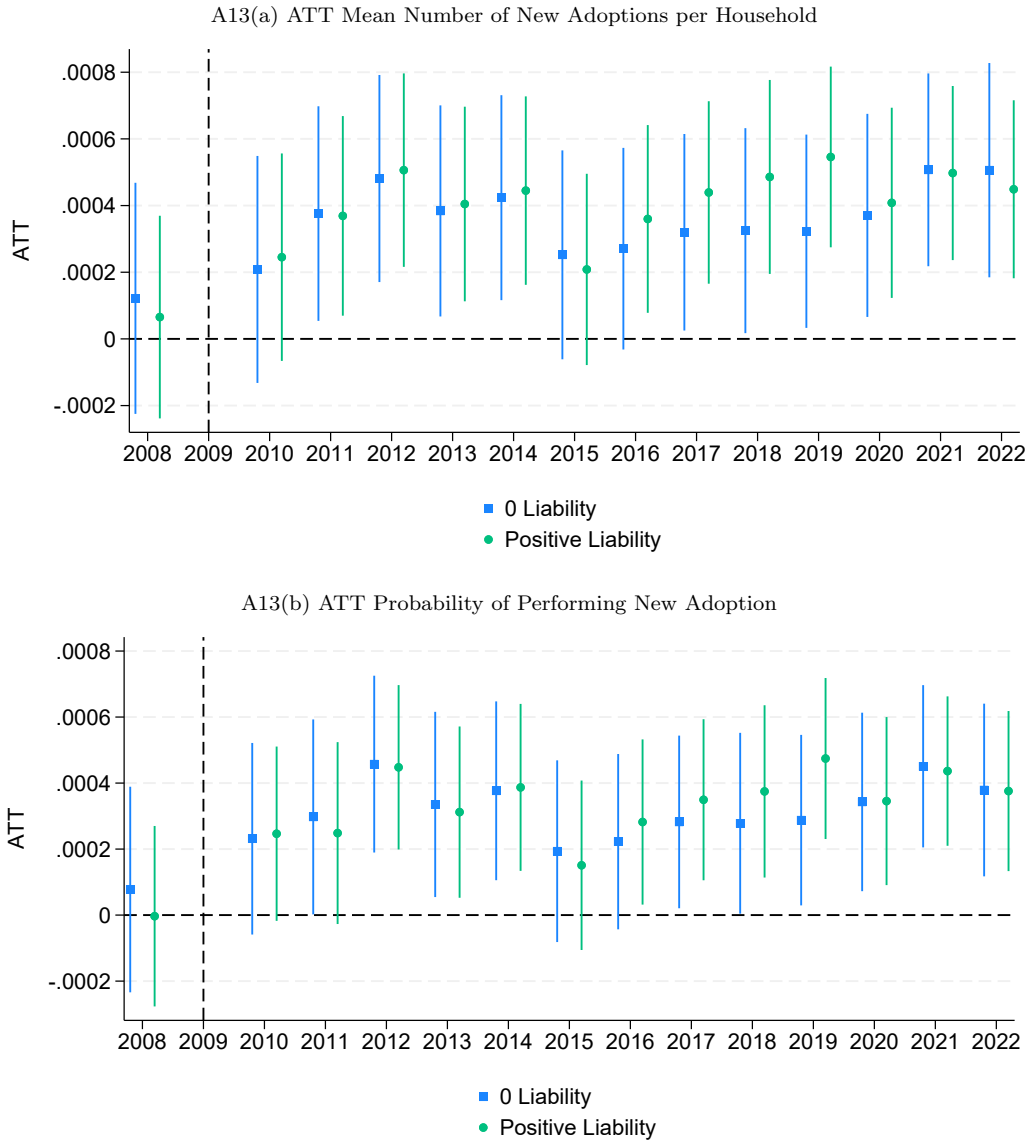


A12(c) ATT Mean Number of New Non-Step Adoptions per Household: 2-Year Intervals



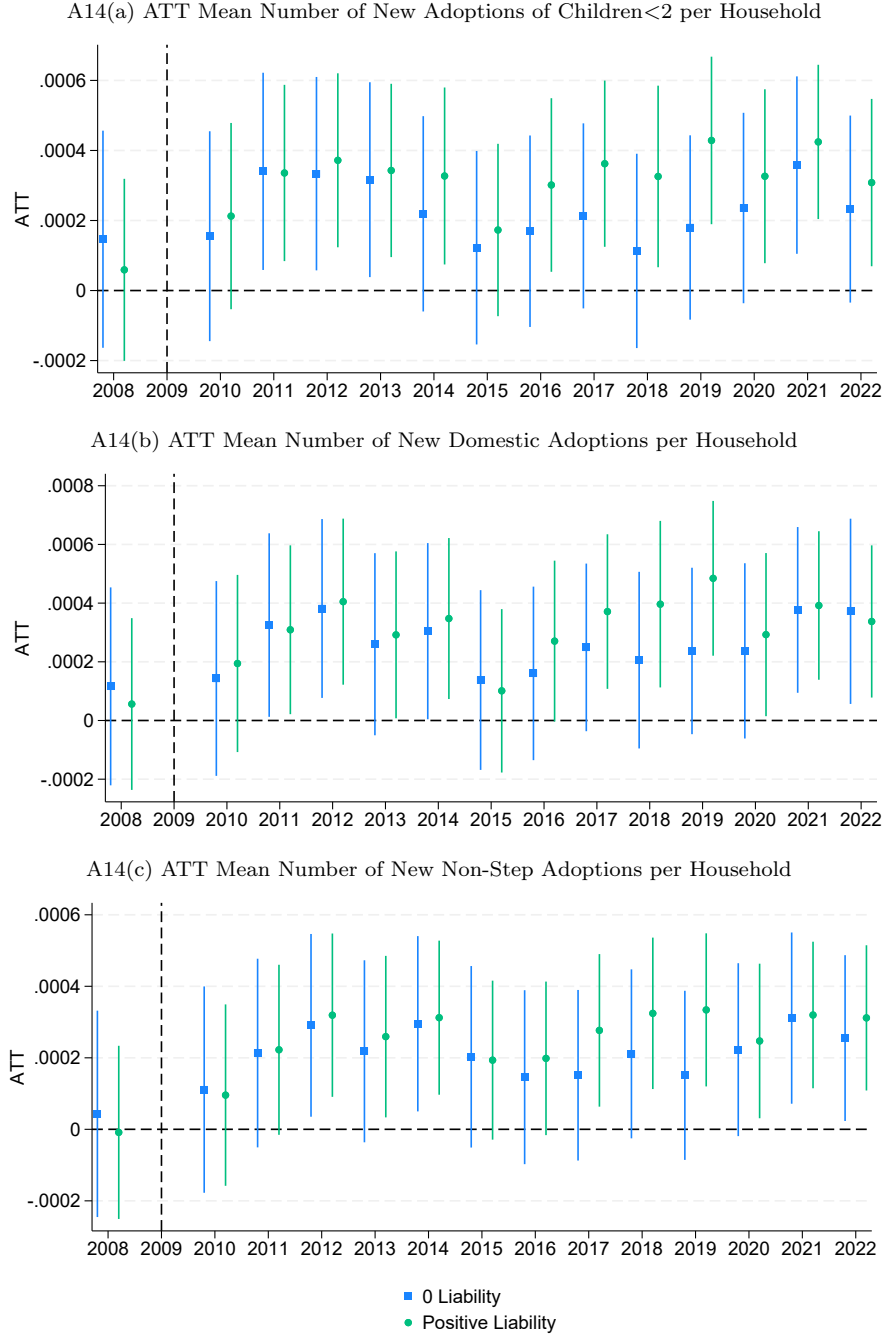
Note: This figure reports the 2-year interval coefficients from the Callaway et al. (2024) specification presented in equation (1) on the mean per-household number of new adoptions: aged <2, classified as domestic adoptions, and classified as non-step-parent adoptions from the 2008-2022 American Community Survey (ACS). The 2010 and 2011 coefficients represent the change in outcomes for treated households relative to control households, as compared to the difference in 2009. There is no estimated coefficient for 2009, since the data only span backward to 2008. For year $Y \in \{2012, \dots, 2022\}$, the placebo coefficients represent the change in outcomes for treated households relative to control households, as compared to year $Y - 2$.

Figure A13: ATT Estimates - Standard TWFE



Note: This figure reports the coefficients from the TWFE specification on the mean number of new adoptions per household and on the probability a household performs a new adoption from the 2008-2022 American Community Survey (ACS). The estimates represent the change in outcomes for treated households relative to control households, as compared to the difference in 2009, the year before reform. The procedure for imputing which children were classified as newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had zero tax liability, and the second treated group includes those with non-zero tax liability.

Figure A14: ATT Estimates - Standard TWFE: Alternate Outcomes



Note: This figure reports the coefficients from the TWFE specification on the mean per-household number of new adoptions: aged <2, classified as domestic adoptions, and classified as non-step-parent adoptions from the 2008-2022 American Community Survey (ACS). The estimates represent the change in outcomes for treated households relative to control households, as compared to the difference in 2009, the year before reform. The procedure for imputing which children were classified as newly adopted is detailed in Figure 1 and in Appendix A. The control group includes households with tax liability greater than \$13,360, which is the maximum credit available in 2011. The first treated group includes those exposed to reform who had zero tax liability, and the second treated group includes those with non-zero tax liability.