HARVARD | BUSINESS | SCHOOL

# Entrepreneurship and Public Health Insurance

Gareth Olds

Working Paper 16-144

# Entrepreneurship and Public Health Insurance

Gareth Olds
Harvard Business School

**Working Paper 16-144**

# ENTREPRENEURSHIP AND PUBLIC HEALTH INSURANCE

Gareth Olds*

This draft: May 2016

**Abstract**

I examine the relationship between public health insurance and firm formation. Developing a variant of regression discontinuity, I find the Child Health Insurance Program lowered the child uninsured rate by 40% and increased self-employment by 15%. Monte-Carlo evidence suggests the technique significantly reduces bias and Type-1 Error. SCHIP increased incorporated ownership by 36% and business share of household income by 12%, implying higher-quality ventures. The mechanism is a reduction in risk rather than credit constraints, and I find no imbalance in observable characteristics between treatment and control groups. These findings strongly suggest social safety nets have spillover benefits on the supply of firms.

*Harvard Business School. Email: golds@hbs.edu. This research was supported in part by the National Science Foundation's IGERT Fellowship and the Hazeltine Fellowship for Research in Entrepreneurship. I am grateful to Ken Chay, Brian Knight and David Weil for their feedback and support. A special thanks to Ronnie Chatterji, Alex Eble, Josh Lerner, Bill Kerr, Dina Pomeranz, and various seminar participants at Harvard and Brown for useful suggestions.

Entrepreneurs occupy a central place in the economist's imagination. Schumpeter called the entrepreneur "certainly the most important person" in his theory of capital interest (Schumpeter, 1961), an observation that sparked a large literature on the role of entrepreneurship in economic growth (Aghion and Howitt, 1997). Empirical evidence suggests young firms are outsize contributors to gross job creation, and firms that survive their early years are important sources of net new jobs (Decker et al., 2014). In recent decades, parallel debates have emerged about the loss of business dynamism as entrepreneurship rates fall and the optimal size of the welfare state. Despite widespread concern about both topics, little is known about how social insurance programs actually affect new firm formation.

This paper provides evidence that public healthcare programs encourage entrepreneurship.[1] I examine the State Child Health Insurance Program (SCHIP), a large-scale public healthcare initiative similar to Medicaid but aimed at children in moderate-income families. Using data from the Current Population Survey (CPS), I show that SCHIP reduced the number of households with uninsured children by 40%, consistent with earlier estimates (Sasso and Buchmueller, 2004; Bansak and Raphael, 2005a). I also show that SCHIP increased the self-employment rate among parents by 15%. These businesses are disproportionately more likely to be high-quality ventures: SCHIP increased the number of incorporated businesses by 36% and the share of household income derived from the business by 12%. This is driven by a 12% increase in firm birth rates and a 26% increase in the rate of incorporated firm formation, and I document only small and insignificant gains in firm survival. I also observe a large increase in labor supply as a result of the program, both on the extensive margin (significantly more wage earners in affected households) and the intensive margin (longer work-years as measured by weeks and hours employed).

To identify the causal effect of the policy, I explore a quasi-experimental research design that enhances regression discontinuity (RD) by incorporating pre-policy data into the estimator. This method, a "difference-in-regression discontinuity" (DRD), identifies the treatment effect using differences in threshold breaks over time.[2] I provide Monte Carlo

[1] I use the terms "self-employment" and "entrepreneurship" interchangeably to signify a household's ownership of a business, which is standard practice in the literature (Parker, 2009). However, since entrepreneurship is frequently associated with risk-taking innovators who incorporate their ventures (Levine and Rubinstein, 2013) I also show that the results are consistent for incorporated business ownership.

[2] Note that this not the same as the first-difference RD estimator (Lemieux and Milligan, 2008) or fixed-effects RD (Pettersson-Lidbom, 2012)—both of which require panel data—since I use cross-sectional information to construct the RD in the post- and pre-policy periods. Like difference-in-differences, this differencing procedure for RD can be implemented on either cross-sectional or panel data, and can be used in conjunction with fixed effects (see the robustness section). Grembi et al. (2012) independently develop a similar estimator, but there are some important differences in identifying assumptions and implementation

evidence that the estimator dramatically reduces bias and Type 1 Error compared to traditional RD and difference-in-differences (DID) estimators, particularly when treatment effects are small or when bandwidths are large. This strategy improves on earlier studies of SCHIP by providing a plausible counterfactual: I find no evidence that treatment and control groups differ in observable characteristics. The core findings are also robust to more traditional quasi-experimental methods, such as propensity score matching.

There are two channels through which the safety net can affect entrepreneurship. Social programs provide insurance against consumption shocks in the case of business failure (the *risk channel*) and free up resources for use as collateral or to fund start-up costs (the *credit channel*). I find the strongest evidence for the risk channel: eligibility for public insurance reduces the risks of entering self-employment, since children can obtain care if the family loses private coverage, which protects the household from paying for potentially expensive procedures out-of-pocket. This is consistent with results in the development literature that uninsured risk, rather than access to credit, is the binding constraint for low-income firms (Karlan et al., 2012; Bianchi and Bobba, 2013). There is also some limited evidence that SCHIP relaxes credit constraints by reducing medical expenditures, allowing households to build up savings and enter self-employment. However, the central mechanism appears to be a reduction in the overall risk of starting a business.

Finally, I examine the affect of the policy on non-parent firm formation as a falsification check. The difference-in-RD estimator can be extended to include a "placebo" population in much the same way that DID models can accommodate a triple-difference estimator. Using both triple-difference and the extended difference-in-RD estimators, I show significant gains in self-employment among parents net of any underlying changes in non-parent entrepreneurship. The extended difference-in-RD method also balances covariates along both the treated/untreated dimension and with respect to parenthood better than the triple-difference estimator, implying a more appropriate treatment-control grouping.

Economists and policymakers both stress the importance of entrepreneurship for economic growth and job creation, but self-employment rates in the US have been steadily declining for five decades. Social insurance programs can promote entrepreneurship by reducing the risks of business ownership and relaxing credit constraints. This paper provides evidence that a large-scale public healthcare program significantly increased the number of entrepreneurs. The findings suggest there may be welfare losses from an inefficiently weak social safety, as public programs have important spillover benefits on the supply of firms.

(see the identification section).

# 1 Background

**Previous literature.**

Very few studies have examined the link between entrepreneurship and health insurance. Fairlie et al. (2011) use differences in healthcare demand between people whose spouses have health insurance and those who do not to identify the relationship between healthcare and self-employment. They also improve upon previous studies by implementing a regression-discontinuity strategy using the change in Medicare eligibility at age 65, motivated by concerns that spousal coverage is not exogenous when couples make employment decisions jointly. Using matched, monthly CPS data from 1996-2006, they find that business ownership increases significantly among men just after they turn 65, increasing by 3.4 percentage points (a 14% increase from the relatively high baseline of 24.6%), suggesting that employer-based healthcare induces some form of "entrepreneurship lock" analogous to a long literature on traditional job lock. However, these results are difficult to generalize for non-elderly populations since new retirees are often at their maximum wealth level and are the least likely to be credit-constrained, as evidenced by their self-employment rate being two times the national average.

However, there is a large literature on the role of public policy more generally in fostering business ownership. Economists have organized their focus along two broad pathways: bankruptcy reform and the tax code. Business ownership is risky, and governments may try to reduce these risks by providing stronger bankruptcy protection to debtors (Fan and White, 2003; Berkowitz and White, 2004; Meh and Terajima, 2008). Tax policy can affect entrepreneurship through marginal tax rates, which change the returns to business success (Gentry and Hubbard, 2000b; Meh, 2005; Bruce and Mohsin, 2006), and wealth taxation, which affects capital accumulation (Cagetti and Nardi, 2009).

One strand of this literature has centered on government interventions into credit markets to reduce rationing—the inability of some borrowers to receive a loan at any interest rate, or restrictions on the maximum loan size a borrower can secure. Credit rationing should generate a positive relationship between entry into entrepreneurship and assets, since banks use collateral as a screening mechanism. Evans and Jovanovic (1989) find a positive and significant effect of assets on entrepreneurship, arguing that 94% of people who are likely to start a business have faced credit rationing, and that 1.3% of the US population has been prevented from starting a business as a result. Holtz-Eakin et al. (1994a), Holtz-Eakin et al. (1994b), and Blanchflower and Oswald (1998) all find a strong link between inherited funds and entrepreneurship, consistent with credit market imperfections that prevent potential entrepreneurs from borrowing against future income.

In the development literature, Bianchi and Bobba (2013) examine a large-scale randomized cash grant program in Mexico and find an 11% increase in the self-employment rate among recipients, though this is driven primarily by a reduction in risk rather than an alleviation of credit constraints. Karlan et al., 2012 also point to the primacy of uninsured risk as a barrier to business formation and expansion by running an insurance and credit intervention among Ghanaian farmers. Finally, governments can guarantee loans or lend directly to alleviate credit rationing, but despite widespread use of loan guarantee programs, there is disagreement about their effectiveness (Parker, 2009).

In addition to influencing lender risk, governments sometimes reduce borrower risk by strengthening bankruptcy protection. There is substantial heterogeneity among states in the amount of assets that are protected under bankruptcy law, and large asset exemptions for personal bankruptcy proceedings—which cover unincorporated businesses—have sizable and significantly positive effects on the number of entrepreneurs (Fan and White, 2003), though this effect may not be linear (Georgellis and Wall, 2006). However, the reduction in risk from entrepreneurs raises the risks faced by lenders, since they cannot recoup some of their losses. These exemptions increase bank rejection rates for new loans, reduce the average value of small business loans, and push up interest rates (Berkowitz and White, 2004). This means government interventions in credit markets to reduce risks for borrowers are at odds with their efforts to reduce lender risk and may in fact increase credit rationing and harm new business formation. Recent evidence suggests that the offsetting effect of increased credit rationing may outweigh the reduction in entrepreneurial risk, so that stronger bankruptcy protection lowers the overall supply of entrepreneurs (Meh and Terajima, 2008).

In some cases governments have enacted social programs to encourage entrepreneurship directly. For example, the United Kingdom ran the Enterprise Allowance Scheme during the 1980's, which provided income support and cash grants to the unemployed for starting businesses. At it's peak in 1987, the program has over 100,000 members and accounted for 30% of the country's new business starts (Parker, 2009). The program was considered cost-effective because it replaced unemployment benefits that the government was already obligated to pay, but the effect on unemployment was small. While the US has not attempted a similar policy, other European countries have established analogous programs with mixed success.

Taxation is the other key channel through which governments influence entrepreneurship. High marginal tax rates may discourage business ownership by reducing the payoff to successful ventures. Some studies have found that both marginal rates and the tax code's overall progressivity negatively impact entry (Gentry and Hubbard, 2000b; Meh,

2005) and innovation (Gentry and Hubbard, 2005), but the effect may be small (Bruce and Mohsin, 2006) or non-monotonic (Georgellis and Wall, 2006). The average rate is positively associated with self-employment rates (Schuetze, 2000), though this is often taken as evidence of tax evasion (Robson and Wren, 1999).

Taxes may also play a part in capital accumulation. Wealth and saving rates are high among the self-employed (Quadrini, 2000), which may in part be driven by the inability of entrepreneurs to finance new businesses with outside funds (Gentry and Hubbard, 2000a). Taxes on wealth can restrict entry into entrepreneurship if entrepreneurs must post collateral for loans or pay a portion of start-up costs out-of-pocket (Cagetti and Nardi, 2006). The effect of tax policy on capital accumulation may be particularly important if credit market imperfections are holding back business formation (King and Levine, 1993). However, more recent work has found a negligible effect of wealth taxes (Cagetti and Nardi, 2009).

Finally, this paper hews closely to extensive work by labor economists on "job lock," the idea that employer-sponsored health insurance discourages workers from switching jobs. Estimates of whether health insurance non-portability creates labor market rigidities have been mixed: some find no effect on job mobility (Holtz-Eakin, 1994; Berger et al., 2004; Fairlie et al., 2012) whereas others find sizable effects (Gruber and Madrian, 1993; Madrian, 1994; Gilleskie and Lutz, 1999; Hamersma and Kim, 2009; Gruber and Madrian, 2002; Bansak and Raphael, 2005b).[3] Recent work by Garthwaite et al. (2013) examines a massive disenrollment in Tennessee's public health insurance program, finding a large increase labor supply. With regard to entrepreneurship, Holtz-Eakin et al. (1996) find that health insurance portability does not affect entry into self-employment, whereas Gumus and Regan (2013) find that tax deductibility for health insurance premiums increased self-employment.

**SCHIP.**

In order to investigate the effect of social insurance programs on entrepreneurship, I examine the State Child Health Insurance Program (now just called the Child Health Insurance Program), which was passed by Congress in July of 1997. The program's original goal was to insure 5 million children using a public health insurance plan modeled in part on an existing plan in Massachusetts. The program was aimed at children whose family income made them ineligible for Medicaid benefits but who nevertheless did not have health insurance. This was motivated by large numbers of uninsured lower- and middle-

[3]See Gruber and Madrian, 2002 for an excellent review of this literature.

class children: in 1998, nearly one in seven families with children had no health insurance coverage for their kids.

SCHIP is structured similarly to Medicaid, with state administration based on federal guidelines and cost-sharing between governments. The law allowed states to administer the program in several ways: they could create their own, separate child health insurance program; they could expand their Medicaid program to cover higher-income children; or they could do a combination of both. Twenty-six states chose a combination of the two, with 17 opting for a separate program and the remainder expanding their Medicaid coverage.

The implementation of SCHIP was relatively swift. The program was originally conceived by Senator Ted Kennedy after Massachusetts passed a similar child health insurance program in July of 1996. President Clinton called for a nationwide program in his January 1997 State of the Union address, and the bill was ratified by the House in June of that year. Forty-one states had SCHIP programs in place by the end of 1998, and there were 1 million enrolled children by the beginning of 1999. The program now covers 7 million children, with about 13% of households reporting a child covered by SCHIP. Eligibility differs by state, but children are usually eligible for coverage up until age 18 if their net family income is below the state income threshold, generally subject to yearly reauthorization. States were also allowed to set these thresholds independently, so there is a wide array of cut-offs with respect to the Federal Poverty Line, ranging from 100% (Arkansas, Tennessee and Texas) to 350% (New Jersey).

For SCHIP to have an effect on employment choices, it is important that the benefits have a significant impact on household budget constraints. While it is less costly to insure a child than an adult, spending on children is still sizable: yearly healthcare expenditures averaged $4,000 for children under 3 and $2,000 for all children, and even high-deductible insurance plans with minimal benefits rarely drop below $1,000.[4] However, the distribution of medical costs has a large upper tail, so parents may be particularly concerned about low-probability events with large price tags. For example, in 2010 the average inpatient procedure for a child cost $12,000, and the average surgical procedure was $35,000. If households place an outsize weight on the upper tail of the cost distribution, public insurance may have a strong impact on labor market decisions even when the immediate shift in the budget constraint is more modest. Further, when households are unable to obtain insurance at any price because of pre-existing conditions they may be reluctant to leave wage employment or reduce hours to start a side-project for fear of losing current

[4]According to the Health Care Cost Institute's Child Health Care Spending Report, 2007-2010 (available at www.healthcostinstitute.org) and the author's calculations.

coverage. Eligibility for SCHIP provides a stop-gap source of insurance for this population, independent of health insurance costs.

**Data.**

I use data from the Current Population Survey (CPS) and the Survey of Income and Program Participation (SIPP) to identify the impact of SCHIP on health insurance status and entrepreneurship. Data come from the 1992-2013 CPS March supplement files, aggregated to the household level since eligibility is determined based on total household income. Samples from both datasets are restricted to non-farm households with at least one child under 18. The main variables of interest are the presence of a child who has health insurance and the presence of a self-employed member. I also make use of a vector of demographic and economics variables in order to test covariate balance; variable descriptions and summary statistics can be found in the appendix.

Eligibility for SCHIP is determined by total household income and the threshold relevant to a family. This income cut-off comes from the state's policy and is based on the Federal Poverty Line; because federal guidelines on poverty vary by household size, there is within-state variation in income thresholds. I use a household's total income from the previous calendar year to determine whether households are eligible at a given date, since families are required to show tax records or pay stubs when applying for SCHIP benefits and when reapplying to maintain coverage in future years.

Finally, data on the state threshold levels and the timing of policy implementation comes from Rosenbach et al. (2001), Mathematica Policy Research's first annual report to the US Department of Health and Human Services on SCHIP implementation.[5] For the 17 states that adopted both a Medicaid expansion and a separate state health insurance program, I use the threshold levels and enrollment dates from the separate program, since the Medicaid expansions tended to have much stricter requirements for child age or birth year.

## 2 Empirical Strategy

In order to estimate the causal impact of SCHIP eligibility on self-employment, I implement a variant of regression discontinuity that uses pre-policy data about the population to incorporate a falsification test directly into the estimator. The intuition for this is straightforward: since eligibility is based on a continuous forcing variable—income—which is col-

[5] The report is available at http://www.mathematica-mpr.com/PDFs/schip1.pdf.

lected both before and after policy implementation, households just above the threshold can be compared both to those just below and to those above and below in the pre-policy regime. Under certain conditions, described below, this procedure eliminates bias that might arise in a traditional RD estimator when the conditional expectation of the outcome variables with respect to the forcing variable are highly non-linear.

Essentially this "difference-in-RD" strategy amounts to first estimating an RD in the post-policy and pre-policy periods and then identifying the differences in the threshold breaks. Because this approach controls for some pre-existing changes in the dependent variable at the threshold in the pre-policy period, this method incorporates a falsification test directly into the estimator in one step. One advantage to this approach is that it does not require that unobserved characteristics of the treated and untreated groups remain comove over time—a variant of the parallel trends assumption from DID. Instead, this only needs to be true for people close to the discontinuity. However, this strategy has the same drawback as all RD designs: it only identifies treatment effects that are local to the threshold, meaning some external validity is sacrificed in order to sharpen the identification. I formalize this procedure below using the potential outcomes notation in Imbens and Lemieux (2008). I will focus on the "sharp" version of RD here, but the results are easily extended to the "fuzzy" case by scaling up the marginal effects along the lines of intent-to-treat models.

**Identification.**

Let $X_i$ be the forcing variable with threshold $c$, and let $Y_{i,P}(1)$ be the outcome for individual $i$ in period $P$ if they are treated and $Y_{i,P}(0)$ be the outcome for the same person if they were untreated, where $P \in \{0, 1\}$ indicates whether $t \geq PolicyYear_s$.

We typically would like to know the population average treatment effect

$$\tau_{ATE} = \mathbb{E}[Y(1) - Y(0)|X]$$

since we never observe $Y_i(1)$ and $Y_i(0)$ for the same person. Regression discontinuity instead identifies the treatment effect at the discontinuity,

$$\tau_{ATE|X=c} = \mathbb{E}[Y_P(1) - Y_P(0)|X = c, P = 1].$$

To estimate this value, RD designs that the potential outcomes for the treated and untreated are continuous in expectation, meaning $\lim_{x\uparrow c} \mathbb{E}[Y_1(0)|X] = \lim_{x\downarrow c} \mathbb{E}[Y_1(0)|X]$ and

$\lim_{x\uparrow c}\mathbb{E}[Y_1(1)|X] = \lim_{x\downarrow c}\mathbb{E}[Y_1(1)|X]$.[6] Under this assumption, $\mathbb{E}[Y_1(0)|X] = \lim_{x\downarrow c}\mathbb{E}[Y_1(0)|X] = \lim_{x\downarrow c}\mathbb{E}[Y_1|X]$, where the first equality is due to continuity and the second to the unconfoundedness assumption that $Y_i(0),\, Y_i(1) \perp\!\!\!\perp Treat_i|X_i$; in the same vein, $\mathbb{E}[Y_1(1)|X] = \lim_{x\uparrow c}\mathbb{E}[Y_1|X]$. The treatment effect at the discontinuity is just

$$\begin{aligned}\tau_{ATE|X=c} &\equiv \mathbb{E}[Y_1(1) - Y_1(0)|X = c] = \mathbb{E}[Y_1(1)|X = c] - \mathbb{E}[Y_1(0)|X = c] \\ &= \lim_{x\uparrow c}\mathbb{E}[Y_1|X] - \lim_{x\downarrow c}\mathbb{E}[Y_1|X] \equiv \tau_{RD}\end{aligned}$$

using the definitions above.

What happens if the continuity assumption isn't satisfied? For example, other policies could have nearby thresholds that cause a jump in the outcome variable.[7] We can use information from before the policy was enacted to get information about the counterfactual for the treatment group. Define

$$\begin{aligned}&\lim_{x\uparrow c}\mathbb{E}[Y_P(0)|X, P = 0] - \lim_{x\downarrow c}\mathbb{E}[Y_P(0)|X, P = 0] \equiv \kappa_{0,0} \neq 0 \\ &\lim_{x\uparrow c}\mathbb{E}[Y_P(1)|X, P = 0] - \lim_{x\downarrow c}\mathbb{E}[Y_P(1)|X, P = 0] \equiv \kappa_{1,0} \neq 0\end{aligned}$$

and define $\kappa_{0,1}$ and $\kappa_{1,1}$ analogously for $P = 1$. Now consider the following assumption:

**Assumption 1.** $\kappa_{1,1} - \kappa_{1,0} = \kappa_{0,1} - \kappa_{0,0}$.

Assumption 1 means that any differences in the expectations of the potential outcome variable for the treated and untreated are the same. Note that this is closely related to the parallel trends assumption, except that instead of the changes in the outcome variables being the same, this assumption requires that the breaks at the discontinuity not change for both groups. Because this version of the parallel trends is local only to the discontinuity, it does not require that potential outcomes for the treated and untreated groups far from the discontinuity grow at the same rates.

**Assumption 2.** $\kappa_{0,1} - \kappa_{0,0} = 0$.

This stronger assumption says that in the absence of treatment, the pre-existing break in the untreated group would remain the same.[8]

[6]For example, this is a restatement of Assumption 2.1 in Imbens and Lemieux (2008).

[7]This could occur even when there is no discontinuity in the actual expectations; there only needs to be a jump in the estimated expectation $\hat{\mathbb{E}}[Y(\cdot)|X_i, P = 1]$ for the problem to arise. Since regression discontinuity designs are typically estimated using a parametric relationship as an approximation to a non-parametric one (to avoid the slow convergence of non-parametric estimators), this assumption of distributional continuity could be true in theory but violated in practice.

[8]Assumption (2) is the same as the first assumption in Grembi et al. (2012). However, their model

Notice that Assumptions 1 and 2 also imply continuity in the differences of the potential outcome functions, even when the expectations themselves are not continuous:

$$\kappa_{0,1} - \kappa_{0,0} = 0 \Leftrightarrow \lim_{x\uparrow c} \mathbb{E}[Y_1(0) - Y_0(0)]|X] = \lim_{x\downarrow c} \mathbb{E}[Y_1(0) - Y_0(0)]|X]$$
$$\kappa_{1,1} - \kappa_{1,0} = 0 \Leftrightarrow \lim_{x\uparrow c} \mathbb{E}[Y_1(1) - Y_0(1)]|X] = \lim_{x\downarrow c} \mathbb{E}[Y_1(1) - Y_0(1)]|X]$$

Using this continuity result and Assumption 1, define the differenced RD estimator as

$$\begin{aligned}
\tau_{DRD} &= \lim_{x\uparrow c} \mathbb{E}[Y_1 - Y_0|X] - \lim_{x\downarrow c} \mathbb{E}[Y_1 - Y_0|X] \\
&= \mathbb{E}[Y_1(1) - Y_0(1)|X = c] - \mathbb{E}[Y_1(0) - Y_0(0)]|X = c] \\
&= \mathbb{E}[Y_1(1) - Y_1(0) - [Y_0(1) - Y_0(0)]|X = c] \\
&= \mathbb{E}[Y_1(1) - Y_1(0)|X = c] \\
&= \tau_{ATE|X=c}
\end{aligned}$$

where the first equality comes from the continuity in the time difference in potential outcomes, the second from properties of expectations, the third from the fact that policies cannot affect outcomes before they are enacted—meaning $\mathbb{E}[Y_0(1) - Y_0(0)|X] = 0$—and the last using the definition of $\tau_{ATE|X=c}$.

To see more clearly how this RD estimator is the one-step equivalent to running regression discontinuity designs in the pre- and post-policy periods, rewrite it as:

$$\tau_{DRD} = \lim_{x\uparrow c} \mathbb{E}[Y_1|X] - \lim_{x\downarrow c} \mathbb{E}[Y_1|X] - \left( \lim_{x\uparrow c} \mathbb{E}[Y_0|X] - \lim_{x\downarrow c} \mathbb{E}[Y_0|X] \right)$$

This is almost the same thing as using the treated group in the pre-period (since the policy did not yet exist, these really are the “would-have-been-treated”) as counterfactuals for the (truly) treated group in the post period. The important difference is that the outcome variable for the pre-policy treatment group is shifted to reflect changes in the untreated

also assumes that the jumps in the conditional expectation are the same in the post period for the treated and untreated groups; using the notation of this paper, the assumption is equivalent to $\kappa_{1,1} = \kappa_{0,1}$. This assumption may not be plausible for contexts in which the underlying characteristics of the treatment group are substantively different than those for the untreated. Assumptions (1) and (2) imply that $\kappa_{1,1} = \kappa_{1,0}$, meaning the underlying jump in the conditional expectation is the same for the treatment group before and after the policy. Here the relationship to the parallel trends assumption becomes more clear: treated and untreated groups need not have the same jumps, only consistent differences over time. In some sense this is weaker than the assumptions in Grembi et al. (2012) because it still guarantees the continuity needed for identification without imposing cross-group restrictions (only within-group ones); however, neither set of assumptions implies the other. The implementation of the estimator (below) also differs, since I do not assume the functional form of the expectation differs in the pre- and post-policy periods.

group over time. Basically, the counterfactual for the treated group in the post-policy period is constructed using two pieces of information: the "treated" pre-period individuals and the trends in the untreated.

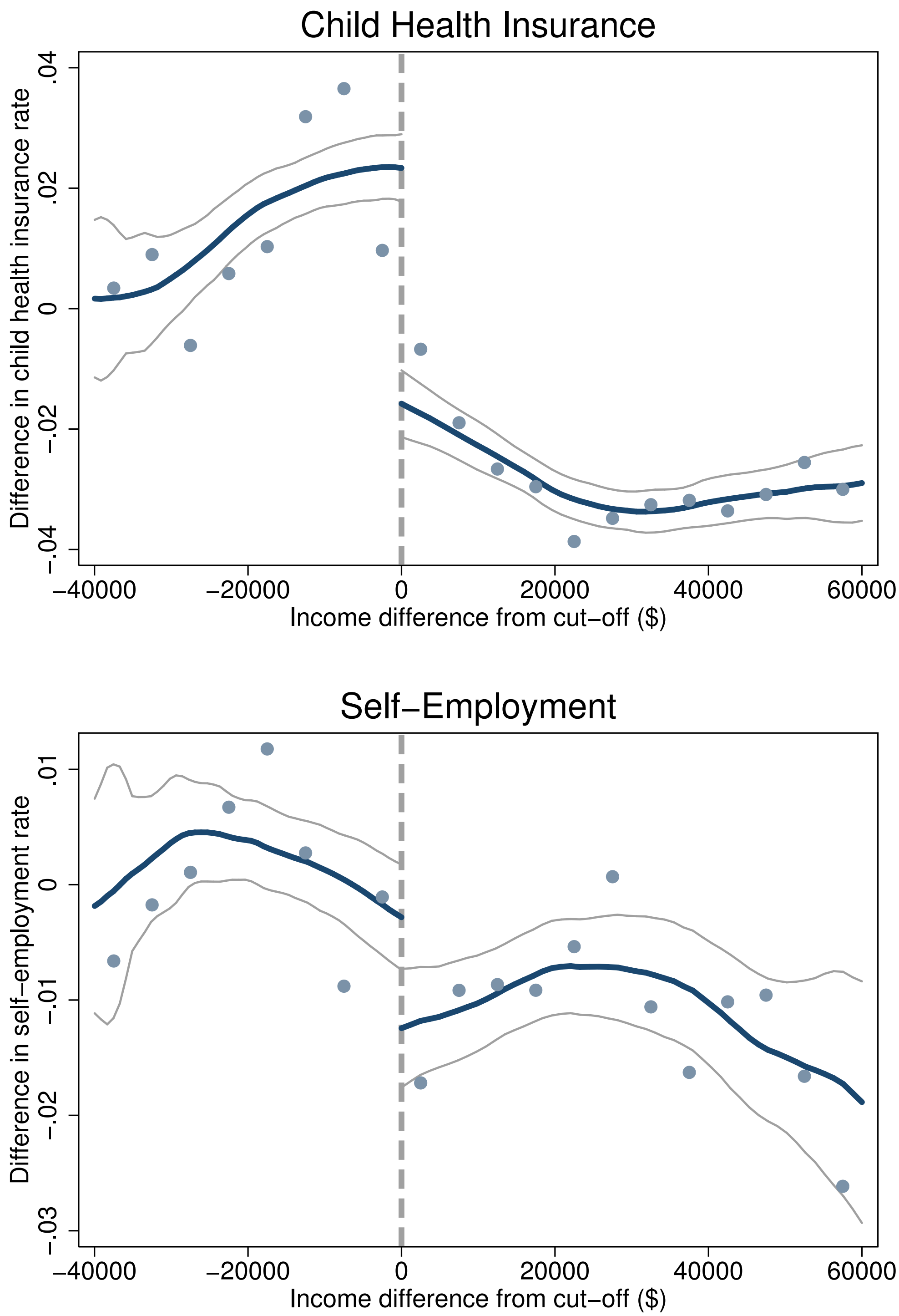
**Intuition and implementation.**

Figure 1 gives some intuition for what this estimator is picking up. The upper panel plots a local polynomial fit for the difference in child health insurance coverage—post-policy minus pre-policy—against the forcing variable—difference between previous year's income and the threshold-–net of state and year fixed effects and state-specific linear trends. These differences are calculated by sorting the assignment rule distribution into equally-sized bins and taking the within-bin time differences in healthcare outcomes, which are then used to fit the lines in Figure 1. The plot also includes average values in \$10,000 intervals, though it is important to note that the fitted lines are based on the localized within-bin differences and not these aggregate values. After the policy was enacted, households just below the cutoff were differentially more likely to have an insured child than those just above the cutoff relative to households in the same area of the distribution before the policy. The lower panel repeats the procedure for self-employment, and the results are similar. While poorer families in general had larger increases in self-employment than their richer counterparts, those just to the left of the threshold were significantly more likely to have a business after the policy, even when accounting for this relationship.

Difference-in-differences essentially takes the average difference in conditional expectations to the left of the break and subtracts the same difference to the right; difference-in-RD, on the other hand, uses only individuals close to the discontinuity while controlling for the slope of the difference in conditional expectations. While the parallel trends assumption in DID makes restrictions on the changes to the whole population below and above the cut-off, the analogous assumption in difference-in-RD only requires that the difference in the expectations below and above the threshold would have remained the same in the post period in the absence of treatment (though the expectations themselves could be shifted).

The estimator can be implemented in the same way as a typical RD. I will focus on the same two methods from the RD literature as before: high-order polynomials and local linear regression. The first is to allow a flexible function $f(\cdot)$ to capture the relationship between the outcome variable and the assignment rule. When households are just below the threshold they are "treated" (i.e. marginally more likely be enrolled, making this a "fuzzy" RD), so an indicator for this captures the marginal increase in the probability of

Figure 1: Regression Discontinuity (Differenced)



enrollment. If $f(\cdot)$ accurately captures the conditional expectation of the outcome variable with respect to the assignment rule, the treatment indicator picks only only differences that arise from being just below the threshold. The second method allows the conditional expectation to vary non-parametrically by using a linear function of the assignment rule and restricting the sample to data within an arbitrarily narrow bandwidth around the threshold. In the context of difference-in-RD, this means estimating for individual $i$ in state $s$ at time $t$:

$$Y_{ist} = \beta_1 + \beta_{DRDpoly} \cdot Treat_{it}Post_{st} + \beta_2 \cdot Treat_{it} + \beta_3 \cdot Post_{st} + f(Rule_{it}) + \nu_s + \eta_t + \gamma \cdot t\nu_s + \varepsilon_{ist} \quad (1)$$

for the polynomial regression and

$$Y_{ist} = \beta_1 + \beta_{DRDlin} \cdot Treat_{it}Post_{st} + \beta_2 \cdot Treat_{it} + \beta_3 \cdot Post_{st} + \beta_4 \cdot Rule_{it} + \nu_s + \eta_t + \gamma \cdot t\nu_s + \varepsilon_{ist} \qquad \text{if } abs(Rule_{it}) \leq W \quad (2)$$

for the local linear specification, where $Rule_{it} = Inc_{it-1} - Thresh_{ist}$ is the forcing variable, $f(\cdot)$ is a high-order function, and $Y_{ist}$ is either child health insurance status or self-employment. The variable $Treat_{it} = \mathbf{1}[Inc_{it-1} \leq Thresh_{ist}]$ is an indicator for whether the household is "treated" (i.e. has income below the relevant policy threshold), the term $Post_{st} = \mathbf{1}[t > PolicyYear_s]$ is an indicator for being observed in the post-policy period, and $Treat_{it}Post_{st}$ is their interaction. The parameters $\nu_s$ and $\eta_t$ are state and year fixed effects, and $\gamma \cdot t\nu_s$ is a state-specific linear time trend.[9] A positive and significant coefficient on $\beta_{DRD}$ in the health insurance regression would mean SCHIP increased the number of insured children; if the same result holds with self-employment on the left-hand side, it suggests SCHIP eligibility positively affected business ownership.

**Results.**

Table 1 shows the parametric estimates using these two methods, with child health insurance coverage as the outcome variable in Panel A and presence of a self-employed individual as the measure in Panel B.[10] The results in column (1) of Panel A suggests that SCHIP increased the proportion of households with an insured child by 4.68 percent-

[9]The assumption of a linear trend is not important here; state-by-year fixed effects produce similar results.

[10]Density plots of the assignment rule in the pre- and post-policy periods appear in the appendix. The functions are smooth near the discontinuity in both periods, meaning there is no evidence that people manipulated the assignment rule in order to select into (or out of) treatment.

age points, which represents a 34% drop in the rate of uninsured children; the inclusion of demographic covariates increasing this number somewhat to 5.5 percentage points. Though the results are all significant statistically and in policy terms, they is markedly lower than the 9 percentage point drop reported in Sasso and Buchmueller (2004) and Bansak and Raphael (2005a). The local linear method, shown in columns (2)-(6), further reduces this amount; restricting the sample to households with income less than $5,000 away from the cutoff drops the treatment effect to a 1.9 percentage points drop in the number of uninsured households, translating to a local average treatment effect of 11%. In treatment-on-the-treated terms, where the pre-policy "treated" are those below the threshold, the drop in the uninsured is closer to 10%. The decrease in the effect size is due to the fact that lower-income households are less likely to be insured in the first place, so fixed percentage-point gains in insurance rates have a proportionally smaller effect on the ranks of the uninsured.

The self-employment are reported in Panel B, and show business ownership increasing by more than 4 percentage points using the polynomial method—column (1)—and nearly 2 percentage points using data close to the cutoff—column (2). These treatment effects imply an increase in self-employment rates of 23% using the polynomial method and 14% using the local linear method. The results are on the same order of magnitude to those in Fairlie et al. (2011), who find a 14% increase in the self-employment rate as a result of Medicare coverage. However, in treatment-on-the-treated terms, the increase in self-employment is between 15% (local linear) and 30% (polynomial), since lower-income households are less likely to own a business. About 11% of the target population had a self-employed member before the policy took effect, in keeping with nationally-representative statistics of business ownership.

# 3 Model Extensions and Evaluation

So far the analysis has focused only on households that have children living at home at the time the family is surveyed. A simple extension of the empirical strategy is to incorporate households that do not have children: if the estimators are truly picking up the causal effect of the policy, the findings should hold after netting out any concurrent changes in the self-employment rate of childless households. These changes might represent underlying bias in the estimator, movement of households who anticipate soon becoming parents into self-employment as a result of the policy, or both. In the first case the inclusion of childless households will simply remove a nuisance parameter, whereas in the second case the estimator excises more indirect effects of the policy to focus on the direct effects

Table 1: Difference-in-Regression Discontinuity Results

| | Polynomial Method | | Local Linear Method | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Panel A: Child Health Insurance | | | | | | |
| $Treat \cdot Post$ | 0.0468** | 0.0550** | 0.0415** | 0.0346** | 0.0193* | 0.0192* |
| | (0.00546) | (0.00470) | (0.00467) | (0.00529) | (0.00796) | (0.00743) |
| Pre-policy avg. | 0.861 | 0.861 | 0.829 | 0.826 | 0.827 | 0.827 |
| LATE (%) | 5.438 | 6.390 | 5.003 | 4.197 | 2.334 | 2.323 |
| Treated avg. | 0.783 | 0.783 | 0.783 | 0.793 | 0.810 | 0.810 |
| TOT (%) | 5.980 | 7.027 | 5.297 | 4.366 | 2.383 | 2.372 |
| Bandwidth | — | — | ±$20,000 | ±$10,000 | ±$5,000 | ±$5,000 |
| Covariates | No | Yes | No | No | No | Yes |
| Observations | 475,546 | 456,756 | 203,411 | 106,034 | 53,566 | 50,942 |
| R-squared | 0.047 | 0.113 | 0.035 | 0.032 | 0.032 | 0.091 |
| Panel B: Self-Employment | | | | | | |
| $Treat \cdot Post$ | 0.0323** | 0.0211** | 0.0142** | 0.0104* | 0.0180** | 0.0185** |
| | (0.00402) | (0.00347) | (0.00303) | (0.00390) | (0.00578) | (0.00542) |
| Pre-policy avg. | 0.140 | 0.140 | 0.120 | 0.125 | 0.127 | 0.127 |
| LATE (%) | 23.10 | 15.12 | 11.88 | 8.307 | 14.19 | 14.55 |
| Treated avg. | 0.102 | 0.102 | 0.105 | 0.116 | 0.117 | 0.117 |
| TOT (%) | 31.52 | 20.63 | 13.58 | 8.967 | 15.35 | 15.75 |
| Bandwidth | — | — | ±$20,000 | ±$10,000 | ±$5,000 | ±$5,000 |
| Covariates | No | Yes | No | No | No | Yes |
| Observations | 481,678 | 462,760 | 206,155 | 107,487 | 54,295 | 51,648 |
| R-squared | 0.023 | 0.049 | 0.009 | 0.008 | 0.008 | 0.054 |

OLS. Dep. var. is presence of insured child (Panel A) or presence of self-employed member (Panel B). Polynomial fits 6th-order function using entire distribution, local linear uses data within bandwidth. Includes time/state FE and state trends. See appendix for covariate list. Data: CPS 1992-2013.
** 0.01, * 0.05, + 0.1. Robust standard errors clustered at the state level in parentheses.

of eligibility (as opposed to anticipated future eligibility). In either case the basic results should maintain if the difference-in-RD estimator signifies a causal relationship.

I use the standard triple-difference estimator as a benchmark:

$$\begin{aligned}Y_{ist} = \beta_1 + \beta_{DDD} \cdot Treat_{it}Post_{st}HasKids_{it} + \beta_2 \cdot Treat_{it}HasKids_{it} \\ + \beta_3 \cdot Post_{st}HasKids_{it} + \beta_4 \cdot Treat_{it}Post_{st} + \beta_5 \cdot Treat_{it} + \beta_6 \cdot Post_{st} \\ + \beta_7 \cdot HasKids_{it} + \nu_s + \eta_t + \gamma \cdot t\nu_s + \varepsilon_{ist}\end{aligned} \quad (3)$$

where $HasKids_{it}$ is a binary indicator of whether household $i$ has at least one child living at home in time $t$, $Treat_{it}$ is an indicator for the household income being below the threshold level, and all other variables defined as before. The triple-difference (DDD) estimator $\beta_1$ will identify the effect SCHIP eligibility on the outcome variable relative both to ineligible households, households in the pre-policy regime, and households that do not have children. The difference-in-RD estimator can be analogously extended to include childless households by interacting $HasKids_{it}$ with all $Treat_{it}$ and $Post_{st}$ variables in equations 1 and 2; the result looks identical to equation 3 except with $Rule_{it}$ terms and applicable bandwidth restrictions. This double-difference-in-RD (DDRD) estimator identifies the marginal effect of being just below the income cut-off relative to households just above the cutoff, households around the threshold in the pre-policy period, and smiler households that have no children.

Table 2 shows the results when childless households are included, using self-employment as the outcome variable.[11] Columns (1) and (2) show the DDD results with and without covariates, which indicate a 0.8 percentage point increase in the self-employment rate as a result of SCHIP eligibility (though this is smaller and insignificant once covariates are introduced). The DDRD estimators, however, report a 0.6 to 1.5 percentage point increase; using the local linear method and including covariates, self-employment increased by 11.3% relative to the pre-policy baseline, or about 15% among the treated population. These are smaller though on the same order of magnitude as the estimates in Table 1, suggesting the difference-in-RD estimator is not being driven by underlying changes in the general population that are unrelated to having a child in the household.

Since the difference-in-RD model includes elements of both DID and RD, one natural way to evaluate the estimator is to compare it to the two more traditional methods.

[11] Child health insurance coverage obviously cannot be used in this case, since the variable is undefined for childless households.

Table 2: Triple-Difference and Double-Difference-in-RD Results

| | DDD | | DDRD, Polynomial | | DDRD, Local Linear | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Dep. variable: Self-employment | | | | | | |
| $Treat \cdot Post$ | 0.00858** | 0.00503 | 0.0127** | 0.00692* | 0.0128+ | 0.0148* |
| $\cdot HasKids$ | (0.00300) | (0.00307) | (0.00296) | (0.00301) | (0.00654) | (0.00628) |
| Pre-policy avg. | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 | 0.131 |
| LATE (%) | 6.549 | 3.839 | 9.707 | 5.277 | 9.735 | 11.30 |
| Treated avg. | 0.0973 | 0.0973 | 0.0973 | 0.0973 | 0.0973 | 0.0973 |
| TOT (%) | 8.826 | 5.174 | 13.08 | 7.112 | 13.12 | 15.23 |
| Bandwidth | — | — | — | — | ±\$5,000 | ±\$5,000 |
| Covariates | No | Yes | No | Yes | No | Yes |
| Observations | 1,039,514 | 999,363 | 1,039,512 | 999,361 | 126,926 | 120,874 |
| R-squared | 0.010 | 0.039 | 0.024 | 0.046 | 0.009 | 0.044 |

OLS. Dependent variable is presence of self-employed member.

Polynomial method fits a 6th-order function using the entire dataset, local linear uses data within the bandwidth.

Time/state FE and state trends included in all specifications. See appendix for covariate list. Data: CPS 1992-2013.

** 0.01, * 0.05, + 0.1. Robust standard errors clustered at the state level in parentheses.

Consider the standard DID estimator $\beta_{DID}$ identified from:

$$Y_{ist} = \beta_1 + \beta_{DID} \cdot Treat_{it} Post_{st} + \beta_2 \cdot Treat_{it} + \beta_3 \cdot Post_{st} + \nu_s + \eta_t + \gamma \cdot t\nu_s + \varepsilon_{ist} \quad (4)$$

and the standard RD estimator $\beta_{RD}$ identified from:

$$Y_{ist} = \beta_1 + \beta_{RDpoly} \cdot Treat_{it} + f(Rule_{it}) + \nu_s + \eta_t + \gamma \cdot t\nu_s + \varepsilon_{ist} \quad \text{if } Post_{st} = 1 \quad (5)$$

for the polynomial method, and

$$Y_{ist} = \beta_1 + \beta_{RDlin} \cdot Treat_{it} + \beta_2 \cdot Rule_{it} + \nu_s + \eta_t + \gamma \cdot t\nu_s + \varepsilon_{ist}$$
$$\text{if } Post_{st} = 1 \text{ and } abs(Rule_{it}) \leq W \quad (6)$$

for the local linear method. As an evaluation metric, consider demographic covariates that should be unaffected by the policy: in an experimental setting, differences between treatment and control groups along these lines is evidence of poor randomization or self-selection into treatment. Similarly, an estimator is less convincing as having isolated the causal effect of an intervention if it also picks up underlying differences in the population, since the synthetic “control” group differs significantly from those affected by the policy. In order to test whether covariates are balanced, I apply the estimators in equations 1, 2 and 3 to a range of demographic variables, and compare the results to the more traditional methods captured in equations 4, 5 and 6.

The results are shown in Table 3. Difference-in-differences estimation, reported in column (1), suggests that SCHIP induced a 3.4 percentage point increase health insurance coverage and a 1.2 percentage point increase in the self-employment rate. However, repeating the specification using each of the covariates as an outcome variables reveals that nine out of 16 of these variables—about 56%—vary in a statistically significant way at the 5% level along the same dimension supposedly being explained by the policy. Even triple-differencing, which should control for any underlying changes in the population so long as they are parallel between childless and non-childless households, shows a significant imbalance between covariates. This may not be surprising as there are several well-known problems with difference-in-difference strategies. Besides issues with over-rejection in the presence of serial correlation (Bertrand et al., 2004), the logic of DID can be misleading: comparing changes in one group against changes in another suggests all else being equal, the changes are good counterfactuals for one another. Put another way, any underlying unobservable characteristics between the two groups should not be changing in different ways (this is the parallel trends assumption).

Note that controlling for these covariates in the DID specification does not solve the problem of non-orthogonality of covariates reported in Table 3: unconditional orthogonality of assignment to the treatment group with respect to observables and unobservables is a stricter requirement than conditional orthogonality (with respect to unobservables) once controls are included. While both rely on assumptions about unobservables, the first can be partially tested by using observables as outcomes; the second requires making an assumption about observables containing sufficient information such that unobservables are as good as random.

Another metric useful for evaluating an estimator is its sensitivity to parameter choices that are within the control of the econometrician. In an RD setting, the researcher typically specifies a bandwidth either using an optimal rule based on the data's density or presents a range of treatment effects for different bandwidth choices. Difference-in-RD eliminates much of the heterogeneity in the treatment effect by bandwidth choice since any bias induced by an inappropriately large window will be cancelled out so long as that bias also exists when estimating the model in the pre-policy regime. In the context of SCHIP, Figure A1.2 in the Appendix presents the estimated treatment effect and 95% confidence intervals for a wide range of bandwidth choices, both for child health insurance coverage and self-employment. The results are generally stable across the entire range, though losing in precision as the window narrows and more data is excluded from the regression. This same finding is not true when the model is estimated using a traditional RD framework (not shown). I present some Monte Carlo evidence in support of this relative bandwidth-invariance later on in the discussion.

### Business source and quality.

Where is this business growth coming from? Table 4 breaks down the self-employment results by previous businesses status, using the falsification procedure described above. The effect for people who already have businesses 1.3 percentage points, suggesting SCHIP caused an insignificant 1.7% increase in the pre-policy yearly firm survival rate of 74%.[12] The effect on non-business-owning households is 0.6 percentage points, a 12% increase in the yearly firm birth rate of 4.7%. The marginal effects are actually highest for the incorporated firm birth rate, which increased by 0.27 percentage points (26% higher than the baseline). This was entirely driven by new firms rather than newly-incorporated existing firms, whose numbers actually fell as a result of the policy. The growth also

[12] These numbers should be interpreted with caution since the sample sizes are much smaller, both because of restrictions based on previous employment status and because the sample of households observed at two time periods is smaller.

Table 3: Covariate Balance Across Models

| Estimator: | | | Polynomial Method | | | Local Linear Method | | |
|---|---|---|---|---|---|---|---|---|
| | DID | DDD | RD | DRD | DDRD | RD | DRD | DDRD |
| Dep. variable | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Child Health Ins. | 0.0338** | — | -0.0352** | 0.0468** | — | 0.00566 | 0.0193* | — |
| Self-Employment | 0.0120** | 0.00858** | -0.000226 | 0.0323** | 0.0127** | 0.00175 | 0.0180** | 0.0128+ |
| Age of Adults | -0.280** | 1.122** | -0.377** | 0.473** | 1.232** | 0.0831 | 0.132 | 0.392 |
| Bachelor's Deg. | -0.0376** | -0.0196** | -0.0500** | 0.0396** | -0.00507 | -0.0224** | 0.00206 | -0.000907 |
| Black | -0.00821 | -0.0248** | 0.0427** | -0.0256** | -0.0284** | 0.0286** | 0.00420 | -0.00313 |
| Children | -0.0518** | -0.0513** | -0.142** | -0.158** | -0.0684** | 0.0408* | -0.0131 | -0.0132 |
| Disabled | 0.00199** | 0.00464** | 0.00297** | 0.00171* | 0.00450** | -0.00183 | -0.000957 | 0.00269 |
| Graduate Deg. | -0.0203** | -0.00770** | 0.00536* | 0.0186** | -0.000562 | -0.000970 | -0.000588 | 0.00539 |
| High School | 0.0131* | -0.0130+ | -0.0632** | 0.0455** | -0.00646 | -0.0215** | 0.00459 | -0.0116 |
| Hispanic | 0.0119 | 0.0206* | 0.0606** | -0.0123 | 0.0166+ | 0.0253** | -0.00271 | 0.000689 |
| Household Size | -0.0451** | -0.0474** | -0.317** | -0.119** | -0.0538** | 0.0635* | -0.00873 | -0.0411+ |
| Married | 0.00436 | 0.0163+ | -0.103** | 0.0526** | 0.0282** | 0.0113 | 0.00591 | 0.000496 |
| Medicaid | -0.0417** | -0.0755** | 0.111** | -0.0754** | -0.0816** | 0.0305** | 0.00499 | -0.000400 |
| Moved | -0.0103+ | -0.0182** | 0.0554** | -0.0205** | -0.0205** | 0.0439** | 0.00996 | 0.0108 |
| Renter | 0.00148 | -0.0303** | 0.165** | -0.0565** | -0.0428** | 0.0461** | 0.00310 | 0.00395 |
| Unemployment | 0.00488 | 0.00235 | 0.0146** | -0.00147 | 0.00164 | 0.0185* | -0.000022 | -0.000842 |
| Urban | 0.00335 | -0.0167** | -0.00263 | 0.0206** | -0.0138* | 0.0108 | -0.00164 | -0.0141+ |
| Veteran | -0.00244** | 0.00709** | -0.00105 | -0.00112 | 0.00736** | 0.00222 | -0.000739 | 0.000674 |
| Bandwidth | — | — | — | — | — | $\pm$\$5,000 | $\pm$\$5,000 | $\pm$\$5,000 |
| Unbalanced vars. | | | | | | | | |
| $\alpha = 5\%$ | 56% | 81% | 88% | 81% | 69% | 63% | 0% | 0% |
| $\alpha = 10\%$ | 63% | 94% | 88% | 81% | 75% | 63% | 0% | 13% |

OLS. Coefficients reported are treatment effects for the applicable model, with row as dependent variable.
Polynomial method fits a 6th-order function using the entire dataset, local linear uses data within the bandwidth.
Time/state fixed effects and state trends included. Data: CPS 1992-2013.
** 0.01, * 0.05, + 0.1 Robust standard errors clustered at the state level are omitted for brevity.

Table 4: Business Source (Difference-in-RD)

| | Business was previously ... | | | | Household was ... | |
|---|---|---|---|---|---|---|
| | Existent | Non-existent | | Unincorp. | Employed | Unemp. |
| Dependent | (1) | (2) | (3) | (4) | (5) | (6) |
| variable: | Self-Emp. | Self-Emp. | Incorp. | Incorp. | Self-Emp. | Self-Emp. |
| $Treat \cdot Post$ | 0.0126 | 0.00579 | 0.00272+ | -0.0159+ | 0.0223** | 0.00709 |
| | (0.0180) | (0.00379) | (0.00162) | (0.00912) | (0.00588) | (0.0168) |
| Pre-policy avg. | 0.742 | 0.0466 | 0.0106 | 0.0106 | 0.158 | 0.0237 |
| LATE | 1.697 | 12.42 | 25.82 | -150.9 | 14.10 | 29.85 |
| Treated avg. | 0.659 | 0.0441 | 0.00451 | 0.00451 | 0.129 | 0.0160 |
| TOT | 1.909 | 13.14 | 60.37 | -352.8 | 17.30 | 44.20 |
| Observations | 18,817 | 111,301 | 111,301 | 12,232 | 122,708 | 7,410 |
| R-squared | 0.033 | 0.008 | 0.016 | 0.048 | 0.022 | 0.047 |

DRD method using a 6th-order polynomial. Data: CPS 1992-2013. Includes time/state FE and state trends.
** 0.01, * 0.05, + 0.1. Robust standard errors clustered at the state level in parentheses.

Table 5: Business Quality (Difference-in-RD)

| Dependent variable: | Incorp. | % HH inc. from bus. | Workers | Wage earners | Weeks | Hours |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Treat \cdot Post$ | 0.0148** | 0.0419** | 0.191** | 0.175** | 2.175** | 0.623** |
| | (0.00150) | (0.0102) | (0.0185) | (0.0181) | (0.154) | (0.0887) |
| Pre-policy avg. | 0.0408 | 0.362 | 1.727 | 1.656 | 45.89 | 39.93 |
| LATE | 36.20 | 11.56 | 11.08 | 10.55 | 4.739 | 1.561 |
| Treated avg. | 0.0138 | 0.417 | 1.286 | 1.227 | 41.86 | 37.78 |
| TOT | 107.2 | 10.03 | 14.88 | 14.25 | 5.194 | 1.650 |
| Observations | 481,678 | 57,290 | 481,678 | 481,678 | 452,090 | 452,090 |
| R-squared | 0.040 | 0.048 | 0.189 | 0.175 | 0.094 | 0.053 |

DRD using 6th-order polynomial. CPS 1992-2013. Includes time/state FE and state trends.
** 0.01, * 0.05, + 0.1. Robust standard errors clustered at the state level in parentheses.

seem to be coming from households with previously employed members rather than the unemployed, though the marginal effects for the unemployed are higher if statistically insignificant.

Another important concern is the quality of these new businesses: are they high-quality ventures that contribute to job growth and innovation, or household side-projects that supplement income but do not hire employees? Without more detailed data on payroll, receipts or long-term survival rates it is difficult to answer this question, but Table 5 in the appendix provides some suggestive evidence. Researchers often use incorporation as a measure of entrepreneur "seriousness" or quality (Levine and Rubinstein, 2013), so Table 5 repeats the specifications above using a household's ownership of an incorporated business as the outcome variables. The effects on incorporation are large: SCHIP caused a 1.5 percentage point increase in incorporation ownership, a 36% increase from the pre-policy baseline (4.1% of households had an incorporated firm before the policy). Since the overall increase in self-employment was only 15%, SCHIP disproportionately encouraged people to open incorporated businesses, shifting the overall composition of firms toward more "serious" ventures.

Table 5 also looks at the percentage of income that self-employed households derive from their businesses as another gauge for seriousness. The business share of income rose by 4.2 percentage points as a result of SCHIP, a 12% increase from the baseline. Table 5 also shows the change in a household's extensive and intensive labor supply margins. More than one in ten households added a working adult or wage earner; combined with the fact that household size was unchanged, this suggests SCHIP induced some individuals who were previously out of the labor market to enter employment. Households also spent more time working because of the policy: the average employed member worked more than two additional weeks out of the year. Taken as a whole, the findings support the interpretation that SCHIP induced entry of high-quality entrepreneurs who devote a significant amount of their time to running a business.

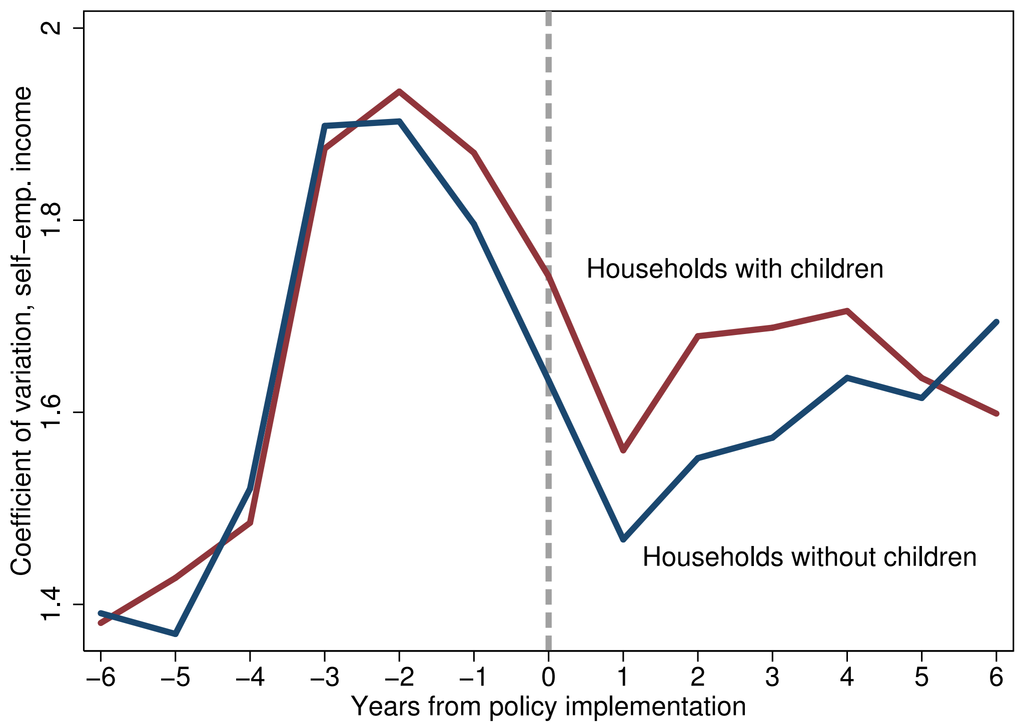
Finally, risk-taking is closely related to the quality and innovative nature of businesses. Are these new businesses high-risk ventures? While this question cannot be answered at the individual level without a longer panel on business income, aggregate relationships provide some interesting results. Figure 2 plots the coefficient of variation—the standard deviation divided by mean—for self-employment income by the number of years since SCHIP was enacted, separately for households with and without children. Using this measure of risk, the two groups track very closely before the policy; for five years following the child health insurance program, however, households with children had more variation in business income relative to childless households. This is consistent with SCHIP enabling households to take on high-risk entrepreneurial ventures they would otherwise have avoided, which suggests these new firms are engaged in more innovative projects.

**Mechanisms.**

How does public health insurance eligibility influence entrepreneurship? To answer this question, I focus on two theoretical mechanisms: the *risk channel* and the *credit channel*. Public programs like SCHIP may operate through risk channel by reducing the uncertainty associated with entering self-employment and losing employer-sponsored health insurance, since public insurance protects the family from medical debts (Baicker et al., 2013) and the resulting consumption shocks. These programs may also work through a credit channel by freeing up income for credit-constrained households, which can be used to post as collateral for business loans or fund start-up costs.

In order to isolate the risk channel, I look at differences between households who were previously insured and those who were not. Households with health insurance experience a change in their exposure to risk—they now have a fall-back source of insurance should they lose private coverage—but no change in their budget constraint since they do not immediately receive benefits. Table 6 estimates the treatment effects separately based on whether a household's children were insured in the previous year. Columns (1) and (2) estimate the treatment effect on the subset of the population that previously reported hav-

Figure 2: Business Risk




ing health insurance for at least some of their children; columns (3) and (4) do the same for fully uninsured households. The results indicate that both previously insured and uninsured children benefited from the program, but that unsurprisingly it was the uninsured who enjoyed the biggest gains in next-year coverage. However, the self-employment increase is largely driven by entrepreneurship among previously *insured* households, who face no immediate changes to their budget constraints but are now eligible for publicly subsidized coverage should they lose their current plan. This is an important piece of evidence in support of the risk channel: the increase in self-employment cannot be explained solely by changes to the budget constraint, because many families move into business ownership even when they do not benefit directly from SCHIP. This is consistent with the risk channel, in which eligibility for public benefits lowers a household's exposure to risk whether or not they avail themselves of the program's benefits. The results suggest the risk channel plays an important role in the effect of health insurance on business formation.

Isolating the credit channel is more difficult without detailed information on the constraints that households face. However, higher-income households are less likely to face credit constraints than poorer ones, so heterogeneity in the treatment effect along income

Table 6: Mechanisms (Difference-in-RD)

| Children were previously: | Uninsured | | Insured | | | | |
|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Dependent variable: | Health ins. | Self-emp. | Health ins. | Self-emp | Self-emp. | Self-emp. | Self-emp. |
| $Treat \cdot Post$ | 0.0516* | 0.0161 | 0.0314** | 0.0305** | 0.0305** | 0.0151+ | 0.0234** |
| | (0.0237) | (0.0146) | (0.00703) | (0.00549) | (0.00549) | (0.00879) | (0.00498) |
| $Treat \cdot Post \cdot Income/10k$ | | | | | -0.00489+ | | |
| | | | | | (0.00268) | | |
| $Treat \cdot Post \cdot Children$ | | | | | | 0.00505+ | |
| | | | | | | (0.00253) | |
| $Treat \cdot Post \cdot Risk$ | | | | | | | -0.0415* |
| | | | | | | | (0.0186) |
| Pre-policy avg. | 0.572 | 0.179 | 0.924 | 0.150 | — | — | — |
| LATE | 9.021 | 9.010 | 3.398 | 20.42 | — | — | — |
| Treated avg. | 0.521 | 0.146 | 0.861 | 0.107 | — | — | — |
| TOT | 9.903 | 11.02 | 3.645 | 28.42 | — | — | — |
| Observations | 11,185 | 11,279 | 104,905 | 105,297 | 130,118 | 481,678 | 481,678 |
| R-squared | 0.046 | 0.031 | 0.031 | 0.025 | 0.030 | 0.025 | 0.023 |

DRD method using a 6th-order polynomial. Data: CPS 1992-2013. Includes time/state FE and state trends.
** 0.01, * 0.05, + 0.1. Robust standard errors clustered at the state level in parentheses.

lines may hint at the role of credit. Column (5) interacts $Treat_{it}Post_{st}$ with the household's total income in the previous year (in tens of thousands of dollars); the negative coefficient means poorer households had larger increases in the self-employment rate as a result of SCHIP, which is consistent with the program alleviating credit constraints. The semi-elasticity of income on the treatment effect is relatively large: an increase in household income by one standard deviation—about $6,800—decreases the marginal effect of SCHIP eligibility on self-employment by 1.09 percentage points, more than one-third of the baseline effect. However the coefficient is only marginally significant, and the result is also consistent with a model in which risk aversion is decreasing in income, so that the treatment effects are largest in the most risk averse (and poorest) households.

Column (6) interacts each of the variables with the number of children in the household. Larger households may be more credit-constrained than smaller ones, though the effective value of SCHIP coverage is larger for big families, so the comparisons aren't exact. The results suggest larger families benefited more from the policy than smaller ones, and the effects are large (an increase in household size by one choice—close to the standard deviation—raises the treatment effect by one-third) though only marginally significant. This could be the result of either canceling effects from the simultaneously more binding credit constraint and bigger benefit pay-out, or from insurance and self-employment decisions that are driven by high-cost, low-risk events—which are relatively invariant to household size—rather than routine care.

Finally, column (7) interacts $Treat_{it}Post_{st}$ with a measure of self-employment risk: the coefficient of variation in self-employment income before the policy was enacted, calculated at the state level. The negative estimate means treatment effects were largest in states with low levels of self-employment income volatility. This might be the result of more risk averse people to entering entrepreneurship as a result of SCHIP, so long as people of similar risk aversion tend to reside in the same states. Alternatively, it could be the case that the reduction in risk was insufficient to induce large changes in self-employment in high-risk states—perhaps because there is a threshold level of risk above which little entrepreneurship takes place—whereas the reduction in lower-risk states was more effective. Overall, Table 6 finds the strongest evidence for the risk channel, with some suggestive evidence that the credit channel matters as well.

# 4 Monte Carlo Analysis and Robustness

Does difference-in-RD generally perform better than traditional RD and DID designs, or is this a quirk of the data and policy? Covariate balance is a useful indicator of randomization quality but provides no information about bias or whether hypothesis testing provides correct inference. In this section I use Monte Carlo analysis to directly control the data-generating process and identify whether the estimator is unbiased in finite samples and if test statistics are correctly sized in a more general setting.

I use the following Monte Carlo procedure:

1. First I draw 10,000 observations of the forcing variable $X \sim F(\theta)$, where $F(\cdot)$ is a probability distribution with parameter vector $\theta$; in practice I use the empirical income distribution from the 2010 and 2011 CPS, but results for the uniform and standard normal distributions are very similar.

2. Next, I assign post-policy status randomly, so that half of the observations fall in the "post" period (with the empirical income distribution I call 2010 the pre-policy year and draw 5,000 observations from each year).[13]

3. I then construct an outcome variable $Y = g(X|\xi) + \varepsilon$ using the mapping $g(\cdot|\xi)$ with parameter vector $\xi$; for the results below I run a non-stochastic version ($\varepsilon \equiv 0$) with polynomials of order $k = [1, ..., 7]$, where the coefficients $\xi_k \sim N(0, 1)$; the natural logarithm; and a sinusoidal function with a period equal to one-fifth of the domain of $X$.

4. The next step is to randomly assign treatment status, which I do by randomly choosing an observation $i$ and assigning treatment to all individuals $j$ such that $x_j \leq x_i$—this is equivalent to choosing a threshold $\bar{x} \sim F(\theta)$—and then subtract $x_i$ from $X$.

5. I then add a treatment effect $\eta$ so that $Y = g(X|\xi) + \eta \cdot \mathbf{1}[x_j \leq x_i] + \varepsilon$; when modeling Type 1 Error, $\eta = 0$, and for Type 2 Error I use a range of $\eta \in [.01, 1]$ where the units are standard deviations of $Y$.

6. Finally, I use several different estimators $T(\cdot)$—DID, RD and differenced RD—and a range of bandwidths to generate predictions $\hat{Y}$.

Each draw $m$ of the $M$ total draws produces an estimate $\hat{Y}_m$, which generate the following statistics:

[13]Since RD uses only post-policy data, this procedure might unfairly stack the deck against RD since it will use half as many observations. I develop two remedies for this. First, I double the number of observations sampled for RD, so that each estimator uses a sample size of 10,000 on average; these are the results shown below. Second, I use an algorithm to identify the differenced RD bandwidth size whose number of observations would match the observations for a given RD bandwidth (usually around half, depending on the density of $X$). The results, which apply only to the local linear method and are not shown here, are very similar.

- Bias: E$[(\hat{Y}-Y)/\sigma_y]$, which has the sample analogue $\frac{1}{M}\sum_{m=1}^{M}(\hat{Y}_m - Y)/\sigma_y$;
- Root-mean-squared error (RMSE): $\sqrt{\text{E}\left[\left(\frac{\hat{Y}-Y}{\sigma_y}\right)^2\right]}$, which is calculated as $\sqrt{\frac{1}{M}\sum_{m=1}^{M}\left(\frac{\hat{Y}_m-Y}{\sigma_y}\right)^2}$;
- Type 1 Error: $p\left[\left(\hat{Y}|x_i \leq x_j\right) \neq \left(\hat{Y}|x_i > x_j\right)|T(\cdot),\alpha,\eta=0\right]$, which has the sample analogue
$$\frac{1}{M}\sum_{m=1}^{M}\mathbf{1}\left[\left(\hat{Y}_m|x_i \leq x_j\right) \neq \left(\hat{Y}_m|x_i > x_j\right)|T(\cdot),\alpha,\eta=0\right];$$
- Statistical Power: $p\left[\left(\hat{Y}|x_i \leq x_j\right) = \left(\hat{Y}|x_i > x_j\right)|T(\cdot),\alpha,\eta\neq 0\right]$, which is calculated as
$$\frac{1}{M}\sum_{m=1}^{M}\mathbf{1}\left[\left(\hat{Y}_m|x_i \leq x_j\right) = \left(\hat{Y}_m|x_i > x_j\right)|T(\cdot),\alpha,\eta\neq 0\right].$$

If the hypothesis testing procedure is correct, the estimated Type 1 Error should be equal to the size of the test, $\alpha$ (I use a 5% level of significance below). Table **??** shows the results of this procedure when there is no treatment effect ($\eta = 0$), using the polynomial versions of RD and differenced RD and the empirical income distribution for $X$.

Panel A pretends as though the econometrician knows the exact functional form of the conditional expectation, meaning when $g(\cdot)$ is quadratic, the estimators include $x$ and $x^2$ terms; when $g(\cdot)$ is logarithmic, the regression includes $\ln(x)$; and so on. This provides a "best case scenario" in which error arises only from the research design itself rather than functional form misspecification. Since this is unrealistic in most settings, Panel B imposes a cubic approximation of $g(\cdot)$, which is one of the more common strategies used by researchers deploying RD. Finally, Panel C presents the results using difference-in-differences as an additional benchmark. The stars denote statistically-significant differences from theoretical values (0.05 in the case of Type 1 Error and zero in the case of bias).

Regression discontinuity dramatically over-rejects the null hypothesis in these simulations, with Type 1 Error rates ranging from 30% to 99% depending on the degree of non-linearity in the conditional expectation. This is true even when the econometrician models the exact functional form; for example, a random quadratic relationship between $Y$ and $X$ will find a "treatment" effect significant at the 5% level 43% of the time. The inaccuracy of the test statistic seems to be increasing with the non-linearity of $g(\cdot)$, and when $Y$ has a periodic relationship to $X$—such as the sine function—regression discontinuity nearly always finds a significant effect. Hypothesis testing under the differenced RD estimator, by contrast, produces a nearly correctly-sized test, with Type 1 Error rates

Table 7: Monte Carlo, No Treatment Effect (Polynomial Method)

| | | $g(x\|\xi)$ | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Polynomial of order $k =$ ... | | | | | | |
| Panel A: Exact $g(\cdot)$ known | | 1 | 2 | 3 | 5 | 7 | Log | Sin |
| Type 1 Error: | RD | 0.29** | 0.43** | 0.47** | 0.46** | 0.33** | 0.69** | 0.99** |
| | Differenced | 0.06 | 0.07* | 0.10** | 0.10** | 0.06 | 0.81** | 0.19** |
| Bias, $sd(y)$: | RD | -0.00 | 0.00 | 0.00 | 0.00 | -0.00 | 0.00 | 0.92** |
| | Differenced | 0.00* | 0.00 | 0.00 | -0.00 | 0.00 | 0.00 | -0.01** |
| RMSE: | RD | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 1.07 |
| | Differenced | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.14 |
| Panel B: Cubic approximation to $g(\cdot)$ | | | | | | | | |
| Type 1 Error: | RD | 0.30** | 0.40** | 0.47** | 0.97** | 0.98** | 0.96** | 0.98** |
| | Differenced | 0.06 | 0.09** | 0.10** | 0.32** | 0.11** | 0.19** | 0.24** |
| Bias, $sd(y)$: | RD | -0.00 | -0.00 | -0.00 | 0.08** | 0.03+ | -0.17** | -0.41** |
| | Differenced | 0.00 | 0.00 | 0.00 | 0.02** | -0.00 | 0.01** | 0.01 |
| RMSE: | RD | 0.00 | 0.00 | 0.00 | 0.40 | 0.60 | 0.64 | 1.91 |
| | Differenced | 0.00 | 0.00 | 0.00 | 0.11 | 0.09 | 0.07 | 0.19 |
| Panel C: Difference-in-Differences | | | | | | | | |
| Type 1 Error: | | 0.08** | 0.08** | 0.13** | 0.23** | 0.58** | 0.10** | 0.21** |
| Bias, $sd(y)$: | | -0.01 | -0.15** | -0.19** | -0.31** | -0.11** | -0.00* | -0.01* |
| RMSE: | | 0.33 | 1.00 | 1.14 | 1.36 | 0.33 | 0.05 | 0.15 |

** 0.01, * 0.05, + 0.1. Robust standard errors calculated using the number of simulations (omitted for brevity).
Tests are for difference from $\alpha$ (for Type 1 Error) or zero (for bias).
$X \sim$ empirical income distribution (2010-2011 CPS). Each cell represents 1,000 draws of 10,000 observations.

generally much lower than their RD counterparts and often not significantly different from 5%.

Incorrectly-sized tests may be relatively unimportant if the estimator is unbiased: researchers will find significant policy effects but conclude the point estimates are so small as to be indistinguishable from zero. To address these concerns, Table **??** also presents the average bias from the Monte Carlo simulations in units of $\sigma_y$. In Panel A, RD is generally unbiased except in the case of the sine function, where regression discontinuity finds a spurious effect of nearly one standard deviation almost every draw. This is particularly relevant to contexts with regular fluctuations—for example, a time series subject to a business cycle—where the researcher runs a discontinuity design based on the timing of a policy. This over-rejection and bias is present even when the regression directly controls for the functional form of the conditional expectation.

More realistically, however, the researcher does not know the exact nature of the relationship $g(\cdot)$ and approximates this function with a cubic polynomial in $X$. Panel B repeats the Monte Carlo simulations using this estimation strategy. First notice that the Type 1 Error rates are almost uniformly larger using this method, especially when the $g(\cdot)$ has a higher degree of non-linearity than is controlled for in the regression. However, while the error rates also increase for differenced version of RD, the rise is much smaller than for traditional RD, with reductions in error of around 75% on average.

Second, RD begins exhibiting statistically significant amounts of bias for conditional expectations of order higher than three. Even a fairly un-exotic quintic function nearly always finds a "treatment" effect of around 8% of a standard deviation, which would be a notable result in the education or labor economics literature. A logarithmic relationship between $Y$ and $X$ causes RD to identify a 0.17 standard deviation *decrease* in the outcome variable for the "treated" group, an underlying bias that would mask even extremely large policy effects. The falsification check version of RD, on the other hand, completely eliminates this bias, with coefficient estimates rarely different than zero and never more than 3% of a standard deviation. The difference-in-difference estimates in Panel C are biased even at low levels of non-linearity, and while Type 1 Error rates are lower than for RD they are generally higher than the differenced version.

Finally, Table **??** also presents the root-mean-squared error (RMSE) from the simulations. While comparing efficiency of the estimators is difficult when one is not unbiased, the uniformly lower RMSE of differenced RD relative to traditional RD suggests the method carries some improvements in precision.

Using only the polynomial method is somewhat misleading, since most researchers supplement this method with local linear estimation. This conforms more closely to the

Table 8: Monte Carlo, No Treatment Effect (Local Linear Method)

| | | Bandwidth (in terms of $\sigma_x$) | | | | | |
|---|---|---|---|---|---|---|---|
| Panel A: Quadratic $g(\cdot)$ | | 1 | 0.5 | 0.25 | 0.1 | 0.05 | 0.01 |
| Type 1 Error: | RD | 0.98** | 0.96** | 0.84** | 0.54** | 0.48** | 0.48** |
| | Differenced | 0.08** | 0.07* | 0.06 | 0.07** | 0.09** | 0.11** |
| Bias, $sd(y)$: | RD | -0.005** | -0.002** | -.0002** | -0.000** | -0.000** | -0.000** |
| | Differenced | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| RMSE: | RD | 0.038 | 0.008 | 0.002 | 0.000 | 0.000 | 0.000 |
| | Differenced | 0.005 | 0.002 | 0.000 | 0.000 | 0.000 | 0.000 |
| Panel B: Logarithmic $g(\cdot)$ | | | | | | | |
| Type 1 Error: | RD | 0.98** | 0.98** | 0.92** | 0.55** | 0.47** | 0.50** |
| | Differenced | 0.07* | 0.06+ | 0.07* | 0.07* | 0.08** | 0.10** |
| Bias, $sd(y)$: | RD | -0.071** | -0.030** | -0.007** | -0.001** | -0.000 | 0.000** |
| | Differenced | 0.001 | 0.002** | 0.001+ | -0.000 | -0.000 | 0.000 |
| RMSE: | RD | 0.349 | 0.158 | 0.057 | 0.009 | 0.002 | 0.000 |
| | Differenced | 0.023 | 0.020 | 0.017 | 0.002 | 0.001 | 0.000 |
| Panel C: Sinusoidal $g(\cdot)$ | | | | | | | |
| Type 1 Error: | RD | 0.97** | 0.96** | 0.80** | 0.56** | 0.49** | 0.49** |
| | Differenced | 0.08** | 0.08** | 0.07* | 0.07* | 0.08** | 0.10** |
| Bias, $sd(y)$: | RD | -0.067** | 0.014** | 0.004** | 0.000** | 0.000** | 0.000** |
| | Differenced | -0.005* | 0.000 | 0.000 | 0.000 | 0.000** | 0.000** |
| RMSE: | RD | 0.760 | 0.149 | 0.023 | 0.003 | 0.001 | 0.000 |
| | Differenced | 0.078 | 0.026 | 0.007 | 0.002 | 0.000 | 0.000 |
| Obs. (approx.) | | 7,130 | 4,488 | 2,492 | 1,046 | 527 | 110 |

** 0.01, * 0.05, + 0.1. Robust s.e. calculated using number of simulations (omitted for brevity).
Tests are difference from $\alpha$ (Type 1 Error) or zero (bias). $X \sim$ empirical income dist. 2010-2011 CPS.
"Obs. (approx.)" is observations used in regressions, averaged over simulations and $g(\cdot)$ functions.
Each cell represents 1,000 draws of 10,000 observations.

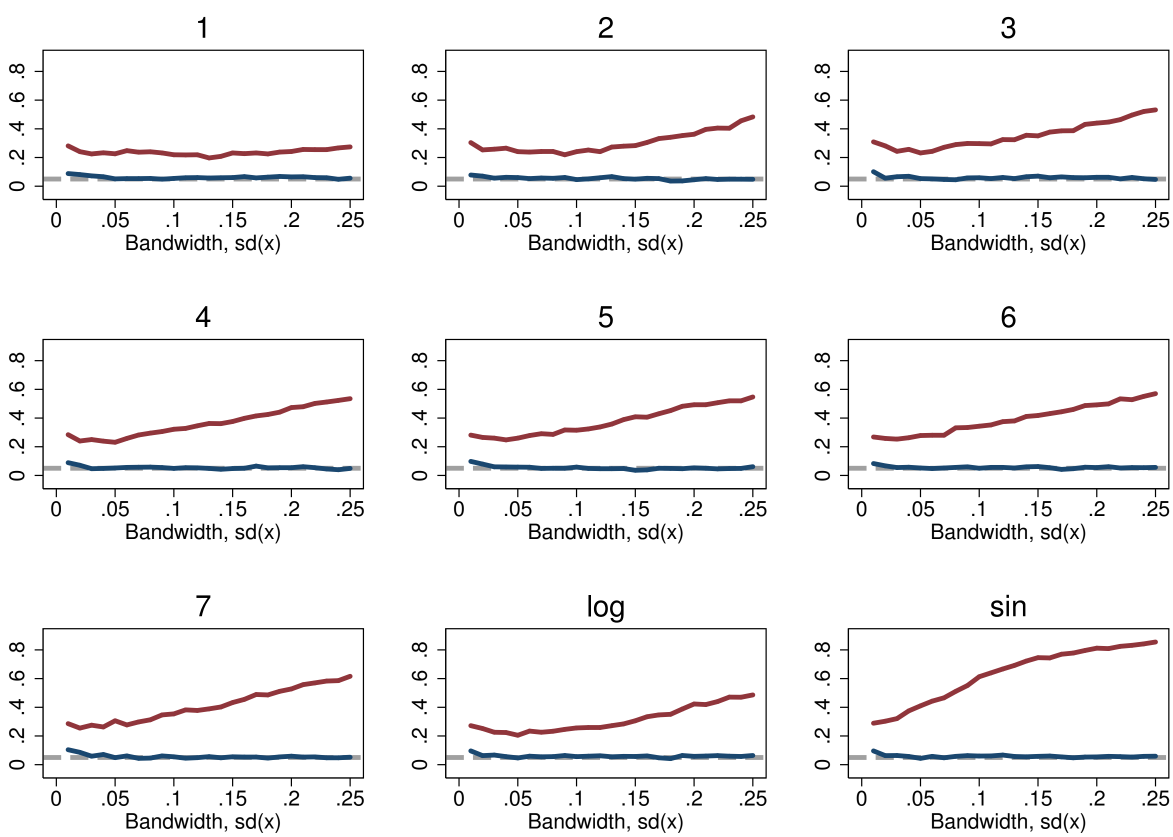
underlying logic of RD, which is that the estimator is unbiased as the bandwidth asymptotically approaches zero. Does this mean that test statistics are correctly sized as the bandwidth narrows? Table **??** shows the Monte Carlo results for a local linear estimator of varying bandwidths for several functional forms of $g(\cdot)$, where bandwidths are measured in standard deviations of $X$. While the bias of RD tends to zero as the bandwidth decreases, Type 1 Error is significantly different than zero and actually begins to increase as the sample size becomes very small. Difference-in-RD, in contrast, rejects the null hypothesis correctly nearly every time while retaining the unbiasedness of RD. In particular, while RD is significantly biased for large bandwidths, differenced RD remains unbiased even when the econometrician chooses an inappropriately large sample for local linear estimation. In this sense, this method is "researcher-proof", meaning the econometrician cannot selectively present bandwidth levels with rejection rates and coefficient estimates that confirm his or her prior.

This bandwidth-choice invariance is a striking feature of Figure 3, which shows the Type 1 Error rates for the estimators using each of the functions $g(\cdot)$ from Table **??**. The horizontal axis for each graph is the bandwidth in terms of standard deviations of $X$, starting at 0.1 (less than 1% of the data) and ending at 0.25 (about 15% of the total observations). While the rates decline monotonically for RD as the bandwidth closes, the levels are always significantly different from 5%—represented by a horizontal line—and begin to rise close to zero. Adding a falsification check directly in to RD, however, uniformly reduces Type 1 Error rates in a way that is generally orthogonal to the bandwidth choice, with estimates that are generally not significantly different than $\alpha$.

Testing the estimator's performance when the treatment effect is non-zero is slightly more difficult, since it requires taking into account two dimensions: treatment effect size $\eta$ (measured in $\sigma_y$) and bandwidth size (measured in $\sigma_x$). A helpful visualization of this mapping is a heat-map, which varies in color depending on the underlying value at a given coordinate. Panel (a) of Figure 4 shows the bias as a fraction of the treatment effect $\eta$ of the basic and difference-corrected RD, for a wide range of bandwidths and treatment effects, using a 4th-order polynomial as the $g(\cdot)$ function (other relationships look very similar). Regression discontinuity produces biased estimates even in relatively small bandwidths, and the problem is particularly severe when treatment effects are small. For example, a true treatment effect of 0.02 $\sigma_y$ estimated with a bandwidth of 0.15 $\sigma_x$, which uses about 15% of the observations, would be overstated by 50%. Even a more commonly sized policy effect for labor or education—something on the order of 0.1 $\sigma_y$—would be biased upwards by 10%.

The difference-corrected RD estimator, shown in the right panel, reduces much of this

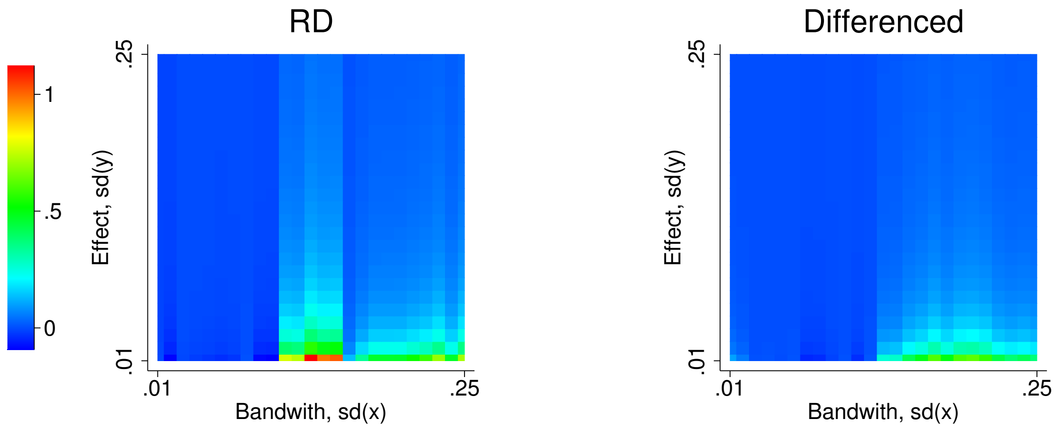
Figure 3: Monte Carlo, Type 1 Error



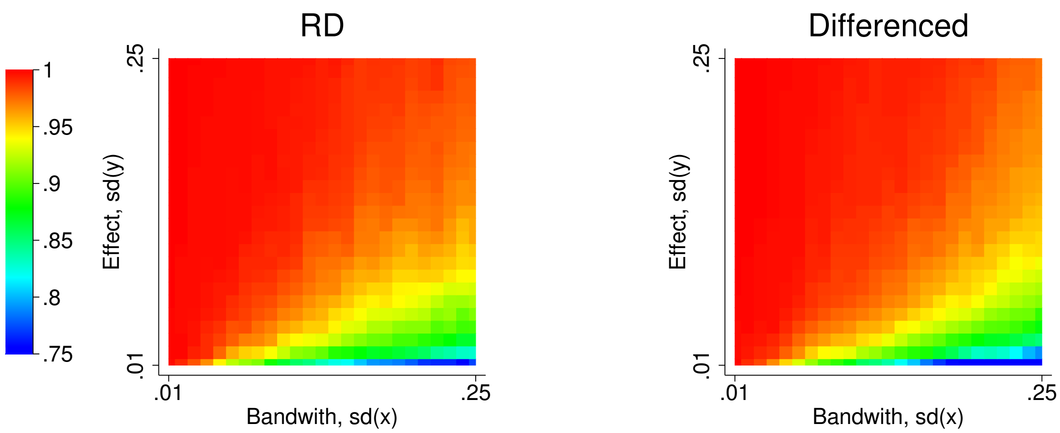

Note: Titles refer to functional form of $g(\cdot)$. Top (red) line is RD; bottom (blue) line has the difference correction; dashed (grey) reference line is at 5%.

Figure 4: Monte Carlo, Positive Treatment Effect (4th Order $g(\cdot)$, Local Linear Method)

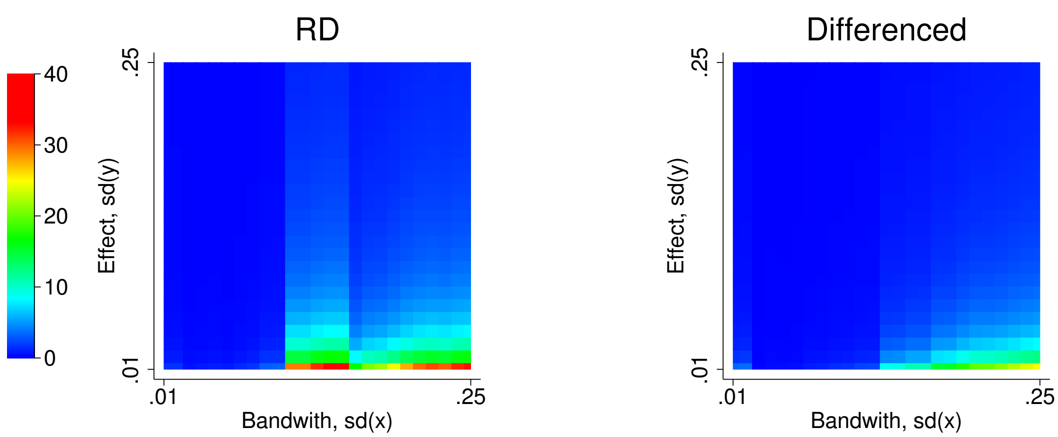
(a) Bias




(b) Power




(c) RMSE



bias. The range of bias values drops by 45%, and much of the distortion is confined to extremely small effects and large bandwidths. And while 9% of the bias values in the RD graph are significantly different from zero at the 5% level (generally on the small bandwidth end of the plane) none of the bias values on the righthand side are significant.

Panel (b) of Figure 4 shows the statistical power (the counterpart to Type 2 Error) of basic and corrected RD for a range of effect sizes and bandwidths, over several different $g(\cdot)$ functions. Not surprisingly, RD is underpowered when the effect size is very small, and this problem is exacerbated when bandwidths are large. Differenced RD has slightly lower power for large bandwidths, but the effects are minimal and the estimator has similar Type 2 Error rates to RD.

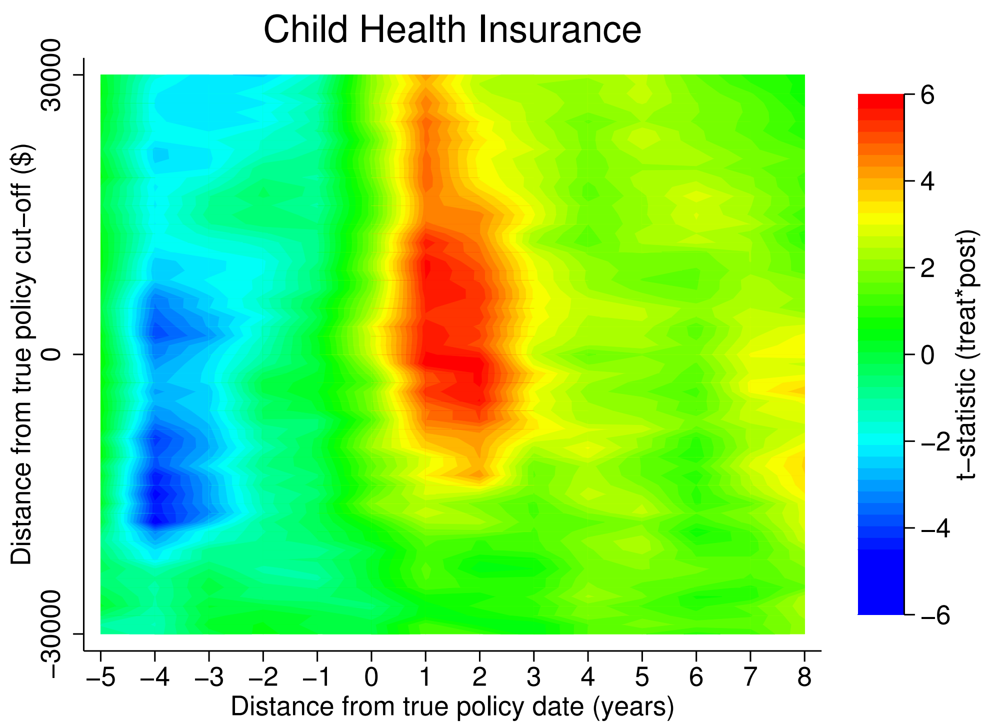
Finally, panel (c) of Figure 4 plots the root-mean-squared error of the two estimators, again using a 4th order polynomial for $g(\cdot)$ and a range of bandwidths and treatment effects. Regression discontinuity estimates are very imprecise out when policy effects are small, irrespective of the bandwidth. The root-mean-squared error of RD with the difference correction is much lower, with the range of values decreasing by more than 25%. While efficiency comparisons aren't useful when both estimators are not unbiased, Figure 4 suggests some gains in precision from including a falsification test into RD.

**Additional falsification and robustness tests.**

An important question when evaluating a policy is whether the research design picks up shifts in the underlying populations that would have occurred without the program. The RD framework provides a natural falsification test: if treatment status is assigned using different points in the forcing variable's domain, the largest jump occur at the true policy threshold. Treatment effects estimated at non-policy levels should pick up only small and insignificant differences.

Incorporating a falsification check into RD adds another dimension of potential testing, since the date the policy was enacted is also a variable of interest. Assigning post-policy status using fake dates should produce insignificant results. However, if the interaction term is significant only when the true policy date is used, the results are more plausibly uncontaminated by violations of the parallel trends assumption. Because the differenced version of RD has two dimensions of falsification—threshold levels and policy years—every combination of these points forms a falsification test. A region of large t-statistics only in the neighborhood of the true policy's date and threshold would support the claim that the estimator isn't picking up spurious changes in the population (see Chay and Munshi, 2013 for a similar method).

Figure 5: Falsification Tests




Contour maps are a natural way to visualize this relationship. Figure 5 shows the results using child health insurance as the outcome variable, with each coordinate representing the falsification threshold and year, and the "height" of each point representing the t-statistic associated with the $\beta_{DRDpoly}$ term.[14] The largest "peak" lies in the first year after the policy was enacted, along a "ridge" that has several apices in the neighborhood of the actual income threshold.

The lower panel shows the self-employment results; the highest peak occurs in the year of policy adoption just below the actual income cutoff.[15] The fact that the most significant increase in child health insurance coverage occurred one year after SCHIP was enacted could be the result of several things. First, there could be recall bias in answering CPS health insurance variables in the CPS, and there is some evidence that income and employment questions are answered based on the household's experiences over the previous calendar year, whereas health insurance variables are point-in-time measures (Davern et al., 2007). Alternatively enrollment might have been low in the first year of the program while the government built awareness.

The relationship between the outcome variable and the forcing variable might be dif-

[14]The sample is restricted sample to a rolling 5-year window.

[15]This is consistent with the small but persistent under-reporting of income in the CPS Weinberg (2006).

ferent for the treated, perhaps because treatment induces changes in economic behavior unrelated to health insurance. This means the conditional expectation changes slope across the treatment threshold; including $Treat_{it} \cdot g(Rule_{it})$ or $Treat_{it}Rule_{it}$ terms into the specifications from Table 1 allows for this flexibility. Table A1.2 in the appendix shows the sensitivity to allowing this difference, and the treatment effects are all robust to this change. A popular method for dealing with selection on observables is propensity score matching, which explicitly balances observables to generate matched groups with the same propensity to be treated. Table A1.3 in the appendix shows the result of including a propensity score (using $Treat_{it}$ as the treatment variable and the covariates discussed earlier as the predictors) in the regressions. The findings are all robust to including a propensity score, which is strongly and negatively correlated with the outcomes of interest.

# 5 Discussion

The social safety net is a crucial source of support for millions of Americans, yet little attention has been paid to its effect on business ownership. This paper tests whether the State Child Health Insurance program impacted firm formation. I develop a quasi-experimental research design to identify the causal effect of the policy and find that SCHIP reduced the number of households with uninsured children by nearly 40%. The program also increased the self-employment rate by 15%, the number of incorporated firms by 36%, and the share of household income derived from self-employment by 12%, indicating that these are relatively high-quality ventures. The increase is driven largely by a 12% rise in firm birth rates and a 26% increase in the birth rate of incorporated firms. I also document a large increase in labor supply, both on the extensive margin (more wage-earners in affected households) and the intensive margin (treated households worked an additional two weeks out of the year). There is no evidence that observable characteristics differ between treatment and control groups, and the results survive a number of falsification and robustness tests. Monte Carlo analysis strongly suggest the difference-in-RD estimator outperforms more traditional estimators in terms of bias and Type 1 Error, particularly when bandwidths are large and treatment effects are small.

Public health insurance may encourage self-employment by protecting households from the risk of catastrophic medical debt if they start a business and lose coverage (the *risk channel*) and by freeing up assets for use as loan collateral or start-up funds (the *credit channel*). I find the strongest evidence for the risk channel: households who did not enroll in SCHIP but were otherwise eligible saw significant increases in self-employment, even though their budget constraint had not shifted. The importance of uninsured risk in constraining firm formation and expansion is consistent with work in the development literature on barriers faced by low-income firms (Bianchi and Bobba, 2013; Karlan et al., 2012).

The link between public health insurance and entrepreneurship is particularly salient in light of recent healthcare reform and the Affordable Care Act, but the theoretical mechanisms apply to a more general class of programs. To the extent that entrepreneurs contribute to innovation, job creation or economic growth, these findings strongly suggest social insurance programs provide spillover benefits by affecting the supply of firms.

# References


**Aghion, Philippe and Peter Howitt**, *Endogenous Growth Theory*, Vol. 1 of *MIT Press Books*, The MIT Press, July 1997.

**Baicker, Katherine, Sarah L. Taubman, Heidi L. Allen, Mira Bernstein, Jonathan H. Gruber, Joseph P. Newhouse, Eric C. Schneider, Bill J. Wright, Alan M. Zaslavsky, and Amy N. Finkelstein**, "The Oregon Experiment: Effects of Medicaid on Clinical Outcomes," *New England Journal of Medicine*, 2013, *368* (18), 1713–1722. PMID: 23635051.

**Bansak, Cynthia and Steven Raphael**, "The Effects of State Policy Design Features on Take Up and Crowd Out Rates for the State Children's Health Insurance Program," Working Papers 0002, San Diego State University, Department of Economics Mar 2005.

**_ and _** , "The State Health Insurance Program and Job Mobility: Identifying Job Lock among Working Parents in Near-Poor Households," Institute for Research on Labor and Employment, Working Paper Series, Institute of Industrial Relations, UC Berkeley May 2005.

**Berger, Mark C., Dan A. Black, and Frank A. Scott**, "Is There Job Lock? Evidence from the Pre-HIPAA Era," *Southern Economic Journal*, 2004, *70* (4), pp. 953–976.

**Berkowitz, Jeremy and Michelle J. White**, "Bankruptcy and Small Firms' Access to Credit," *RAND Journal of Economics*, Spring 2004, *35* (1), 69–84.

**Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan**, "How Much Should We Trust Differences-in-Differences Estimates?," *The Quarterly Journal of Economics*, February 2004, *119* (1), 249–275.

**Bianchi, Milo and Matteo Bobba**, "Liquidity, Risk, and Occupational Choices," *Review of Economic Studies*, 2013, *80* (2), 491–511.

**Blanchflower, David G and Andrew J Oswald**, "What Makes an Entrepreneur?," *Journal of Labor Economics*, January 1998, *16* (1), 26–60.

**Bruce, Donald and Mohammed Mohsin**, "Tax Policy and Entrepreneurship: New Time Series Evidence," *Small Business Economics*, 2006, *26* (5), pp. 409–425.

**Cagetti, Marco and Mariacristina De Nardi**, "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, October 2006, *114* (5), 835–870.

**_ and _** , "Estate Taxation, Entrepreneurship, and Wealth," *American Economic Review*, March 2009, *99* (1), 85–111.

**Chay, Kenneth and Kaivan Munshi**, "Black Networks After Emancipation: Evidence from Reconstruction and the Great Migration," *Mimeo*, 2013.

**Davern, Michael, Gestur Davidson, Jeanette Ziegenfuss, Stephanie Jarosek, Brian Lee, Tzy-Chyi Yu, and Lynn Blewett**, *A comparison of the health insurance coverage estimates from four national surveys and six state surveys: a discussion of measurement issues and policy implications*, University of Minnesota School of Public Health, State Health Access Data Assistance Center, 2007.

**Decker, Ryan, John Haltiwanger, Ron Jarmin, and Javier Miranda**, “The Role of Entrepreneurship in US Job Creation and Economic Dynamism,” *Journal of Economic Perspectives*, 2014, *28* (3), 3–24.

**Evans, David S and Boyan Jovanovic**, “An Estimated Model of Entrepreneurial Choice under Liquidity Constraints,” *Journal of Political Economy*, August 1989, *97* (4), 808–27.

**Fairlie, Robert W., Kanika Kapur, and Susan Gates**, “Is employer-based health insurance a barrier to entrepreneurship?,” *Journal of Health Economics*, 2011, *30* (1), 146 – 162.

**Fairlie, Robert W, Kanika Kapur, and Susan Gates**, “Job Lock: Evidence from a Regression Discontinuity Design,” Working Papers 201215, School Of Economics, University College Dublin Apr 2012.

**Fan, Wei and Michelle J White**, “Personal Bankruptcy and the Level of Entrepreneurial Activity,” *Journal of Law and Economics*, October 2003, *46* (2), 543–67.

**Garthwaite, Craig, Tal Gross, and Matthew J. Notowidigdo**, “Public Health Insurance, Labor Supply, and Employment Lock,” NBER Working Papers 19220, National Bureau of Economic Research, Inc July 2013.

**Gentry, William M. and R. Glenn Hubbard**, “Entrepreneurship and Household Saving,” NBER Working Papers 7894, National Bureau of Economic Research, Inc September 2000.

_ **and** _ , “Tax Policy and Entrepreneurial Entry,” *The American Economic Review*, 2000, *90* (2), pp. 283–287.

_ **and** _ , “"Success Taxes," Entrepreneurial Entry, and Innovation,” in “Innovation Policy and the Economy, Volume 5” NBER Chapters, National Bureau of Economic Research, Inc, August 2005, pp. 87–108.

**Georgellis, Yannis and Howard J. Wall**, “Entrepreneurship and the policy environment,” *Review*, 2006, (Mar), 95–112.

**Gilleskie, Donna B. and Byron F. Lutz**, “The Impact of Employer-Provided Health Insurance on Dynamic Employment Transitions,” Working Paper 7307, National Bureau of Economic Research August 1999.

**Grembi, Veronica, Tommaso Nannicini, and Ugo Troiano**, “Policy Responses to Fiscal Restraints: A Difference-in-Discontinuities Design,” IZA Discussion Papers 6952, Institute for the Study of Labor (IZA) October 2012.

**Gruber, Jonathan and Brigitte C. Madrian**, "Limited Insurance Portability and Job Mobility: The Effects of Public Policy on Job-Lock," NBER Working Papers 4479, National Bureau of Economic Research, Inc Sep 1993.

_ **and** _ , "Health Insurance, Labor Supply, and Job Mobility: A Critical Review of the Literature," NBER Working Papers 8817, National Bureau of Economic Research, Inc February 2002.

**Gumus, Gulcin and Tracy Regan**, "Self-Employment and the Role of Health Insurance," Discussion Paper 3952, IZA 2013.

**Hamersma, Sarah and Matthew Kim**, "The effect of parental Medicaid expansions on job mobility," *Journal of Health Economics*, July 2009, *28* (4), 761–770.

**Holtz-Eakin, Douglas**, "Health Insurance Provision and Labor Market Efficiency in the United States and Germany," in "Social Protection versus Economic Flexibility: Is There a Trade-Off?" NBER Chapters, National Bureau of Economic Research, Inc, April 1994, pp. 157–188.

_ **, David Joulfaian, and Harvey S. Rosen**, "Entrepreneurial Decisions and Liquidity Constraints," NBER Working Papers 4526, National Bureau of Economic Research, Inc June 1994.

_ **,** _ **, and** _ , "Sticking It Out: Entrepreneurial Survival and Liquidity Constraints," *Journal of Political Economy*, February 1994, *102* (1), 53–75.

_ **, John R. Penrod, and Harvey S. Rosen**, "Health insurance and the supply of entrepreneurs," *Journal of Public Economics*, October 1996, *62* (1-2), 209–235.

**Imbens, Guido W. and Thomas Lemieux**, "Regression discontinuity designs: A guide to practice," *Journal of Econometrics*, February 2008, *142* (2), 615–635.

**Karlan, Dean, Robert Osei, Isaac Osei-Akoto, and Chris Udry**, "Agricultural Decisions after Relaxing Credit and Risk Constraints," Working paper 2012.

**King, Robert G and Ross Levine**, "Finance and Growth: Schumpeter Might Be Right," *The Quarterly Journal of Economics*, August 1993, *108* (3), 717–37.

**Lemieux, Thomas and Kevin Milligan**, "Incentive effects of social assistance: A regression discontinuity approach," *Journal of Econometrics*, February 2008, *142* (2), 807–828.

**Levine, Ross and Yona Rubinstein**, "Smart and Illicit: Who Becomes an Entrepreneur and Does it Pay?," Working Paper 19276, National Bureau of Economic Research August 2013.

**Madrian, Brigitte C**, "Employment-Based Health Insurance and Job Mobility: Is There Evidence of Job-Lock?," *The Quarterly Journal of Economics*, February 1994, *109* (1), 27–54.

**Meh, Cesaire**, “Entrepreneurship, Wealth Inequality, and Taxation,” *Review of Economic Dynamics*, July 2005, *8* (3), 688–719.

**Meh, Cesaire A. and Yaz Terajima**, “Unsecured Debt, Consumer Bankruptcy, and Small Business,” Technical Report 2008.

**Parker, S.C.**, *The Economics of Entrepreneurship* The Economics of Entrepreneurship, Cambridge University Press, 2009.

**Pettersson-Lidbom, Per**, “Does the size of the legislature affect the size of government? Evidence from two natural experiments,” *Journal of Public Economics*, 2012, *96* (3), 269–278.

**Quadrini, Vincenzo**, “Entrepreneurship, Saving and Social Mobility,” *Review of Economic Dynamics*, January 2000, *3* (1), 1–40.

**Robson, Martin T. and Colin Wren**, “Marginal and Average Tax Rates and the Incentive for Self-Employment,” *Southern Economic Journal*, 1999, *65* (4), pp. 757–773.

**Rosenbach, Margo, Marilyn Ellwood, John Czajka, Carol Irvin, Wendy Coupe, and Brian Quinn**, “Implementation of the State Children’s Health Insurance Program: Momentum is increasing after a modest start,” *Prepared for the Centers for Medicare and Medicaid Services, Mathematica Policy Research*, 2001.

**Sasso, Anthony T. Lo and Thomas C. Buchmueller**, “The effect of the state children’s health insurance program on health insurance coverage,” *Journal of Health Economics*, 2004, *23* (5), 1059–1082.

**Schuetze, Herb J**, “Taxes, economic conditions and recent trends in male self-employment: a Canada-US comparison,” *Labour Economics*, 2000, *7* (5), 507 – 544.

**Schumpeter, J.A.**, *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle* A galaxy book. GB 55, Oxford University Press, 1961.

**Weinberg, Daniel H**, “Income data quality issues in the CPS,” *Monthly Lab. Rev.*, 2006, *129*, 38.

# Appendix (for online publication)

Figure A1.1: Density

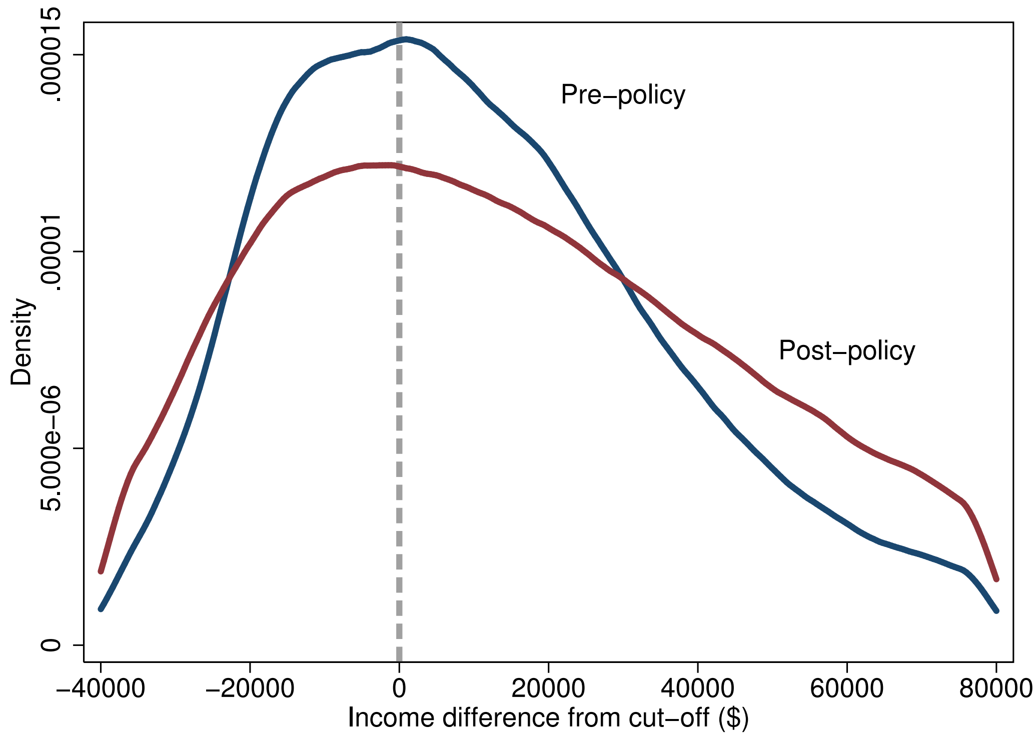


Table A1.1: CPS Summary Statistics

| Variable | Description | Obs. | Mean | Std. Dev. |
|---|---|---|---|---|
| *Child Health Insurance* | Has child covered by health insurance | 494551 | 0.8585747 | 0.34846 |
| *Self-Employment* | Has self-employed member | 1039514 | 0.1268891 | 0.3328489 |
| *Income* | Total household income ($) | 1039514 | 63094.25 | 64607.19 |
| *Black* | Has black family member | 1039514 | 0.1277164 | 0.3337739 |
| *Hispanic* | Has Hispanic family member | 1039514 | 0.1469947 | 0.3541008 |
| *Household Size* | Household size | 1039514 | 2.892961 | 1.505989 |
| *Children* | Number of children | 1039514 | 0.8808222 | 1.168134 |
| *Age of Adults* | Household average adult age | 1039514 | 41.97899 | 11.67393 |
| *Married* | Has married members | 1039514 | 0.5757142 | 0.4942343 |
| *Has Kids* | Has children living at home | 1039514 | 0.4633685 | 0.4986566 |
| *Moved* | Moved in the last year | 1039514 | 0.174142 | 0.3792317 |
| *Urban* | Lives in urban area | 1031598 | 0.7871341 | 0.4093338 |
| *Renter* | Rents domicile | 1039514 | 0.3431277 | 0.4747539 |
| *High School* | Has member with high school degree | 1039514 | 0.8802419 | 0.2902053 |
| *Bachelor's Degree* | Has member with Bachelor's degree | 1006809 | 0.3616908 | 0.428968 |
| *Graduate Degree* | Has member with graduate degree | 1006809 | 0.0899282 | 0.2490771 |
| *Unemployment* | Receives unemployment insurance | 1039514 | 0.079896 | 0.2711322 |
| *Disabled* | Receives disability benefits | 1039514 | 0.015994 | 0.1254521 |
| *Veteran* | Receives veterans benefits | 1039514 | 0.0199016 | 0.1396623 |
| *Medicaid* | Receives Medicaid benefits | 1039514 | 0.1130461 | 0.3166493 |
| *Employed* | Has employed member | 1039514 | 0.8923151 | 0.3099822 |
| *Risk* | Coeff. of variation in SE (state-level) | 1039514 | 1.667465 | 0.1799696 |
| *Self-emp. income* | Percent of income from business | 115256 | 28769.95 | 50912.09 |
| *Workers* | Number of working adults | 1039514 | 1.561936 | 0.9262905 |
| *Wage earners* | Number of wage earners in the household | 1039514 | 1.490459 | 0.9376618 |
| *Weeks* | Weeks worked during the year | 931307 | 46.94887 | 9.693488 |
| *Hours* | Hours worked in typical week | 931307 | 40.09455 | 9.492656 |
| *Incorporated* | Has an incorporated business | 1039514 | 0.0421986 | 0.201042 |
| *Year* | Reference year for household | 1039514 | 2002.239 | 6.319757 |
| *Policy Year* | Year policy was enacted | 1039514 | 1998.221 | 0.6387923 |
| *Threshold* | Policy threshold (% of FPL) | 1039514 | 207.9845 | 45.46071 |
| *Rule* | Last year's *Income* minus *Threshold* | 1039514 | 31251.66 | 62774.42 |
| *Treat* | Treatment status (*rule* $\leq$ 0) | 1039514 | 0.2908205 | 0.4541411 |
| *Post* | Post-policy (year $\geq$ postyear) | 1039514 | 0.7210504 | 0.4484829 |

Table A1.2: Separate Polynomials across Threshold (Difference-in-RD)

| | Child Health Insurance | | | Self-Employment | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Treat \cdot Post$ | 0.0462** | 0.0347** | 0.0193* | 0.0307** | 0.0101* | 0.0180** |
| | (0.00558) | (0.00531) | (0.00803) | (0.00395) | (0.00390) | (0.00576) |
| Pre-policy avg. | 0.861 | 0.826 | 0.827 | 0.140 | 0.125 | 0.127 |
| LATE | 5.369 | 4.207 | 2.331 | 21.93 | 8.118 | 14.18 |
| Treated avg. | 0.783 | 0.793 | 0.810 | 0.102 | 0.116 | 0.117 |
| TOT | 5.904 | 4.377 | 2.381 | 29.93 | 8.762 | 15.34 |
| Bandwidth: | Polynomial | ±$10,000 | ±$5,000 | Polynomial | ±$10,000 | ±$5,000 |
| Observations | 475,546 | 106,034 | 53,566 | 481,678 | 107,487 | 54,295 |
| R-squared | 0.048 | 0.032 | 0.032 | 0.023 | 0.008 | 0.008 |

Time/state FE and state trends included. Data: CPS 1992-2013. ** 0.01, * 0.05, + 0.1.
Polynomial method fits a 6th-order function using the entire dataset, local linear uses data within the bandwidth.
Polynomial specifications include $Treat \cdot g(Rule)$ and local linear specifications include $Treat \cdot Rule$.
Robust standard errors clustered at the state level in parentheses.

Table A1.3: Propensity Score

| | Child Health Insurance | | | Self-Employment | | |
|---|---|---|---|---|---|---|
| | DID | RD (Differenced) | | DID | RD (Differenced) | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| $Treat \cdot Post$ | 0.0332** | 0.0512** | 0.0198* | 0.00983** | 0.0199** | 0.0186** |
| | (0.00532) | (0.00539) | (0.00817) | (0.00344) | (0.00351) | (0.00558) |
| $Prop.\ score$ | 0.0323** | 0.0797** | 0.0234* | -0.197** | -0.179** | -0.214** |
| | (0.00843) | (0.00871) | (0.0105) | (0.00444) | (0.00482) | (0.00811) |
| Pre-policy avg. | 0.861 | 0.861 | 0.827 | 0.140 | 0.140 | 0.127 |
| LATE | 3.858 | 5.948 | 2.397 | 7.029 | 14.26 | 14.65 |
| Treated avg. | 0.783 | 0.783 | 0.810 | 0.102 | 0.102 | 0.117 |
| TOT | 4.242 | 6.541 | 2.448 | 9.593 | 19.47 | 15.85 |
| Bandwidth: | — | Polynomial | ±$5,000 | — | Polynomial | ±$5,000 |
| Observations | 456,756 | 456,756 | 50,942 | 462,760 | 462,760 | 51,648 |
| R-squared | 0.043 | 0.049 | 0.033 | 0.025 | 0.033 | 0.029 |

OLS. Time/state FE and state trends included. Data: CPS 1992-2013.
Polynomial method fits a 6th-order function using the entire dataset, local linear uses data within the bandwidth.
** 0.01, * 0.05, + 0.1. Robust standard errors clustered at the state level in parentheses.

Figure A1.2: Treatment Effects by Bandwidth, RD (Differenced)

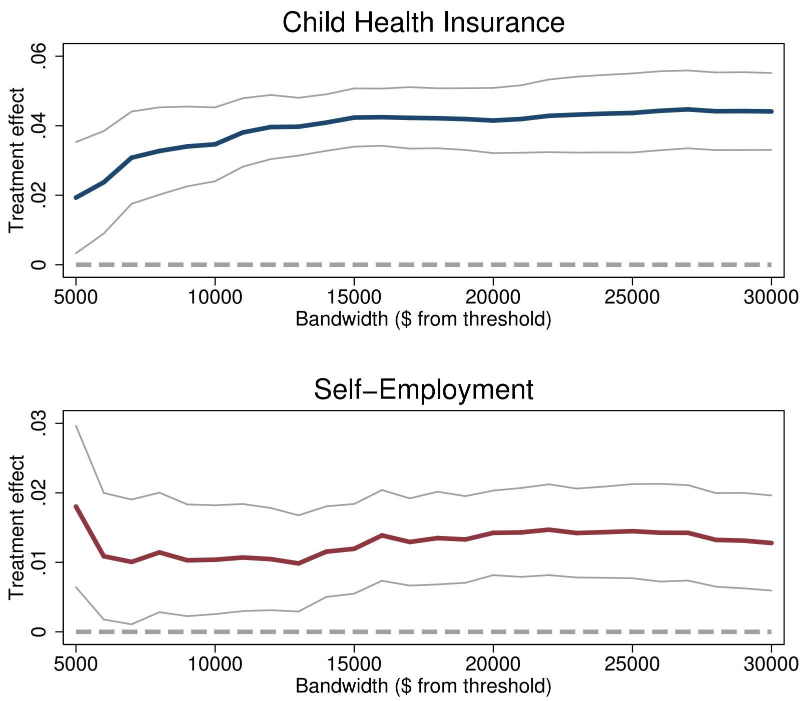
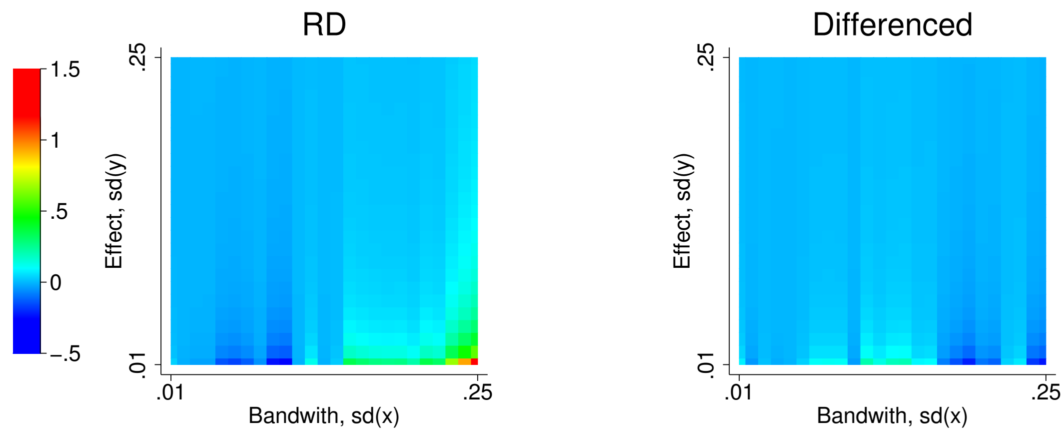
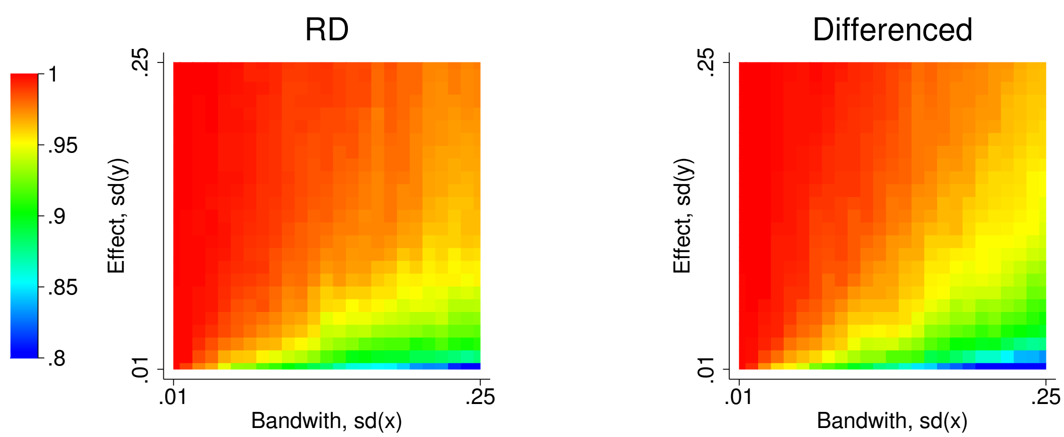


Figure A1.3: Monte Carlo, Positive Treatment Effect (6th Order $g(\cdot)$, Local Linear Method)

(a) Bias




(b) Power




(c) RMSE

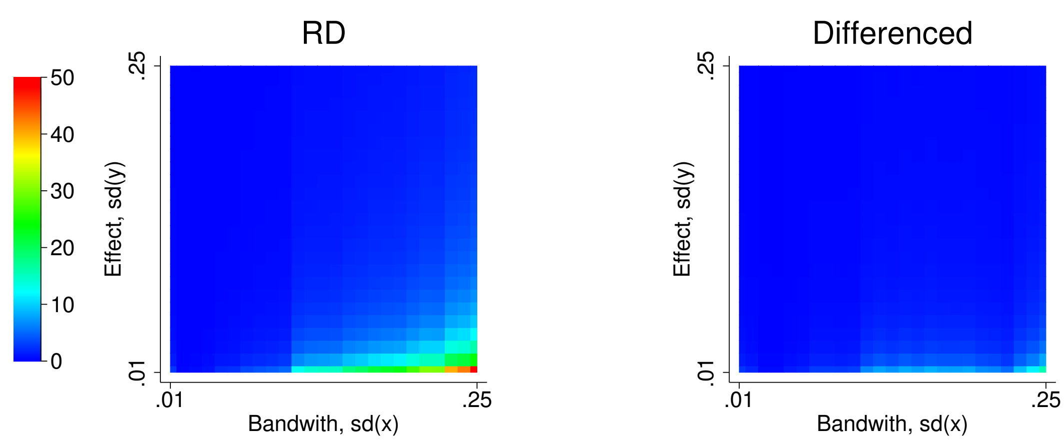


Figure A1.4: Monte Carlo, Positive Treatment Effect (Sinusoidal $g(\cdot)$, Local Linear Method)