

# The Causal Effect of SCHIP Implementation on Insurance Coverage

Soren Dunn, Gloria Huang, Walker Lewis

December 13, 2022

## 1 Introduction

The State Children’s Health Insurance Program (SCHIP) is a federal law which was enacted in August of 1997. Prior to 1997, while all states provided Medicaid coverage for children, many stopped that coverage once the child turned 14. SCHIP provided federal funding to help states voluntarily extend their Medicaid coverage to include eligible people under 19. Because state-level implementations of SCHIP expanded Medicaid coverage for people just under 19 while not altering Medicaid coverage for people just over 19, these programs allow for the natural implementation of a differences-in-differences (DiD) strategy. Levine et al. (2011) used data from the March Current Population Survey (CPS) to estimate SCHIP’s causal effect on youth insurance coverage using Two-Way Fixed Effects. We extend their results by re-estimating the causal effects of SCHIP using a semi-parametric differences-in-differences model with three different treated groups. We find that states that implemented SCHIP at the state level in 1997 and 1998 eventually obtained a 5% increase in public insurance coverage for people aged 14-18, while states that implemented SCHIP in 1999 showed an ATT of nearly half that. States that implemented SCHIP in 1997 found an a 4 percent, while states that implemented SCHIP in 1998 or 1999 showed no significant ATT. Our results are robust to a variety of modelling decisions.

## 2 Original Analysis and Data

### 2.1 Levine et al.

Levine et al. (2011) used the following Two-Way Fixed Effects Model:

$$\text{Insurance}_{iast} = \alpha + \beta * \text{Below19}_a * \text{Post1998}_t + \delta_a + \tau_t + \pi_s + \rho * \text{Unemp}_{st} + X_{iast} * \theta + \varepsilon_{iast}$$

The left hand side is an indicator for the (public or private)<sup>1</sup> insurance status of individual  $i$  at age  $a$  in state  $s$  at time  $t$ . This is regressed on an indicator variable which is equal to 1 if an individual was below 19 during 1998, and 0 otherwise. They also accounted for state-level unemployment rates  $\text{Unemp}_{st}$ , along with a variety of individual ( $X_{iast}$ ), age ( $\delta_a$ ), year ( $\tau_t$ ), and state-level ( $\pi_s$ ) fixed effects. Levine et al. (2011) found that public insurance coverage eventually increased to 6% in the 14-18 year old population, and private insurance coverage decreased by 1%.

<sup>1</sup>While Levine et al. (2011) also estimate this regression using an indicator for the individual having any insurance, we only estimate public and private insurance rates. This is because the third option is ultimately just a weighted average of the first two findings, and is therefore uninformative.

Their analysis has two potential flaws. First, by treating all states which passed an SCHIP-related law as if they implemented it in 1998, they have eliminated the possibility for heterogenous treatment effects across the three groups. Because the passage of these laws was voluntary, states which did so earlier may have done so because they expected it to be more useful than states which implemented it later. Second, TWFE has been found to be biased when estimating long run effects. Per Borusyak et al. (2021), this is caused by negative weights inflicted by the “forbidden comparisons” which occur when TWFE compares states treated at time  $t$  to states which were treated at time  $t' < t$ . To address both issues, we estimate the causal effect semi-parametrically, following the methods proposed by Callaway and Sant’anna (2021).

## 2.2 Data

We use the March CPS data provided by Levine et al (2011). States passed laws in 1997 (10 states), 1998 (25 states), 1999 (9 states), or 2001 (1 state). One state, West Virginia, never implemented SCHIP. Finally, 5 states had already extended Medicaid eligibility to all people under 19 before SCHIP was passed. We remove those 5 states because they are uninformative in our analysis of the causal effects of SCHIP. (Callaway and Sant’anna, 2011). We also exclude Arizona from our analysis. Arizona, the only state which passed a law in 2001, passed two SCHIP related laws (one in 1998, one in 2001), each of which affected Medicaid eligibility. Following Levine et al. (2011), we also exclude Massachusetts from our analysis because it passed multiple pieces of unrelated pieces of healthcare legislation during this time period.

One limitation of the CPS data is that it is unpaired. CPS data is anonymized, and participants in the survey are drawn randomly from the population. That means that different individuals participate in the survey every year, and we cannot track the same individual across multiple years. We explain how we deal with this issue in Section 3.2.4.

## 3 Multi-Period Differences in Differences

### 3.1 Causal Estimand

We examine states which implemented SCHIP in 1997, 1998, and 1999. As such, we divide each person under 19 into three groups:  $g = 1997, 1998, 1999$ , based on which state they resided in when they were polled for the CPS. Because we have 18 years (1991-2008) of cross-sectional data, we can estimate the average treatment on each group over a long horizon, which we summarize using an event study graph in the results section.

Following Callaway and Sant’anna (2021), we use the following notation. Our data covers  $t = 1991, \dots, 2008$ . We let  $T_{i,t} = 1$  if we observe individual  $i$  at time  $T$ . By design,  $T_{i,t} = 1$  for each individual exactly once. If an individual is in group  $g$  then we let  $G_{i,g} = 1$ . Otherwise, we let  $C_i = 1$  if individual  $i$  is never treated, either because they are at least 19 years of age, or they live in West Virginia. We use potential outcome notation to define our causal estimand. We let  $Y_i(t, g)$  equal 1 if individual  $i$  in group  $g$  has insurance at time  $t$ , and 0 else. We let  $Y_i(t, 0)$  equal

1 if individual  $i$  is never treated and has insurance in time  $t$  (and 0 else).

Following Callaway and Sant'anna, our causal estimand of interest is:

$$ATT(g, t) = E(Y_i(t, g) - Y_i(t, 0)|G_g = 1)$$

Which, for  $t \geq g$ , equals the average treatment on the treated (ATT) for individuals in group  $g$  at time  $t$ .

### 3.2 Identification

Identifying  $ATT(g, t)$  requires 5 assumptions. We empirically test each assumption when appropriate, and provide justification for the assumption when not.

#### 3.2.1 Irreversible Treatment

We assume that states which have implemented an SCHIP-related law by time  $t$  have not repealed that law by time  $t + 1$ . This assumption holds for all states over the years sampled. However, Arizona became the first state to freeze its SCHIP program in 2010, which forbade any new enrollments in SCHIP. (Burak and Fu, 2016) As discussed earlier, we drop Arizona for other issues.

#### 3.2.2 Anticipation of Treatment and Parallel Trends

We also assume that individuals do not anticipate treatment. More formally, we require:

$$E(Y_i(t, g)|X, G_g = 1) \stackrel{\text{a.s.}}{=} E(Y_i(t, 0)|X, G_g = 1) \quad \forall g, t \text{ such that } t < g \quad (1)$$

We also require conditional parallel trends with respect to the never treated group. This requires that:

$$E(Y_i(t, 0) - Y_i(t - 1, 0)|X, G_{i,g} = 1) \stackrel{\text{a.s.}}{=} E(Y_i(t, 0) - Y_i(t - 1, 0)|X, C_i = 1) \quad \forall g, t \text{ such that } g \leq t \quad (2)$$

This is the standard conditional parallel trends assumption, extended to the case in which there are multiple treated groups. These two assumptions can be jointly tested, which we defer to Section 5, as that test requires first defining our estimator and fitting our models.

#### 3.2.3 Overlap

We modify the overlap condition by requiring that for each  $g$  there exists  $\varepsilon > 0$  such that:

$$p_g(X) = P(G_g|X, G_g + C = 1) < 1 - \varepsilon \quad (3)$$

We can test this by estimating  $p_g(X)$  for each  $g$  with confounders for the number of never-married children under 18, marital status, being female, having children, ratio of family income to low-income level, disability, having a high-school degree, and having a college degree. This was the largest set of confounders that we were confident would not violate the standard DiD DAG. We report the density plots for  $p_{1998}(X)$  for three different models in Figure 1

and their Accuracy and Cross-Entropy in [Figure 2](#). We find that Random Forests (maximum depth=10, number of estimators=100) has the highest accuracy, and the lowest cross-entropy. Furthermore, while there is some evidence of overlap violations for Logistic Regression, there are no obvious issues with using either Random Forests or XGBoost for estimation. We use Random Forests to estimate  $p_g(X)$  for the rest of this analysis because it performs best, and has no issues surrounding overlap. Finally, as discussed in the following section, our IPW/AIPTW estimates only contains terms proportional to  $\frac{p_g(X)}{1-p_g(X)}$  which means that propensity scores near 0 are unproblematic.

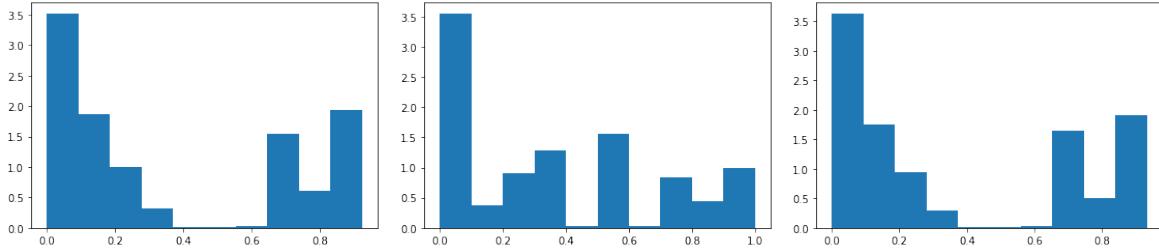


Figure 1: Density Plots of (In Order): Random Forest, Logistic Regression, and XGBoost

	model	g_ce	baseline_ce	accuracy_score
0	Rf_depth2	0.499453	0.657917	0.858877
1	RF_depth10	0.354816	0.657917	0.861143
2	KNN	1.699834	0.657917	0.826930
3	LogReg	0.383680	0.657917	0.856638
4	XGBoost	0.354318	0.657917	0.860973

Figure 2: Accuracy and Cross-Entropy for Each G-Model

### 3.2.4 Dealing with Cross-Sectional Data

The standard semi-parametric DiD setup requires observing the same individuals before and after treatment. However, as the CPS dataset is a cross-section of the population, we never observe the same individual more than once. In order to identify  $ATT(g, t)$  under these conditions, we follow Appendix B in Callaway and Sant'anna (2021), which requires an additional assumption. We assume that for each individual:

$$(Y_{i,t}, X_i, G_{i,g}) \stackrel{\text{iid}}{\sim} F_{Y,X,G}$$

Here, while we assume that the marginal distribution  $F_Y$  does depend on time, we also assume that  $F_{X,G_g}$  is time invariant. This assumption is necessary, as otherwise we cannot accurately estimate  $p_g(X)$  because it is a time-invariant feature of  $F_{X,G_g}$ . Unfortunately, this assumption does not hold in this case. In [Figure 3](#), we plot how each confounder used in our final analysis has changed over time. While many of the confounders are somewhat stable over time, there are two clear outliers: “hsgrad” (an indicator for having graduated high school) and “anykids” (an indicator for having at least one child). As a robustness check, we redo our analysis without hsgrad or anykids in

Appendix B. We find that it does not affect our analysis or conclusions.

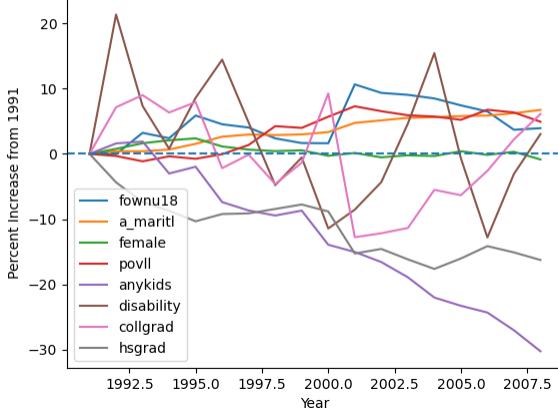


Figure 3: Change in the composition of confounders over time

### 3.2.5 Estimation

We follow Appendix B in Callaway and Sant’anna in estimating an AIPTW estimator suited for cross-sectional data. This estimator requires a little more notation to formalize. Let  $m_{g,t}^{\text{treat}} = E(Y|X, T = t, G_g = 1)$  and  $m_{g,t}^{\text{nev}} = E(Y|X, T = t, C = 1)$ . These are the year by group probabilities of being insured, conditional on confounders and either being in group  $g$  or never receiving treatment. We let our weights be equal to:

$$w^{\text{treat}}(g, t) = \frac{G_g * T_t}{E(G_g * T_t)} \quad w^{\text{nev}}(g, t) = \frac{T_t * p_g(X) * C}{1 - p_g(X)} / E\left(\frac{T_t * p_g(X) * C}{1 - p_g(X)}\right)$$

Each of those weights are natural extensions of the propensity-weights found in the standard AIPTW estimator. We combine these into a final AIPTW estimator:

$$\begin{aligned} \widetilde{ATT}(g, t) &= E\left(\frac{G_g}{E(G_g)} \left( m_{g,t}^{\text{treat}}(X) - m_{g,g-1}^{\text{treat}}(X) - \left( m_{g,t}^{\text{nev}}(X) - m_{g,g-1}^{\text{nev}}(X) \right) \right)\right) + \\ &E\left(w_{g,t}^{\text{treat}} \left( Y - m_{g,t}^{\text{treat}}(X) \right) - w_{g,g-1}^{\text{treat}} \left( Y - m_{g,g-1}^{\text{treat}}(X) \right)\right) - E\left(w_{g,t}^{\text{nev}} \left( Y - m_{g,t}^{\text{nev}}(X) \right) - w_{g,g-1}^{\text{nev}} \left( Y - m_{g,g-1}^{\text{nev}}(X) \right)\right) \end{aligned}$$

Under our assumptions,  $\widetilde{ATT}(g, t) = ATT(g, t)$ .<sup>2</sup> This model uses  $g - 1$  as the base period against which we compare outcomes for all  $t \geq g$ . This gives the  $ATT(g, t)$  the interpretation of the average treatment on the treated at time  $t$  since time  $g - 1$ . This is often referred to as a long-difference. This estimator can be viewed as an analogue of the standard ATT-AIPTW estimator (Murphy, 2022). The first term is analogous to  $Q(1, X) - Q(0, X)$  where we estimate each term separately to get around the short-comings of cross-sectional data. The remaining terms are analogous to  $A \frac{Y - Q(1, X)}{g(x)} - (1 - A) \frac{Y - Q(0, X)}{1 - g(x)}$ . We also estimate G-only (IPW) and Q-only versions of this estimator.

Those models appear in Appendix A. We find that our point estimates replicate reasonably well across choice of

<sup>2</sup>This estimator’s influence curve is defined analogously to class. See Callaway and Sant’anna for a more thorough treatment of the influence curve in the panel-data case.

model. As discussed in the following section, we obtain very poor MSE when estimating our  $Q$  model, which results in implausibly small standard errors. As such, we rely on the AIPTW estimator due to its attractive properties when we may have misspecified  $Q$ .

## 4 Fitting Our $Q$ Estimator

We fit  $Q$  with several different models including Random Forests with a depth of 10, K-Nearest-Neighbors, Extreme Gradient Boosting, and Logistic Regression, which are shown in Figure 5. While our  $Q$  models have good accuracy scores that fall within a range of 0.9 to 0.91, the MSE of our models are consistently larger than the baseline model, suggesting that our confounders are not informative when predicting insurance coverage. These issues persist no matter which model or confounders we use. Given the apparent lack of information in our confounders, it is likely that the  $Q$ -only model will not perform well, which is confirmed in practice.<sup>3</sup> Because the AIPTW estimator is non-parametrically efficient and doubly robust, we still proceed with our main analysis. As XGBoost performed the best with an accuracy of 0.913 and a MSE of 0.087, we use that model for the remainder of the paper.

	model	test_mse	baseline_mse	accuracy_score
0	RandomForest	0.087820	0.085292	0.912180
1	KNN	0.095923	0.085292	0.904077
2	LogReg	0.094093	0.085292	0.905907
3	XGBoost	0.087036	0.085292	0.912964

Figure 4: Accuracy and Mean-Square Error for Each  $Q$ -Model

## 5 Testing No-Anticipation and Conditional Parallel Trends

Using our estimate of  $ATT(g, t)$  We can simultaneously test both no-anticipation and conditional pre-trends by estimating  $ATT(g, t)$  for  $t < g$ . Both assumptions imply  $ATT(g, t) = 0$  for all  $t < g$ , which provides a natural test of both assumptions. In this case, when we estimate  $\widetilde{ATT}(g, t)$ , we let our baseline period be  $t - 1$ , instead of  $g - 1$ . This is equivalent to us pretending that treatment occurred between  $t - 1$  and  $t$ , and estimating the ATT due to that “fake-treatment.” We plot  $ATT(g, t)$  in the pre-treatment period in Figure 5 It is important to note that this is only test of pre-trends, rather than parallel-trends. We observe pre-trends violations for both public and private insurance. For private insurance, the pre-trends violations are large and frequent. Fortunately, for public insurance, the pre-trends violations tend to be smaller and fairly rare. Both violations are cause for concern, as this provides evidence that there is either anticipation or violations of parallel-trends, both of which could bias our analysis.

---

<sup>3</sup>See Appendix A

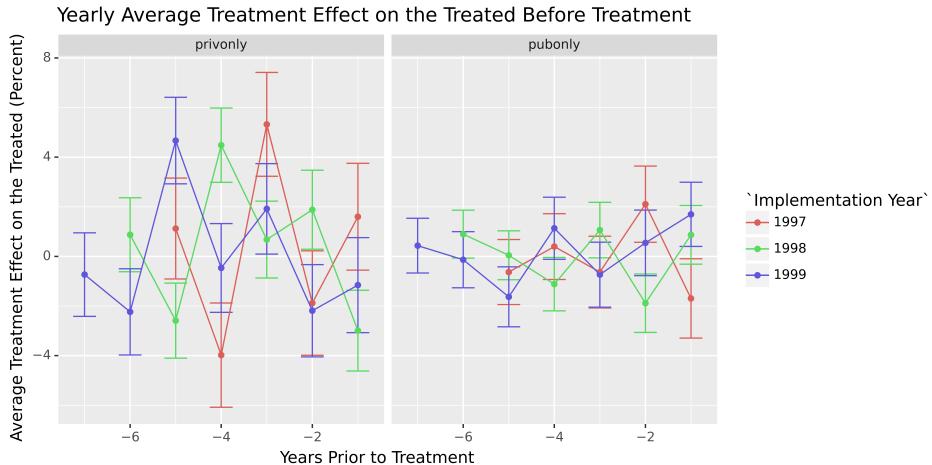


Figure 5: Pre-Treatment  $\text{ATT}(g,t)$

## 6 Results

We estimate the  $\text{ATT}(g,t)$  using a AIPTW model fitted with XGBoost for the Q-model and Random Forests for the g-model. We graph the resulting  $\text{ATT}(g,t)$  in the post-treatment period in [Figure 6](#) for both public and private insurance. Insurance rates for all groups generally increase over the 10 years for public insurance. States which implemented SCHIP in 1997 saw increases up to nearly 6%, 1998 states saw increases up to 5% , and 1999 states saw increases of roughly 2% by 2008. These effects are plausible (but smaller than the effects observed by Levine et al), as policies which make it easier for individuals to stay on health insurance would result in higher health insurance rates among that group. These effects are largest for  $g = 1997$ , smaller for  $g = 1998$ , and quite small for  $g = 1999$ . This may be due to selection effects. States which implemented SCHIP later may have put off the implementation because they expected small effects. SCHIP implementation had a negative effect on private insurance coverage for  $g = 1997$  but no significant decrease for  $g = 1998$  or  $g = 1999$ . This result implies that some of the large public insurance rate gains in  $g = 1997$  may be due to individuals switching from private to public insurance (which wasn't seen for  $g = 1998, 1999$ ).

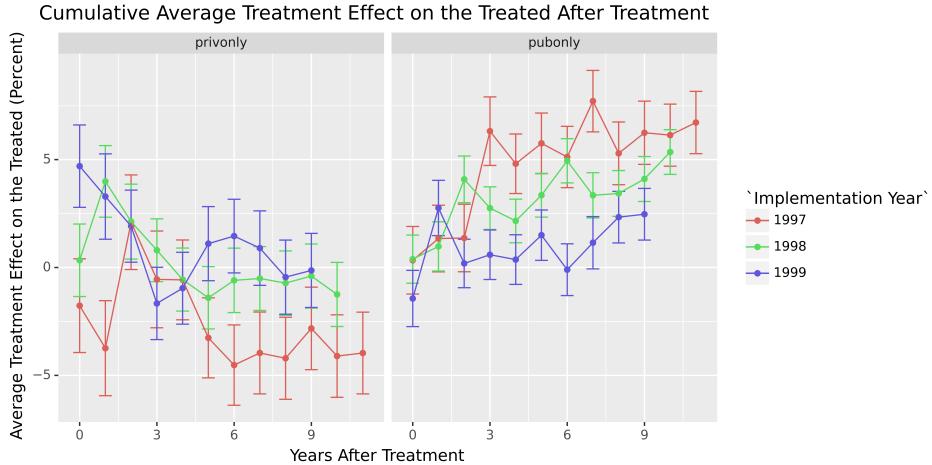


Figure 6: Post-Treatment  $\text{ATT}(g,t)$

## 7 Conclusion

We use semi-parametric differences-in-differences to estimate the causal effects of the implementation on SCHIP on youth insurance rates, and find large, positive effects on public insurance rates for individuals under 19, and smaller, negative effects on private insurance rates. This suggests that SCHIP's impressive gains in public insurance rates are partially driven by re-sorting of youth insurance coverage from private to public insurance rates. We find that early treated ( $g = 1997$ ) states see the greatest effects, with large positive effects on public insurance and large negative effects for private insurance. We find that states which adopted SCHIP later ( $g = 1998, 1999$ ) had smaller effects on both private and public insurance rates. We re-estimate our model with a subset of our confounders, and different choices of  $\text{ATT}(g,t)$  model.

While Levine et al. use a simple TWFE model with a single treated group, our results are consistent with their findings. While this is encouraging, our analysis is incomplete in multiple ways. We observe pre-trends violations, and are unable to fit a Q-estimator which does any better than the baseline. Furthermore, the assumption that confounders are stable in proportion over time does not hold, which may lead to biased and inconsistent estimates of our propensity scores. Finding new datasets which can better account for those issues may be a fruitful avenue for future work.

## 8 Replication

All files required for replication can be found [here](#).

## References

- B. Callaway and P.H.C., Sant'Anna. Differences-in-Differences with Multiple Time Periods. *SSRN*, Dec 2020.
- K. Murphy. Probabilistic Machine Learning Advanced Topics. MIT Press, 2022.
- K. Borusyak, X. Jaravel, and J. Spiess. Revisiting Event Study Designs: Robust and Efficient Estimation. *arXiv*, Apr 2022.
- E. Burak and J. Fu. Children's Health Coverage in Arizona: How Are Children Doing Without KidsCare. *Georgetown University*, Jan 2016.
- P. Levine, R. McKnight, S. Herp. How Effective are Public Policies to Increase Health Insurance Coverage Among Young Adults? *American Economic Journal: Economic Policy*, Feb 2011.

## 9 Appendix A

Recall that  $m_{g,t}^{\text{treat}} = E(Y|X, T = t, G_g = 1)$  and  $m_{g,t}^{\text{nev}} = E(Y|X, T = t, C = 1)$ . Let:

$$w^{\text{treat}}(g, t) = \frac{G_g * T_t}{E(G_g * T_t)} \quad w^{\text{nev}}(g, t) = \frac{T_t * p_g(X) * C}{1 - p_g(X)} / E\left(\frac{T_t * p_g(X) * C}{1 - p_g(X)}\right)$$

We estimate the Q-only estimator as:

$$Q - ATT(g, t) = E\left(\frac{G_g}{E(G_g)} \left( m_{g,t}^{\text{treat}}(X) - m_{g,g-1}^{\text{treat}}(X) - (m_{g,t}^{\text{nev}}(X) - m_{g,g-1}^{\text{nev}}(X)) \right)\right)$$

And the IPW estimator as:

$$G - ATT(g, t) = E\left(\left(w^{\text{treat}}(g, t) - w^{\text{treat}}(g, g-1)\right)Y\right) - E\left(\left(w^{\text{nev}}(g, t) - w^{\text{nev}}(g, g-1)\right)Y\right)$$

[Figure 7](#) and [Figure 8](#) below show the pre- and post-treatment  $ATT(g,t)$  plots the Q-only and IPW models. We observe that the IPW model is very similar to the AIPTW model. However, we see that the Q-only model has standard errors very close to 0. Based on the MSE which we observed when fitting Q, this is likely because XGBoost is just recovering the average for each group/year. That means that it  $ATT(g,t)$  (for each individual) does not vary much with the confounders, which explains the small standard errors.

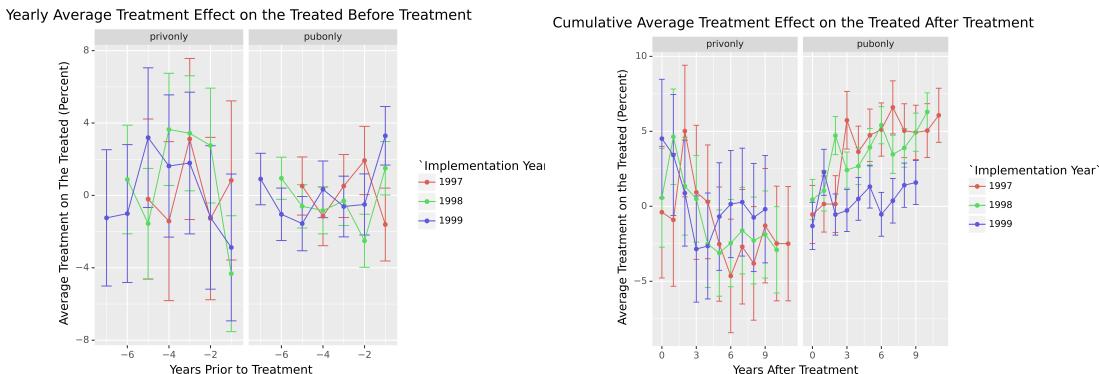


Figure 7: Pre- and Post- Treatment Plots of IPW model

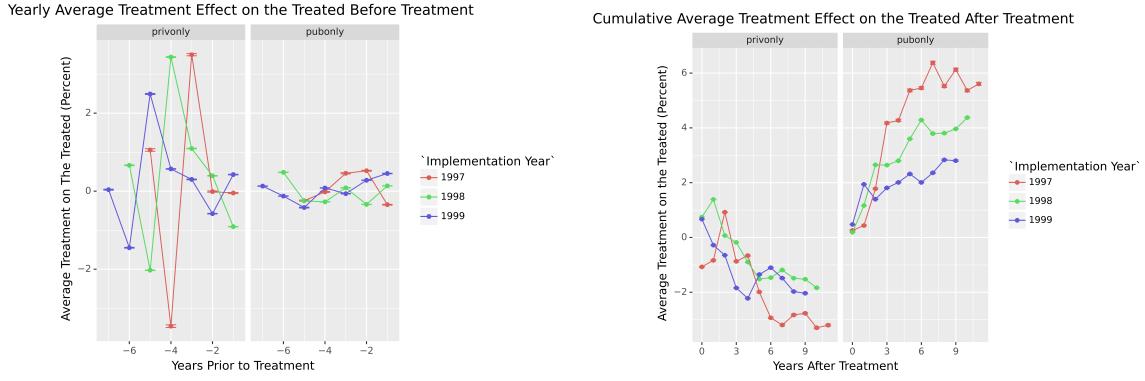


Figure 8: Pre- and Post- Treatment Plots of Q model

## 10 Appendix B

Dropping “anykids” and “hsgrad” does not significantly alter our results, or their interpretation. In Figure 9, we observe that dropping those confounders has almost no effect on pre-trends. In Figure 10, we observe that dropping those confounders has a small positive effect on the public insurance coverage for  $g = 1997$ , and no significant effects for any other group or outcome. Overall these plots demonstrate that our results are quite robust to dropping the confounders which most violated our assumption that  $F_{X,G_g}$  is time invariant.

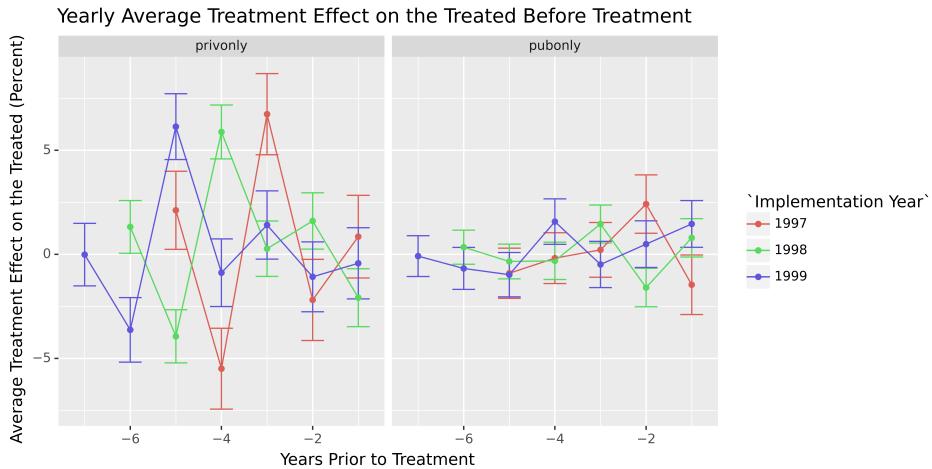


Figure 9: Pre-Treatment  $\text{ATT}(g,t)$  without the confounders: anykids or hsgrad

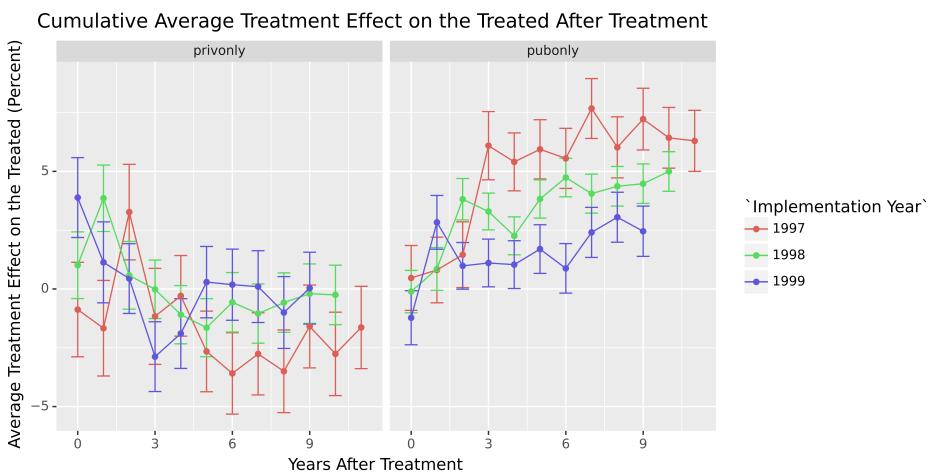


Figure 10: Post-Treatment  $\text{ATT}(g,t)$  without the confounders: anykids or hsgrad