The (In)Effectiveness of State R&D Grants

Evidence from the Kentucky SBIR State Match Program

Caleb Maresca

August 31, 2021

Abstract

A number of states have enacted R&D subsidy programs that match funds from the federal Small Business Innovation Research program. State politicians claim that these programs catalyze local R&D and entrepreneurship. I test these claims using a synthetic control for Kentucky's State Match Program. Results find a large, yet statistically insignificant, positive effect on private R&D five years after the program's inauguration and a surprising short term negative effect on the quantity of new business registrations. Due to the nature of the model, the results are interpreted as upper bounds on the effect of State Match Programs.

1 Introduction

Governments often use subsidies to incentivize private research and development (R&D). The common rationale behind this action is that the private rate of return for R&D is significantly lower than the social rate of return (Jones and Williams, 1998). Various papers have studied the effect of R&D subsidies (Lach, 2002; Marino et al., 2016) and have generally found them to be effective, that is, that they complement private R&D instead of crowding it out. These studies, however, have generally focused on federal R&D subsidies; the literature on state level programs is much smaller. This study seeks to evaluate the effectiveness of state R&D subsidies on macroeconomic outcomes, specifically private R&D and entrepreneurship, by using Kentucky's Small Business Innovation Research (SBIR) State Match program as a case study.

SBIR is a federal program that disburses billions of dollars per year to small businesses of up to 500 employees in order to promote innovation by helping small businesses perform R&D and commercialize their products. Established in 1982, the program has given over 170,000 awards summing almost $50 billion as of the beginning of 2021. Multiple academic studies have praised the federal program (Keller and Block, 2013; Audretsch et al., 2019; Link and Scott, 2012). In response to the successful federal program, many states have instituted State Match programs that match federal SBIR awards given to companies located in their state with the goal of invigorating local innovation and entrepreneurship.

Investigating state programs gives the advantage of heterogeneity between states with and without the program as well as heterogeneity between programs in different states.

Even given these advantages, little attention has been given to the subject. Lanahan (2016) empirically evaluates State Match programs and finds them effective in helping recipient firms reach the next phase of the national program but stops short of assigning causality. Lanahan and Feldman (2018) compare project level data between Kentucky and North Carolina which have State Match, and the surrounding states that do not and again find evidence that the State Match helps projects reach the next phase of the federal program.

This paper seeks to evaluate the macroeconomic effects of State Match Programs to see if state governments are receiving the results that they expect, namely more private R&D and higher levels of entrepreneurship. Specifically, I analyze the effect of SBIR State Match programs on three outcome variables of interest: private R&D, quantity of new businesses and quality adjusted quantity of new businesses.

This analysis is difficult for two main reasons. First, the decision to establish a State Match program is likely endogenous, Lanahan and Feldman (2015) find that the decision to implement a State Match program is correlated with the programs of neighboring states, having a Democratic governor, higher revenue, lagging high tech employment and other important factors that could confound the effects of the program on private R&D or entrepreneurship. Second, detailed information on State programs is hard to find. Lanahan (2016) in her nationwide comparison resorted to using a dummy variable for states with the program due to lack of detailed information on each states program. This is problematic because though most State Match programs are similar in structure, they can vary greatly in their level of support and funding.

To combat these issues, this study compares Kentucky's State Match using a synthetic control as pioneered by Abadie et al. (2010). I also use recent improvements to the method as developed in Klößner and Pfeifer (2015) and Kaul et al. (2021). Kentucky had one of the most aggressive State Match programs in the time-frame studied (our data is until 2013) making it a good candidate for a case study of the program's effectiveness. The synthetic control is a convex combination of states that did not have a State Match program before 2013, balanced to match Kentucky on a number of socio-political variables. The control provides a convincing "alternative Kentucky" that didn't establish a State Match Program. Using Kentucky as a case study limits the external validity of our results but given the heterogeneity between different state programs and the inadequate information available on them, limiting the analysis to Kentucky allows for more precise identification of the treatment policy and greater internal validity.

The results of the synthetic control estimations are rather surprising. I find a large point estimate on private R&D, especially in the final years measured, with private R&D being 40% higher in real Kentucky in comparison to its synthetic counterpart in 2013. However, I do not find a significant positive effect on any of the outcomes examined. In fact, introduction of the State Match program may have *reduced* the number of new business registrations in Kentucky. Due to Kentucky's relatively strong State Match program, I interpret these estimates as upper bounds on the program's effects in other states.

The article continues as follows. Section 2 gives the history of the national SBIR program as well as State Matching Programs. Section 3 details the research design, introducing the synthetic control model and its variants. Section 4 introduces the data used for the empirical analysis. Section 5 reviews the results of the econometric analysis. Section 6 concludes and Appendix A details alternative models used at robustness tests.

2 SBIR and State Match Programs

The SBIR program was first established in 1982 with the passage of the Small Business Innovation Development Act. Its goal was fourfold: ”(1) to stimulate technological innovation; (2) to use small business to meet Federal research and development needs; (3) to foster and encourage participation by minority and disadvantaged persons in technological innovation; and (4) to increase private sector commercialization innovations derived from Federal research and development.” (Public Law 97-219, 1982). In particular, SBIR helps small firms overcome the difficult transition period between investment and sales. As of the beginning of 2021, the program has disbursed almost $50 billion.

The program is coordinated through the Small Business Administration (SBA), though the program is mostly run by the participating agencies, which include all federal agencies with an R&D budget of excess of $100 million. Currently, there are 13 participating agencies, but the large majority of SBIR grants are awarded by the Department of Defense, the Department of Health and Human Services, the Department of Energy, the National Aeronautics and Space Agency and the National Science Foundation.

The SBIR program is separated into three phases, Phase I, Phase II and Phase III. SBIR offers grants for Phase I and Phase II, but not for Phase III. Firms that reach Phase III may receive grants or contracts from other federal programs. In Phase I, firms receive a grant to establish a proof of concept and commercialization potential of their idea. Firms that reach Phase II receive much higher levels of funding in order to make their idea into a reality. Finally, in Phase III, firms transition from R&D into commercialization of their product.

Many scholars have praised the program. Keller and Block (2013) applaud the national SBIR program for catalyzing the growth of small business R&D. Audretsch et al. (2019) evaluate Department of Defense SBIR awards and conclude that the program is stimulating R&D, successfully leads to commercialization that otherwise would not have occurred, and that the social benefits of the program are large. Link and Scott (2012) find large but not statistically significant effects of SBIR funding on employment in award winning firms. Wessner (2008) finds that "the SBIR program is sound in concept and effective in practice. It can also be improved. Currently, the program is delivering results that meet most of the congressional objectives." On the other hand, Wallsten (2000) finds that SBIR awards do not lead to increased employment in the receiving firm and that they crowd out private R&D dollar for dollar.

Numerous states, seeing the success of the federal program, instituted what are now known as State Match Programs (SMPs). These programs are numerous and diverse, both in their level of financial funding and in their structure. The common theme between them, however, is that SMPs provide additional funds to firms that win grants from the national SBIR program. The primary objectives of SMPs is to direct national funds into the state, encourage new businesses to form within the state, or to move to the state from elsewhere, and to catalyze local R&D.

Fewer studies have been conducted on SMPs. Lanahan and Feldman (2015) examine factors that predict the establishment of State Matching Programs, as a response to federal policy, other states' policies or due to internal state politics. Lanahan (2016) evaluates State Matching Programs in their effectiveness at helping projects progress from Phase I of SBIR to Phase II in a cross-sectional state-level analysis. Lanahan and Feldman (2018) use project level data to create semi-exogenous variation between projects from firms in Kentucky or North Carolina which have State Match programs to projects from firms in the surrounding states that did not have the program, finding that SMP funds are most effective when given to firms with little previous experience with SBIR grants.

3 Empirical Strategy

Whereas previous studies have examined the microeconomic effects of SMPs, such as their ability to help firms bridge the gap between Phase I and Phase II of the national SBIR program (Lanahan, 2016), this study seeks to examine the macroeconomic effects of SMPs on state economies. In particular, I examine whether SMPs have fulfilled their goals of catalyzing private R&D and entrepreneurship. The large heterogeneity of different SMPs and the limited information available on them makes cross-sectional analysis difficult. In addition, the endogeneity of the policy decision to enact a SMP presents a challenge to causal inference from regressions (even fixed effects models will not solve this issue, as there are likely to be omitted variables that are time-variant).

In order to overcome these limitations I decided to focus on Kentucky's SMP, which is known to be one of the most aggressive and well funded, and to use a synthetic control method (SCM) to compare the state to a "counterfactual" Kentucky that did not enact a SMP. Due to the relative strength of Kentucky's SMP, I interpret my results as an upper bound on the effect of SMPs in other states. The required assumption for causal inference is that the synthetic control provides an accurate estimate of the counterfactual outcome in the treated unit had the treatment not occurred. A good fit between the treated unit and the synthetic control is a good indication that this assumption holds.

The SCM is a generalization of the classic difference-in-difference model in which a data-driven methods are used to construct a convex combination of the available control units. The SCM was first used by Abadie and Gardeazabal (2003) and was later theoretically developed and refined in Abadie et al. (2010). Klößner and Pfeifer (2015) and Kaul et al. (2021) introduced a number of improvements into this estimation technique. Klößner and Pfeifer (2015) show that the basic SCM can augmented to include time series data, tightening the fit of the resulting synthetic control. Kaul et al. (2021) show that using the full roster of pre-intervention outcomes as predictors renders the rest of the predictors meaningless, as all of the weight will be placed on the lagged outcomes, and can bias the model.

I integrate these insights into my analysis by running a time series synthetic control without using the pre-intervention outcomes as predictors. In the appendix I run four additional models for robustness and find similar results. I run the model separately for each of the three dependent variables, leading to a total of three estimations. I then run a placebo test on all of the control units for which the mean squared prediction error (MSPE) is at most five times as large as the MSPE of synthetic Kentucky for quantity and quality adjusted quantity of new business registrations and ten times as large for private R&D (due to the small number of states that fit the previous criteria under this specification). For estimation, I used algorithms described in Becker and Klößner (2018).

For the potential control units, I use states that did not have a SMP prior to 2014 (my data on SMPs is updated as of 2013, see subsection 4.3). In addition, Missouri, New Hampshire and West Virginia were excluded due to lack of information on the dependent variables, and Delaware was excluded as an outlier because a disproportionally high number of startups are registered there. Figure 1 displays which states did and did not have a SMP in this time frame. I use the same predictors for all three specifications and let the data decide which predictors are more or less important for each dependent variable. Specifically, I use federal R&D obligations, venture capital investment, state GDP, employment in science engineering and technology (SET) industries, federal SBIR funding, private sector output, population and population holding at least a bachelor's degree as predictors. See Table 1 for a list of predictors.

Though all of the predictors are added to all of the specifications and I let the data decide which are more important for each specification, the motivation between the predictors vary. Employment in SET industries is the primary input in R&D and private sector output is the primary funder of private R&D. These variables are therefore expected to be good predictors of private R&D. Federal R&D obligations and federal SBIR funding could have spillover effects on private R&D. The purpose of venture capital is to fund new startups, therefore I expect it to be a good predictor of my entrepreneurship variables. Chatterji et al. (2014) find that population and the general education of the population are good predictors of entrepreneurship. Lastly, State GDP is added as a general indicator of the economy.

4 Data

This study relies on data from two main sources: the National Science Foundation (NSF) State Indicators and the Startup Cartography Project (SCP). Specifically, the data on private

R&D and on most of the control variables comes from the NSF State Indicators, whereas information on the other two outcome variables comes from the SCP. In Subsection 4.1, I review the data gathered from the NSF State Indicators. Next, in Subsection 4.2, I review the SCP and its associated data. Lastly, in Subsection 4.3, I review two other sources of data that were used.

4.1 NSF State Indicators

The National Science Foundation compiles administrative data and government surveys into sixty different state-level indicators. The indicators are divided into six different categories: ”Elementary and Secondary Education”, ”Higher Education, Workforce”, ”Financial Research and Development Inputs”, ”Research and Development Outputs” and ”Science and Technology in the Economy”. Many of the indicators are created by combining multiple different variables. For example, the indicator ”Business-Performed R&D as a Percentage of Private-Industry Output” is created by dividing a measure of private R&D by a measure of private sector output. The raw data available on their website contains data on each of the underlying variables as well as the final indicator.

From the indicator in the example above, I derive data on both private R&D and on private sector output. Data on federal R&D spending, state GDP, venture capital investments, (national) SBIR funding, employment in SET industries, and bachelor's degree holders are likewise derived from the ”Federal R&D Obligations per Employed Worker”, ”Venture Capital Disbursed per $1 Million of Gross Domestic Product”, ”Average Annual Federal Small Business Innovation Research and Small Business Technology

Transfer Funding per $1 Million of Gross Domestic Product", "Employment in High Science, Engineering, and Technology Employment Establishments as a Percentage of Total Employment" and "Bachelor's Degree Holders in the Labor Force" indicators.

Most of the indicators have data going back to the 1990s, however, some important variables have data only starting in the 2000s. In particular, SET industry employment data starts in 2003, and information on degree holders starts in 2005. This means that I only have three and one year(s) of data for each of these variables, respectively, before the inauguration of Kentucky's State Match program in 2006. A list of the variables generated from the NSF State Indicators, start date for each variable and their minimum and maximum values can be found in the middle section of Table 1.

4.2 Startup Cartography Project

The Startup Cartography Project collects business registration records and builds geographic datasets tracking information on startup formation, quality, and performance. Data is available at many different levels of aggregation, of which I use their state level data. The SCP data is used for our two measures of entrepreneurship, namely, new business registrations and quality adjusted new business registrations.

The SCP variable Startup Formation Rate (SFR) measures the number of new business registrations in each state-year. The variable Regional Entrepreneurship Cohort Potential Index (RECPI) measures quality adjusted business registrations. This second variable is constructed by multiplying the SFR by the Entrepreneurial Quality Index (EQI), SCP's indicator of entrepreneurial quality in a local area. Both variables are available starting in

1988. Their summary statistics can be found in the top section of Table 1.

4.3 Other Data

Two other sources were used for my analysis. The final control variable, population, was gathered from Census data and can be found at the bottom of Table 1. Lastly, Prof. Lauren Lanahan graciously shared unpublished data on State Match programs. Specifically, which states have them, when they began and ended (if they ended) and details on each one. This data was used to exclude states that had a State Match program in the time-frame studied. The data on SMPs is updated as of 2013, and thus this is the last year to which I run my synthetic control models.

5 Results

Table 2 displays the weights given to each predictor for each of the three synthetic controls. When private R&D is the dependent variable, the model gives a 89.8% weight on private sector output and a 10.1% weight on national SBIR funding in the state while giving marginal weights to the other predictors. This seems reasonable, as private sector output is the main source of income firms use to fund their R&D and the purpose of the national SBIR program is to encourage R&D, thus the fact that these two variables are optimal for predicting private R&D is not surprising.

Using SFR as the dependent variable, the model gives a 73.2% weight to venture capital funding, a 17.5% weight on state GDP, a 4.9% weight on population, a 3.8% weight on SBIR funding, with marginal weights given to the other predictors. Again, these results are reasonable, as venture capital is expressly given to startup firms, nor is it surprising that state GDP is correlated with new business registrations.

Lastly, when RECPI is the dependent variable, the model gives a 99.97% weight to population. This is concerning as I would have expected to see some of the other variables come into play. This is the first sign of a problem with the fit for RECPI (more on this below).

Moving on to weight placed on each state, Table 3 displays the results for each of the three synthetic controls. When private R&D is the dependent variable, a 67.9% weight is given to Mississippi, a 14.2% weight to Georgia, a 10% weight to Tennessee, a 4.3% weight to South Carolina and a 3.5% weight to Louisiana. With SFR as the dependent variable, a 34.8% weight is given to Louisiana, 33.1% to Arkansas, 14.7% to Mississippi, 12.1% to Tennessee, 4.9% to Wisconsin and 0.3% to California. Finally, with RECPI as the dependent variable, the model gives a 55.5% weight to Iowa, a 24.3% to Louisiana, 11.6% to Tennessee, 3.4% to Idaho, 2.7% to Texas, 1.6% to Mississippi and 0.9% to Arkansas. Though the weights shift between dependent variables, there is a common theme. They all give a positive percentage to Louisiana, Mississippi and Tennessee.

An argument can be made that it is problematic to give different weights to different predictors and states depending on the dependent variable. After all, all three of the dependent variables are economic in nature, therefore preferably there would be one counterfactual Kentucky that mimics Kentucky's economy. To alleviate this concern, I run a model in the appendix that predicts private R&D and SFR concurrently. RECPI was removed from this model due to the poor ability of the SCM to predict this variable (as shown below).

Next, Table 4 examines the ability of the synthetic control to match Kentucky in the values of the predictor variables. Due to the time series nature of the predictors there are a very large number of predictors. To conserve space, only the last five entries (where available) are displayed for each predictor. Across the different specifications, we generally find a good fit for private sector output, state GDP, employment in SET industries, population and population with at least a bachelor's degree. Federal R&D commitments, venture capital funds, and national SBIR funding tend to be further away from their true values (though depending on the specification-year SBIR funding is sometimes very close to its true value). In all of these variables, Kentucky appears to be an outlier, and is much lower than the sample mean. Though the synthetic control tends to be higher than Kentucky in these variables, it is much lower than the sample mean. Lastly, we find that between the specifications, RECPI tends to be furthest away from the true values, this is not surprising since it placed a near 100% weight on population.

Now, for the main results, Figure 2 displays a line graph with two lines. The black line gives the yearly values of private R&D in real Kentucky, and the red line in synthetic Kentucky. In the pre-intervention period, synthetic Kentucky mostly matches real Kentucky, though it is less volatile, underestimating the hump between the years 1999 and 2000, and missing the dip in the years 2003 and 2004. In the post-intervention period, synthetic Kentucky remains below, but relatively close to real Kentucky between the years 2006 and 2009. Starting in 2010, the lines begin to diverge more, with real Kentucky taking a large lead over synthetic Kentucky. In the final year private R&D in real Kentucky is 40% above its counterpart in synthetic Kentucky, a real accomplishment. Figure 3 displays the results of the placebo test. Though the point estimate was large, the result is not significant, with three to four states (out of 17 including Kentucky) above Kentucky for the duration of the post-intervention time period.

Next, Figure 4 follows the same format as Figure 2 for the SFR specification. In the pre-intervention period, the fit is very tight, although the real and synthetic units begin to diverge in 2005, a year before the implementation of the new program. This could potentially be due to an anticipation effect by local (would be) business owners. In the post-intervention time period, we find a negative effect on SFR until the last two years studied, at which time this negative effect disappears. The largest gap appears immediately in 2006, at which time SFR in real Kentucky is 21% lower than SFR in synthetic Kentucky. The placebo test in Figure 5 shows that this negative effect is significant in 2006, but ceases to remain significant in the duration of the time period studied.

Finally, Figure 6 displays the results for the RECPI specification. As hinted to earlier, the fit for RECPI is rather poor. In the post-intervention time period, real Kentucky is constantly below synthetic Kentucky. Figure 7 shows that this outcome is significant during the first two post-intervention time periods (2006 and 2007) although the poor fit in the pre-intervention time period throws doubt on this result.

6 Conclusion

This paper has examined the impact of Kentucky's State Match Program on private R&D and entrepreneurship in the state. This was done by building a synthetic control comprising of states that did not install a State Match Program in the time frame studied. Results show a large yet statistically insignificant effect on private R&D, especially five to seven years after the establishment of the program. Counter-intuitively, the program depressed new business registrations in the short term, to a significant degree in the first year. Results also show a significant negative effect on quality adjusted new business registrations, yet this finding is not reliable due to the poor fit of the synthetic control for this variable in the pre-intervention period.

One of the assumptions of the synthetic control model is non-interference between units, meaning that the treatment does not affect outcomes in the control units. This assumption certainly does not hold for this study. New businesses may decide to open in Kentucky instead of other states due to Kentucky's State Match Program, thus depressing entrepreneurship in the control states. State politicians bragged about how the State Match program convinced companies to move R&D activities to Kentucky, thus depressing R&D in control states. These factors would depress the dependent variables in the synthetic control, thus leading to an upward bias in the estimated treatment effect. In addition, Kentucky's State Match Program is better funded than similar programs in other states. For these two reasons, I interpret the estimated treatment effects as upper bounds on the effectiveness of SMPs.

The main limitation of this study is a lack of reliable data on SMPs. In particular, the lack of information on which states do and do not have SMPs after 2013 impedes any analysis of the long-term effects of the program. This is a significant limitation, as such programs can take a number of years to have a large and significant effect. For example, Fazio et al. (2020) look at state-level R&D tax credits and find significant effects on entrepreneurship (using the same variables from the Startup Cartography Project) starting only eight years after the inception of the program. States have much to gain from making this information publicly available.

References

Abadie, A., A. Diamond, and J. Hainmueller (2010). Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association 105*(490), 493–505.

Abadie, A. and J. Gardeazabal (2003). The economic costs of conflict: A case study of the basque country. *American economic review 93*(1), 113–132.

Audretsch, D. B., A. N. Link, and J. T. Scott (2019). Public/private technology partnerships: evaluating sbir-supported research. In *The Social Value of New Technology*. Edward Elgar Publishing.

Becker, M. and S. Klößner (2018). Fast and reliable computation of generalized synthetic controls. *Econometrics and Statistics 5*, 1–19.

Chatterji, A., E. Glaeser, and W. Kerr (2014). Clusters of entrepreneurship and innovation. *Innovation policy and the economy 14*(1), 129–166.

Fazio, C., J. Guzman, and S. Stern (2020). The impact of state-level research and development tax credits on the quantity and quality of entrepreneurship. *Economic Development Quarterly 34*(2), 188–208.

Jones, C. I. and J. C. Williams (1998). Measuring the social return to r&d. *The Quarterly Journal of Economics 113*(4), 1119–1135.

Kaul, A., S. Klößner, G. Pfeifer, and M. Schieler (2021). Standard synthetic control methods: The case of using all pre-intervention outcomes together with covariates. *Journal of*

Business & Economic Statistics (just-accepted), 1–34.

Keller, M. R. and F. Block (2013). Explaining the transformation in the us innovation system: the impact of a small government program. *Socio-Economic Review 11*(4), 629–656.

Klößner, S. and G. Pfeifer (2015). Synthesizing cash for clunkers: Stabilizing the car market, hurting the environment. Number F13-V1 in Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Automobiles and the Environment. ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft.

Lach, S. (2002). Do r&d subsidies stimulate or displace private r&d? evidence from israel. *The journal of industrial economics 50*(4), 369–390.

Lanahan, L. (2016). Multilevel public funding for small business innovation: A review of us state sbir match programs. *The Journal of Technology Transfer 41*(2), 220–249.

Lanahan, L. and M. P. Feldman (2015). Multilevel innovation policy mix: A closer look at state policies that augment the federal sbir program. *Research Policy 44*(7), 1387–1402.

Lanahan, L. and M. P. Feldman (2018). Approximating exogenous variation in r&d: Evidence from the kentucky and north carolina sbir state match programs. *Review of Economics and Statistics 100*(4), 740–752.

Link, A. N. and J. T. Scott (2012). Employment growth from public support of innovation in small firms. *Economics of Innovation and New Technology 21*(7), 655–678.

Marino, M., S. Lhuillery, P. Parrotta, and D. Sala (2016). Additionality or crowding-out? an overall evaluation of public r&d subsidy on private r&d expenditure. *Research Policy 45*(9), 1715–1730.

Public Law 97-219 (1982). Small business innovation development act of 1982.

Wallsten, S. J. (2000). The effects of government-industry r&d programs on private r&d: the case of the small business innovation research program. *The RAND Journal of Economics 31*(1), 82–100.

Wessner, C. W. (2008). An assessment of the sbir program at the national science foundation.

Table 1: Summary Statistics

	Start Year	Min	Max
From SCP:			
SFR	1988	40	287591
RECPI	1988	0.018	345.10
From NSF State Indicators:			
Private R&D	1997	20	89373
Fed R&D	1991	15.73	27226.90
Venture cap	1995	0	21843.15
State GDP	1995	13891	2262771
SET ind. emp.	2003	14820	1858727
SBIR	1991	0.011	477.21
Private sector output	1997	12366	1988551
Degree holders	2005	29405	3481290
From Census:			
Population	1988	453690	36961229

Notes: This table gives summary statistics on all of the dependent and predictor variables. The first two are from the Startup Cartography Project, the third through second to last are from the National Science Foundation's State Indicators, and lastly, population statistics are from the Census. The "Start Year" column gives the first year for which we have information for that variable available (other than the Census, for which 1988 was chosen as the first year to match the other variables). The "Min" and "Max" columns give the minimum and maximum values for each variable (meaning across all state-years).

Table 2: Predictor Weights

Predictor	R&D SCM	SFR SCM	RECPI SCM
Fed R&D	1.25310E-05	6.35165E-03	1.99463E-06
Private sector output	8.98466E-01	2.92963E-06	1.72730E-08
State GDP	8.55106E-06	1.75293E-01	1.72730E-08
Venture cap.	8.55106E-06	7.32206E-01	2.14868E-04
SBIR funding	1.01104E-01	3.75438E-02	4.43246E-05
SET ind. emp.	8.55106E-06	2.92963E-06	3.19559E-06
Population	3.83697E-04	4.85969E-02	9.99725E-01
Degree Holders	8.55106E-06	2.92963E-06	1.08729E-05

Notes: Each column gives the weights given to each predictor in a particular specification. In the second column from the left the dependent variable is private R&D, in the next column it is new business registrations, and finally in the rightmost column, the dependent variable is quality adjusted new business registrations. SCM is short for synthetic control model.

Table 3: State Weights

	R&D SCM	SFR SCM	RECPI SCM
Alabama	0.000	0.000	0.000
Alaska	0.000	0.000	0.000
Arizona	0.000	0.000	0.000
Arkansas	0.000	0.331	0.009
California	0.000	0.003	0.000
Colorado	0.000	0.000	0.000
Georgia	0.142	0.000	0.000
Idaho	0.000	0.000	0.034
Iowa	0.000	0.000	0.555
Louisiana	0.035	0.348	0.243
Maine	0.000	0.000	0.000
Minnesota	0.000	0.000	0.000
Mississippi	0.679	0.147	0.016
Nevada	0.000	0.000	0.000
New Mexico	0.000	0.000	0.000
North Dakota	0.000	0.000	0.000
Ohio	0.000	0.000	0.000
Oregon	0.000	0.000	0.000
Pennsylvania	0.000	0.000	0.000
Rhode Island	0.000	0.000	0.000
South Carolina	0.043	0.000	0.000
South Dakota	0.000	0.000	0.000
Tennessee	0.100	0.121	0.116
Texas	0.000	0.000	0.027
Utah	0.000	0.000	0.000
Vermont	0.000	0.000	0.000
Washington	0.000	0.000	0.000
Wisconsin	0.000	0.049	0.000
Wyoming	0.000	0.000	0.000

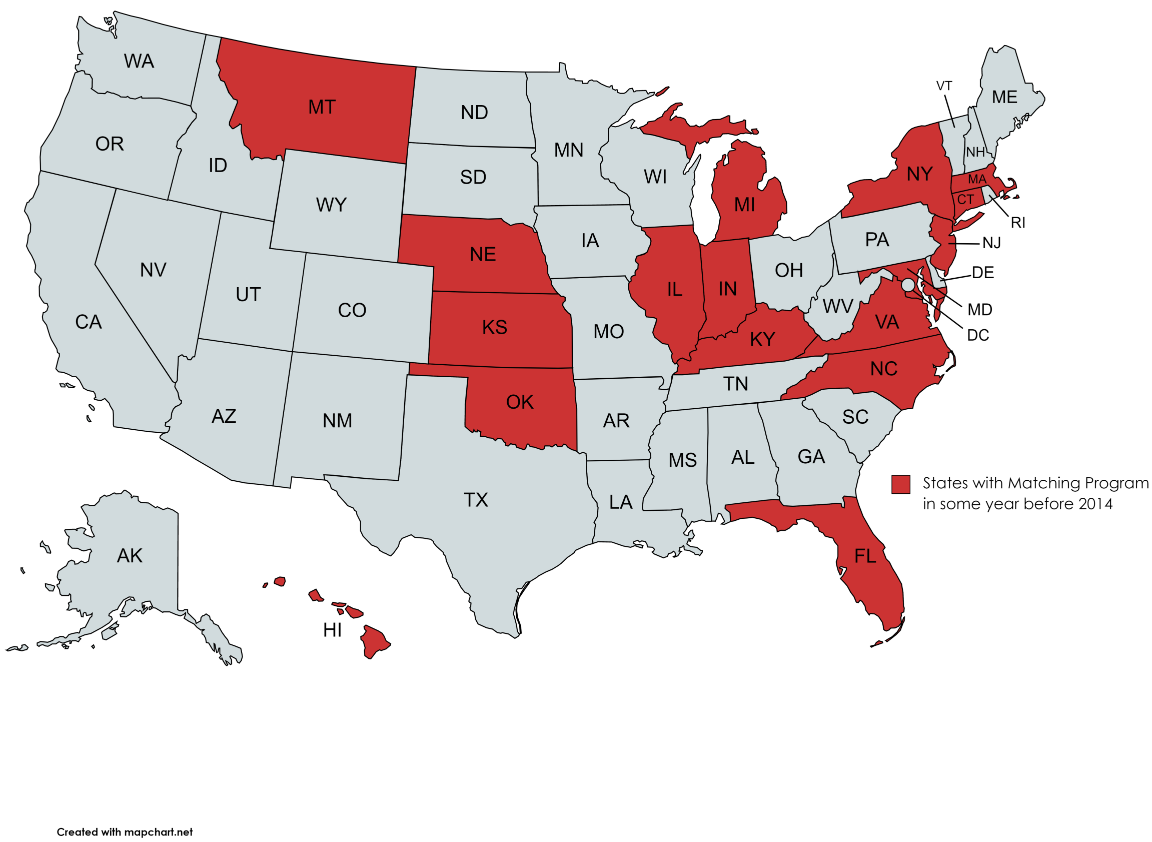
Notes: Each column gives the weights given to each state in a particular specification. In the second column from the left the dependent variable is private R&D, in the next column it is new business registrations, and finally in the rightmost column, the dependent variable is quality adjusted new business registrations. SCM is short for synthetic control model.

Table 4: Predictor Values

Predictor	Treated	R&D SCM	SFR SCM	RECPI SCM	Sample Mean
Fed RD 2001	370.38	898.72	411.72	488.91	1475.01
Fed RD 2002	321.54	890.47	523.83	610.93	1728.53
Fed RD 2003	246.66	1254.32	616.95	690.91	1945.47
Fed RD 2004	251.76	1527.74	703.34	751.44	1968.36
Fed RD 2005	296.31	787.56	545.54	701.02	2005.78
Private sector output 2001	101368.00	102771.70	102513.87	115272.33	169362.48
Private sector output 2002	105037.00	105182.26	105059.32	118565.92	174507.14
Private sector output 2003	109008.00	110702.57	113116.53	126989.73	183858.28
Private sector output 2004	115443.00	117988.77	122862.72	140226.83	197604.24
Private sector output 2005	122267.00	124892.35	135836.52	152573.53	212632.38
State GDP 2001	117431.00	119469.25	117592.98	131306.81	193891.97
State GDP 2002	122193.00	123100.47	121227.00	135641.11	200578.21
State GDP 2003	126971.00	129697.44	130207.54	144921.41	211315.45
State GDP 2004	134194.00	137901.95	140960.12	159048.96	226355.72
State GDP 2005	142649.00	145986.11	154956.75	172431.88	242836.76
Venture cap. 2001	8.68	92.54	55.88	60.59	601.99
Venture cap. 2002	4.63	126.98	57.12	46.31	470.15
Venture cap. 2003	6.25	51.42	43.30	44.22	424.64
Venture cap. 2004	27.38	88.34	54.03	49.01	513.78
Venture cap. 2005	213.46	82.88	62.26	49.37	561.05
SBIR funding 1997	2.57	2.78	3.19	2.98	18.97
SBIR funding 1999	2.48	4.14	3.75	3.61	20.51
SBIR funding 2001	3.78	5.11	4.83	5.08	24.34
SBIR funding 2003	5.05	6.80	6.74	7.49	36.14
SBIR funding 2005	5.80	8.24	10.41	9.54	44.69
empl set ind 2003	121838.00	137896.60	133103.56	150456.96	257121.83
empl set ind 2004	119167.00	133997.29	132008.52	143251.95	250619.59
empl set ind 2005	118969.00	137845.50	134397.84	147260.13	253449.03
Population 2001	4068132.00	4038266.90	3947539.98	4067430.80	5436449.28
Population 2002	4089875.00	4067838.50	3968545.65	4089457.79	5493724.55
Population 2003	4117170.00	4098534.63	3993733.72	4115788.58	5548190.59
Population 2004	4146101.00	4143442.93	4026298.42	4148373.79	5607248.55
Population 2005	4182742.00	4188669.15	4059837.87	4181685.56	5668022.34
Degree holders 2005	256210.00	282176.22	246086.27	285971.93	442870.52

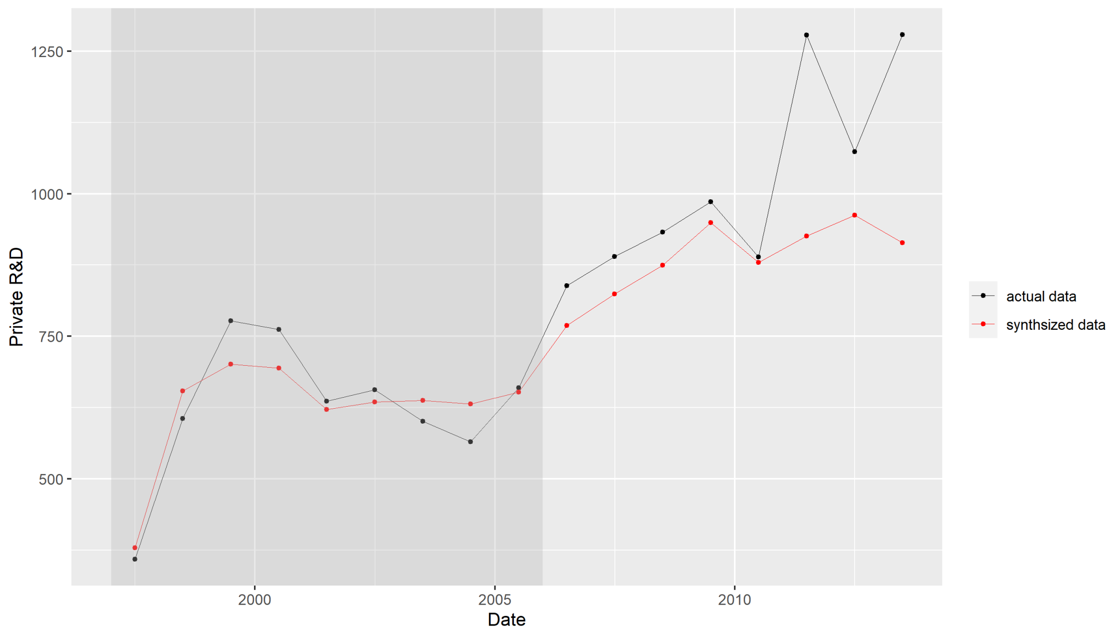
Notes: Each column gives the value of a predictor in a particular specification or unit. The second column from the left gives the predictor value in the treated unit (in other word, real Kentucky), the next column gives the predictor values for the synthetic control estimated for private R&D, followed by new business registrations in the next column, quality adjusted new business registrations in the column after that, and finally in the rightmost column, the sample mean of that predictor (between all potential control states) is given. Each predictor is shown for five period back starting from 2005, where available. SCM is short for synthetic control model.

Figure 1: Map of State Match Programs



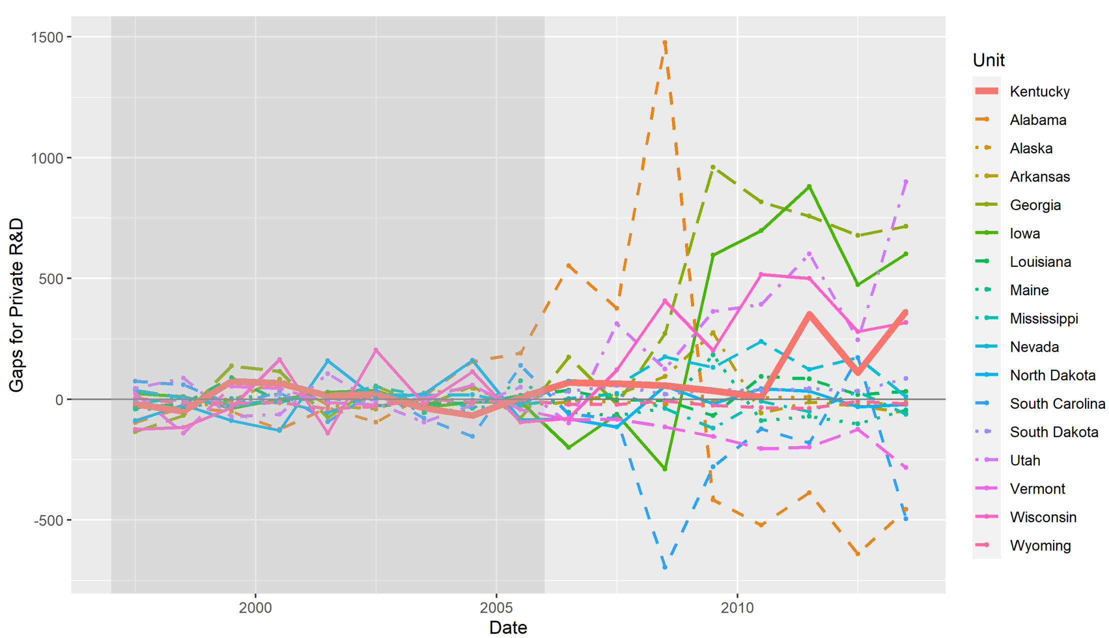
Note: Map of states with a State Match Program in some year prior to 2014. The grey states are the potential control units (with the exception of Delaware, Missouri, New Hampshire and West Virginia, see subsection 4.3) used for the synthetic controls.

Figure 2: Private R&D Synthetic Control



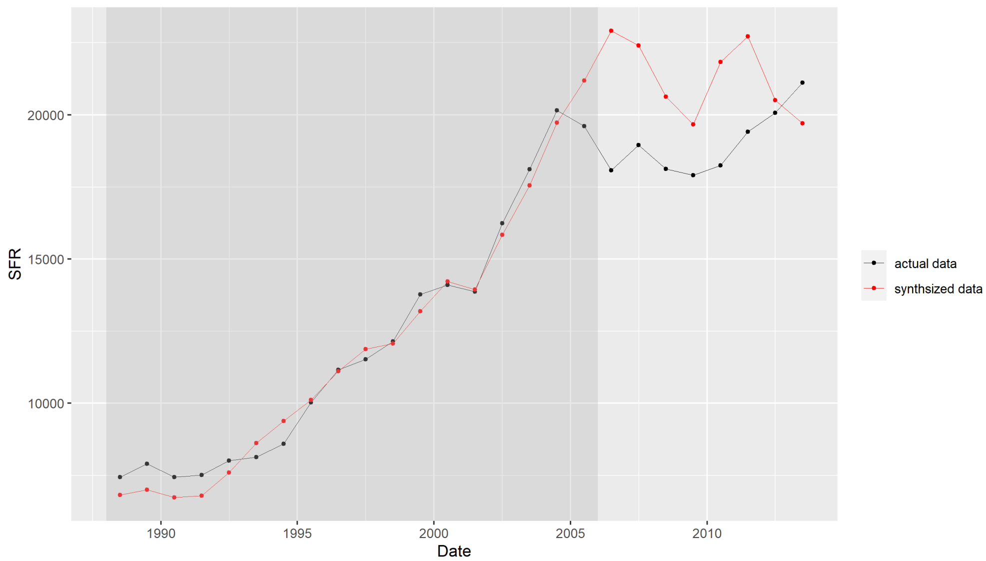
Note: The shaded area indicates the pre-intervention period and the unshaded area the post intervention period. The treatment effect is equal to the vertical distance between the black line and the red line (negative if the red line is above the black). Private R&D is measured in millions of USD.

Figure 3: Private R&D Placebo Test



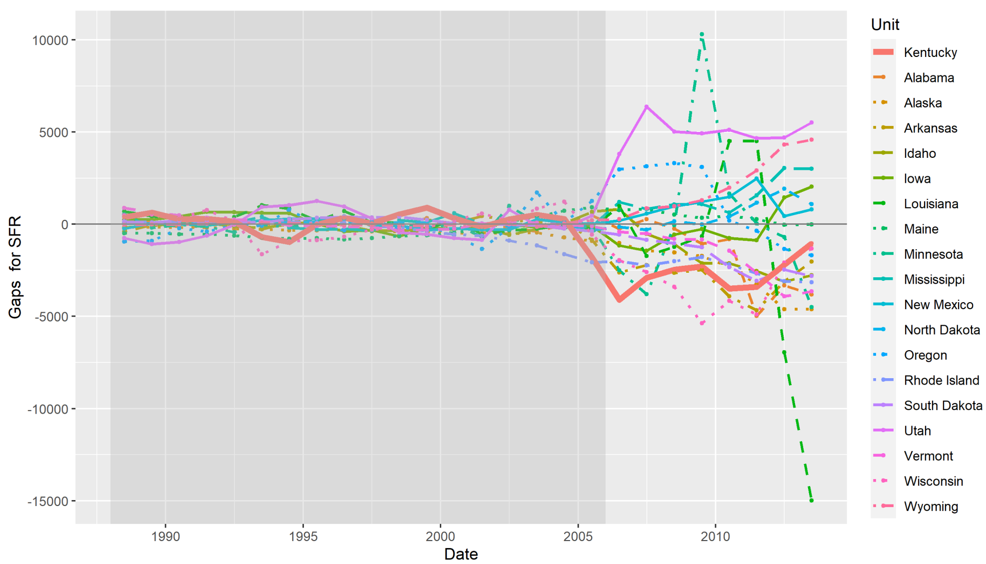
Note: The shaded area indicates the pre-intervention period and the unshaded area the post intervention period. Each line represents the gap between the actual data for that state and the synthetic control built for that state. To be significant, the treatment effect in Kentucky needs to be greater in magnitude than the placebo effects. Private R&D is measured in millions of USD.

Figure 4: SFR Synthetic Control



Note: The shaded area indicates the pre-intervention period and the unshaded area the post intervention period. The treatment effect is equal to the vertical distance between the black line and the red line (negative if the red line is above the black). SFR is short for Startup Formation Rate, and is equal to the number of new businesses opened in the state that year.

Figure 5: SFR Placebo Test

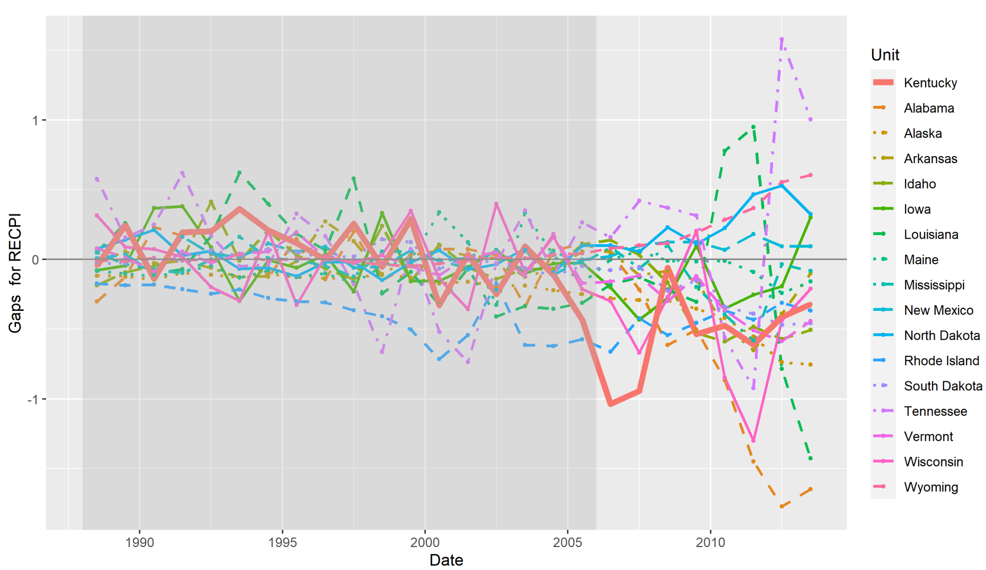


Note: The shaded area indicates the pre-intervention period and the unshaded area the post intervention period. Each line represents the gap between the actual data for that state and the synthetic control built for that state. To be significant, the treatment effect in Kentucky needs to be greater in magnitude than the placebo effects. SFR is short for Startup Formation Rate, and is equal to the number of new businesses opened in the state that year.

Figure 6: RECPI Synthetic Control

Note: The shaded area indicates the pre-intervention period and the unshaded area the post intervention period. The treatment effect is equal to the vertical distance between the black line and the red line (negative if the red line is above the black). RECPI is short for ”Regional Entrepreneurship Cohort Potential Index”, a measure of quality adjusted new business registrations.

Figure 7: RECPI Placebo Test



Note: The shaded area indicates the pre-intervention period and the unshaded area the post intervention period. Each line represents the gap between the actual data for that state and the synthetic control built for that state. To be significant, the treatment effect in Kentucky needs to be greater in magnitude than the placebo effects. RECPI is short for ”Regional Entrepreneurship Cohort Potential Index”, a measure of quality adjusted new business registrations.

A Robustness Tests

In order to check the robustness of my results, I run four additional models. The first model (hereinafter Model A1) is a "simple" SCM which uses the mean of all the prediction variables including the dependent variable as predictors. The second model (hereinafter Model A2) uses these same prediction variables as time series data instead of averaging them. In other words, the only difference between Model A2 and the main model is the inclusion of the dependent variable as a predictor. In order to ensure similarity between the treated unit and the control, the third model (hereinafter Model A3) is identical to the main model, except the predictors are matched only up to the year 2002, instead of 2005. A valid control would continue to match the treated unit for the remaining three years, and only begin to diverge afterwards. Data on the population of bachelor's degree holders and employment in SET industries is only available after 2002, and therefore these predictors were omitted.

Figure A.1, Figure A.2 and Figure A.3 compare models A1, A2 and A3 for R&D, SFR and RECPI respectively. The results of the alternative models are generally in line with the main model. For private R&D, model A1 gives almost identical results to the main model, while models A2 and A3 are higher than the main model, implying an smaller and sometimes negative treatment effect. Model A3 misses the dip in years 2002 and 2003, which is a cause for caution. However, the post-intervention outcomes are similar to the other models, in particular to Model A2.

The results for SFR are very robust. All models have a very tight fit in the pre-intervention period. In particular, though Model A3 diverges from the treated unit in

2005 like the other models (and to approximately the same magnitude), it maintains a good fit in years 2003 and 2004. All three alternative models show a negative effect on SFR, with Model A1 mimicking the main model almost perfectly, Model A2 slightly below it and Model A3 slightly above it. Trivially, all four models (including the main model) show a poor fit to RECPI.

Figure A.4, Figure A.5 and Figure A.6 display the results of placebo tests for Model A2 for each of the outcome variables. We again find a non-significant positive effect on private R&D a non-significant negative effect on SFR. The effect on RECPI is negative, and at some points significant, however, the poor fit makes this finding unreliable. Placebo tests for the other specifications find similar results and are available upon request.

Table A.2, Table A.4 and Table A.6 display the state weights when private R&D, SFR and RECPI are the dependent variables, respectively. Each column represents a different model. Though the weights vary, the same general group of states is used between models. Table A.1, Table A.3 and Table A.5 present the predictor weights for private R&D, SFR and RECPI, respectively, with each column representing a different model.

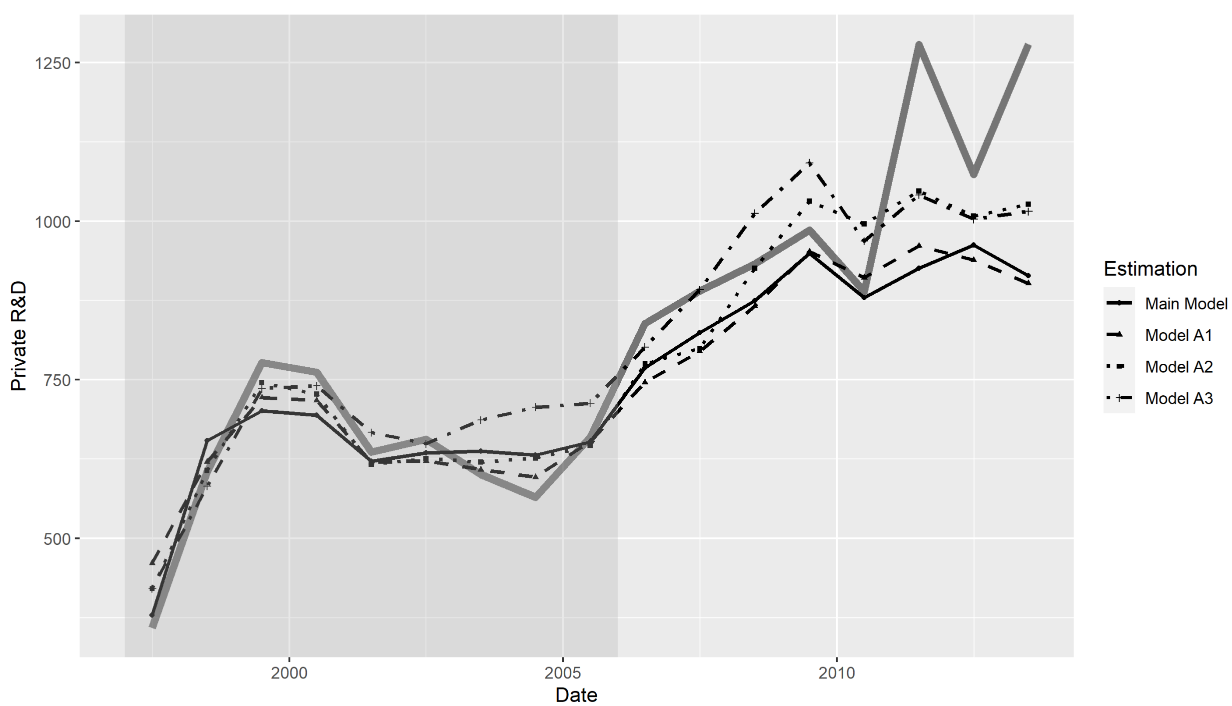
In all of the models above, separate synthetic controls are created for each dependent variable. As stated in Section 5, this could be criticized as problematic. Model A4 addresses this concern by building one synthetic control that attempts to predict both private R&D and SFR concurrently. RECPI was removed from this model due to the poor ability of the SCM to predict this variable. This additional restriction has an expected adverse effect on the fit of the model, however the general results in Section 5 continue to hold. Figure A.7 displays the result of Model A4 on both variables of interest, and Figure A.8 a placebo test. The point estimates on private R&D, especially in later years, grow to be quite large, however this effect is not significant. The negative effect on SFR is comparable to the other models, though this time sometimes bordering on significant.

Table A.1: R&D Predictors and Fit

Predictor	Model A1	Model A2	Model A3	Main Model
State R&D	1.36296E-01	1.00000E+00	–	–
Fed R&D	1.90414E-08	2.16793E-08	1.30279E-02	1.25310E-05
Private sector output	1.32904E-04	1.00000E-08	7.56595E-01	8.98466E-01
State GDP	7.34364E-01	1.00000E-08	6.87712E-08	8.55106E-06
Venture cap.	1.29049E-01	1.01520E-08	2.04045E-01	8.55106E-06
SBIR funding	1.64206E-05	1.02587E-08	2.63317E-02	1.01104E-01
SET ind. emp.	8.98652E-07	1.00000E-08	–	8.55106E-06
Population	1.78974E-07	1.00000E-08	6.87712E-08	3.83697E-04
Degree holders	1.40706E-04	1.00000E-08	–	8.55106E-06

Notes: Dependent variable: Private R&D. Each column gives the weights given to each predictor in a particular model. The "Main Model" is the model described in section 3, and for information on models A1, A2 and A3, see Appendix A. Not all models include all predictors, where not included, the weight is marked with "–".

Figure A.1: Private R&D Comparison of Estimations



Note: The shaded area indicates the pre-intervention period and the unshaded area the post intervention period. Each line type represents a different model. The "Main Model" is the model described in section 3, and for information on models A1, A2 and A3, see Appendix A. Private R&D is measured in millions of USD.

Table A.2: R&D State Weights

	Model A1	Model A2	Model A3	Main Model
Alabama	0.000	0.000	0.000	0.000
Alaska	0.000	0.000	0.000	0.000
Arizona	0.000	0.000	0.000	0.000
Arkansas	0.000	0.000	0.404	0.000
California	0.000	0.000	0.001	0.000
Colorado	0.000	0.000	0.000	0.000
Georgia	0.067	0.173	0.000	0.142
Idaho	0.109	0.080	0.000	0.000
Iowa	0.000	0.000	0.000	0.000
Louisiana	0.437	0.493	0.445	0.035
Maine	0.000	0.000	0.000	0.000
Minnesota	0.000	0.000	0.000	0.000
Mississippi	0.296	0.251	0.000	0.679
Nevada	0.000	0.000	0.000	0.000
New Mexico	0.000	0.000	0.000	0.000
North Dakota	0.000	0.000	0.000	0.000
Ohio	0.000	0.000	0.000	0.000
Oregon	0.000	0.000	0.000	0.000
Pennsylvania	0.000	0.000	0.000	0.000
Rhode Island	0.000	0.000	0.000	0.000
South Carolina	0.078	0.000	0.000	0.043
South Dakota	0.000	0.000	0.000	0.000
Tennessee	0.000	0.003	0.000	0.100
Texas	0.000	0.000	0.000	0.000
Utah	0.000	0.000	0.000	0.000
Vermont	0.000	0.000	0.000	0.000
Washington	0.014	0.000	0.000	0.000
Wisconsin	0.000	0.000	0.150	0.000
Wyoming	0.000	0.000	0.000	0.000

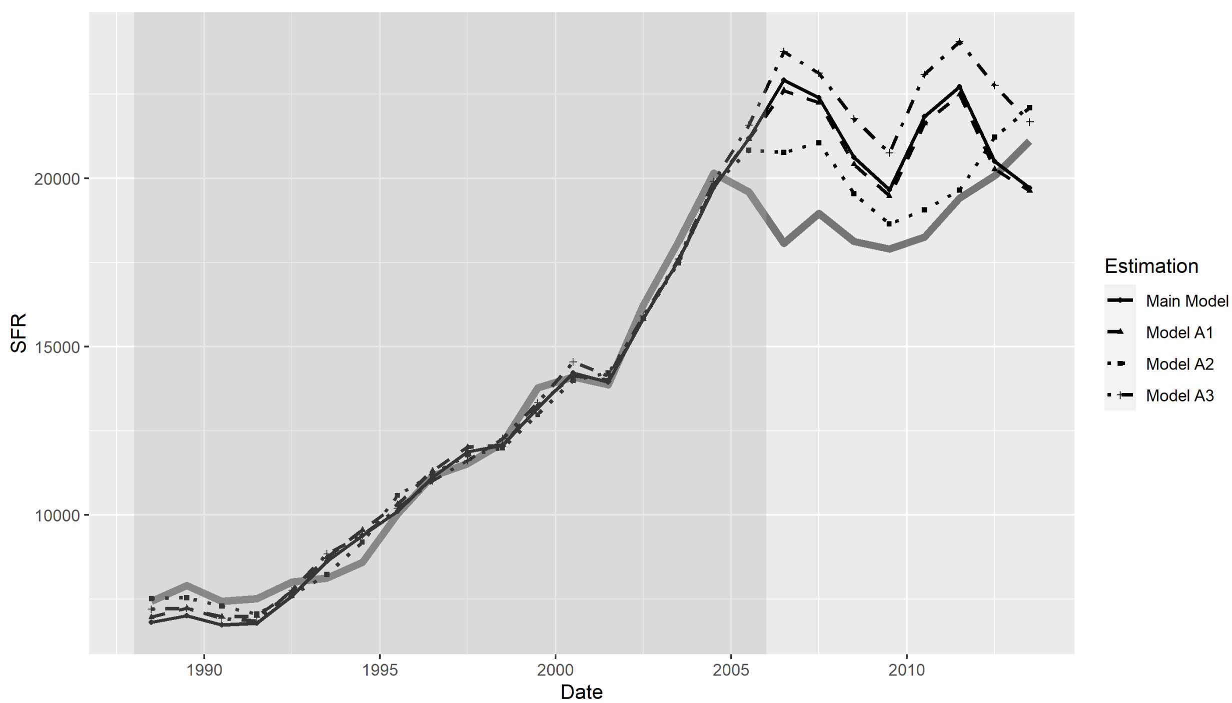
Notes: Dependent variable: Private R&D. Each column gives the weights given to each state in a particular model. The ”Main Model” is the model described in section 3, and for information on models A1, A2 and A3, see Appendix A.

Table A.3: SFR Predictors and Fit

Predictor	Model A1	Model A2	Model A3	Main Model
State R&D	6.09412E-01	1.00000E+00	–	–
Fed R&D	7.44882E-05	1.00000E-08	3.38159E-06	6.35165E-03
Private sector output	1.16091E-06	1.00000E-08	9.29314E-01	2.92963E-06
State GDP	1.16091E-06	1.00000E-08	4.51632E-04	1.75293E-01
Venture cap.	8.60179E-04	1.00228E-08	6.48327E-02	7.32206E-01
SBIR funding	2.83282E-05	1.00000E-08	5.39485E-03	3.75438E-02
SET ind. emp.	1.16091E-06	1.00000E-08	–	2.92963E-06
Population	3.15183E-01	1.00122E-08	3.38159E-06	4.85969E-02
Degree holders	7.44384E-02	1.00150E-08	–	2.92963E-06

Notes: Dependent variable: quantity of new business registrations. Each column gives the weights given to each predictor in a particular model. The "Main Model" is the model described in section 3, and for information on models A1, A2 and A3, see Appendix A. Not all models include all predictors, where not included, the weight is marked with "–".

Figure A.2: SFR Comparison of Estimations



Note: The shaded area indicates the pre-intervention period and the unshaded area the post intervention period. Each line type represents a different model. The "Main Model" is the model described in section 3, and for information on models A1, A2 and A3, see Appendix A. SFR is short for Startup Formation Rate, and is equal to the number of new businesses opened in the state that year.

Table A.4: SFR State Weights

	Model A1	Model A2	Model A3	Main Model
Alabama	0.000	0.000	0.000	0.000
Alaska	0.000	0.000	0.000	0.000
Arizona	0.000	0.000	0.000	0.000
Arkansas	0.513	0.400	0.000	0.331
California	0.000	0.011	0.000	0.003
Colorado	0.000	0.000	0.000	0.000
Georgia	0.000	0.000	0.000	0.000
Idaho	0.000	0.000	0.000	0.000
Iowa	0.000	0.000	0.000	0.000
Louisiana	0.326	0.000	0.347	0.348
Maine	0.000	0.000	0.000	0.000
Minnesota	0.000	0.000	0.000	0.000
Mississippi	0.000	0.000	0.568	0.147
Nevada	0.000	0.000	0.000	0.000
New Mexico	0.000	0.000	0.000	0.000
North Dakota	0.000	0.000	0.000	0.000
Ohio	0.000	0.145	0.065	0.000
Oregon	0.000	0.000	0.000	0.000
Pennsylvania	0.033	0.027	0.020	0.000
Rhode Island	0.000	0.000	0.000	0.000
South Carolina	0.000	0.000	0.000	0.000
South Dakota	0.000	0.000	0.000	0.000
Tennessee	0.122	0.106	0.000	0.121
Texas	0.005	0.012	0.000	0.000
Utah	0.000	0.000	0.000	0.000
Vermont	0.000	0.298	0.000	0.000
Washington	0.000	0.000	0.000	0.000
Wisconsin	0.000	0.002	0.000	0.049
Wyoming	0.000	0.000	0.000	0.000

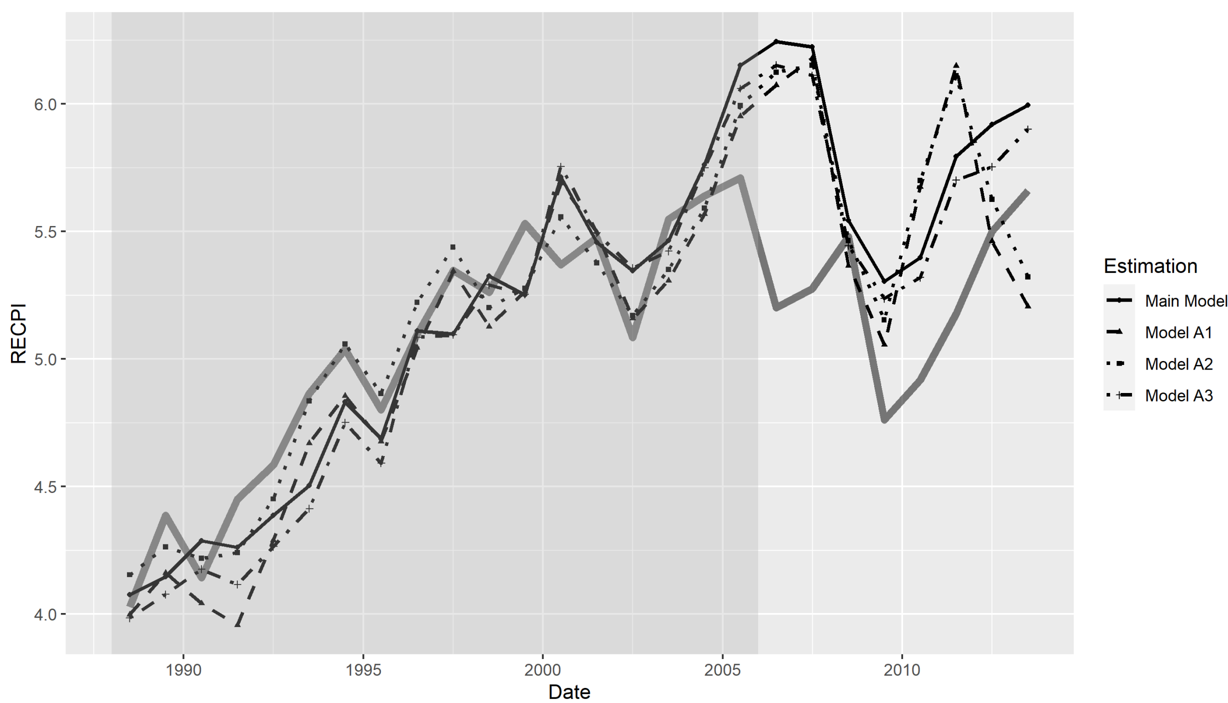
Notes: Dependent variable: quantity of new business registrations. Each column gives the weights given to each state in a particular model. The ”Main Model” is the model described in section 3, and for information on models A1, A2 and A3, see Appendix A.

Table A.5: RECPI Predictors and Fit

Predictor	Model A1	Model A2	Model A3	Main Model
State R&D	3.64309E-06	1.00000E+00	–	–
Fed R&D	1.14356E-07	1.00112E-08	8.47772E-07	1.99463E-06
Private sector output	4.95386E-08	1.00000E-08	1.91080E-08	1.72730E-08
State GDP	4.95386E-08	1.00000E-08	1.15973E-08	1.72730E-08
Venture cap.	3.71424E-01	1.13999E-08	1.92162E-04	2.14868E-04
SBIR funding	2.85807E-06	1.00831E-08	1.06535E-05	4.43246E-05
SET ind. emp.	4.95386E-08	1.00000E-08	–	3.19559E-06
Population	6.28569E-01	1.00000E-08	9.99796E-01	9.99725E-01
Degree holders	4.95386E-08	1.00000E-08	–	1.08729E-05

Notes: Dependent variable: quality adjusted quantity of new business registrations. Each column gives the weights given to each predictor in a particular model. The ”Main Model” is the model described in section 3, and for information on models A1, A2 and A3, see Appendix A. Not all models include all predictors, where not included, the weight is marked with ”–”.

Figure A.3: RECPI Comparison of Estimations



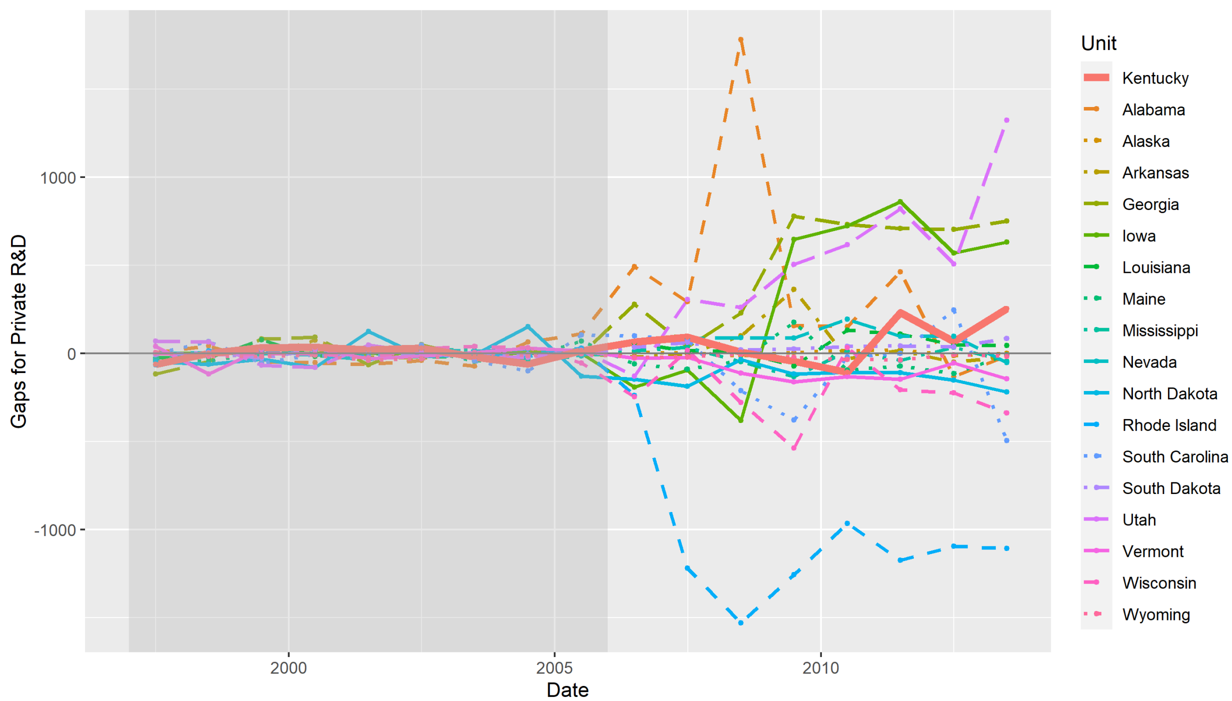
Note: The shaded area indicates the pre-intervention period and the unshaded area the post intervention period. Each line type represents a different model. The ”Main Model” is the model described in section 3, and for information on models A1, A2 and A3, see Appendix A. RECPI is short for ”Regional Entrepreneurship Cohort Potential Index”, a measure of quality adjusted new business registrations.

Table A.6: RECPI State Weights

	Model A1	Model A2	Model A3	Main Model
Alabama	0.000	0.000	0.000	0.000
Alaska	0.000	0.000	0.000	0.000
Arizona	0.000	0.000	0.000	0.000
Arkansas	0.000	0.000	0.138	0.009
California	0.000	0.000	0.000	0.000
Colorado	0.000	0.000	0.000	0.000
Georgia	0.000	0.000	0.000	0.000
Idaho	0.000	0.000	0.000	0.034
Iowa	0.000	0.218	0.503	0.555
Louisiana	0.707	0.653	0.234	0.243
Maine	0.000	0.000	0.000	0.000
Minnesota	0.000	0.000	0.000	0.000
Mississippi	0.000	0.000	0.000	0.016
Nevada	0.010	0.000	0.000	0.000
New Mexico	0.000	0.000	0.000	0.000
North Dakota	0.000	0.000	0.000	0.000
Ohio	0.000	0.000	0.000	0.000
Oregon	0.000	0.000	0.000	0.000
Pennsylvania	0.000	0.000	0.000	0.000
Rhode Island	0.000	0.000	0.000	0.000
South Carolina	0.000	0.011	0.000	0.000
South Dakota	0.252	0.000	0.000	0.000
Tennessee	0.000	0.100	0.054	0.116
Texas	0.031	0.018	0.030	0.027
Utah	0.000	0.000	0.000	0.000
Vermont	0.000	0.000	0.000	0.000
Washington	0.000	0.000	0.000	0.000
Wisconsin	0.000	0.000	0.042	0.000
Wyoming	0.000	0.000	0.000	0.000

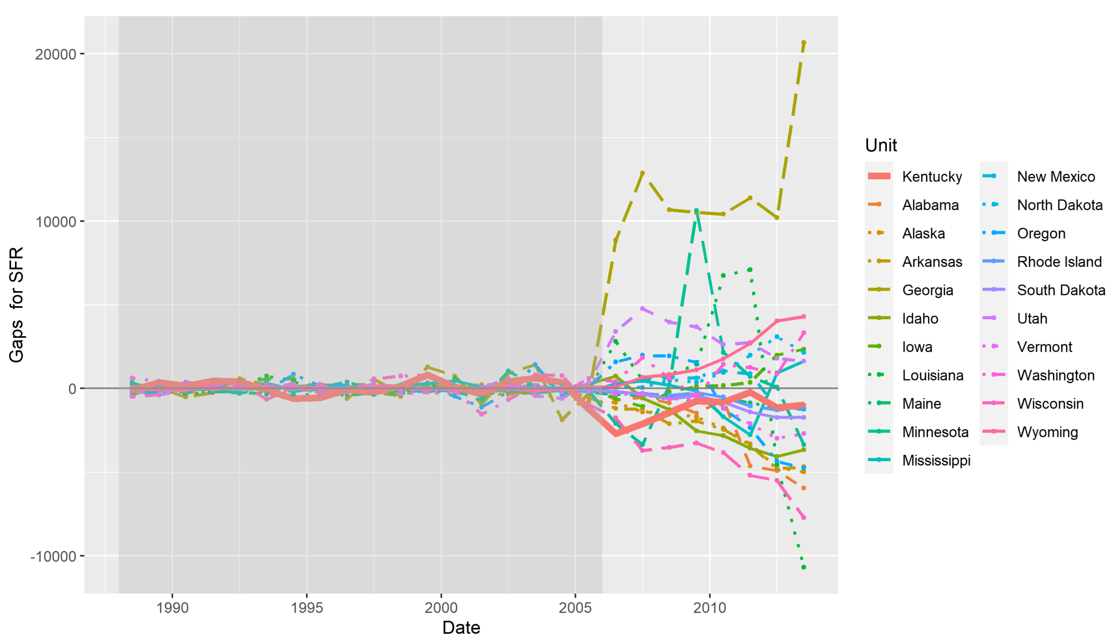
Notes: Dependent variable: quality adjusted quantity of new business registrations. Each column gives the weights given to each state in a particular model. The ”Main Model” is the model described in section 3, and for information on models A1, A2 and A3, see Appendix A.

Figure A.4: Private R&D Model A2 Placebo Test



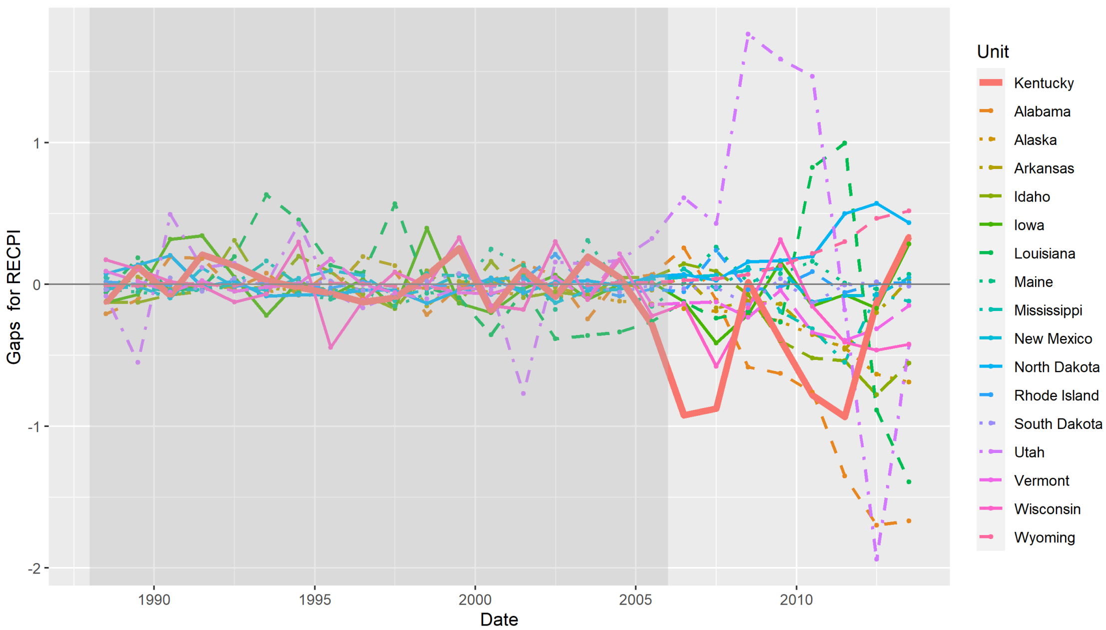
Note: Placebo test using Model A2. For information on the model, see Appendix A. For information on how to read the figure, see Figure 3.

Figure A.5: SFR Model A2 Placebo Test



Note: Placebo test using Model A2. For information on the model, see Appendix A. For information on how to read the figure, see Figure 5.

Figure A.6: RECPI Model A2 Placebo Test

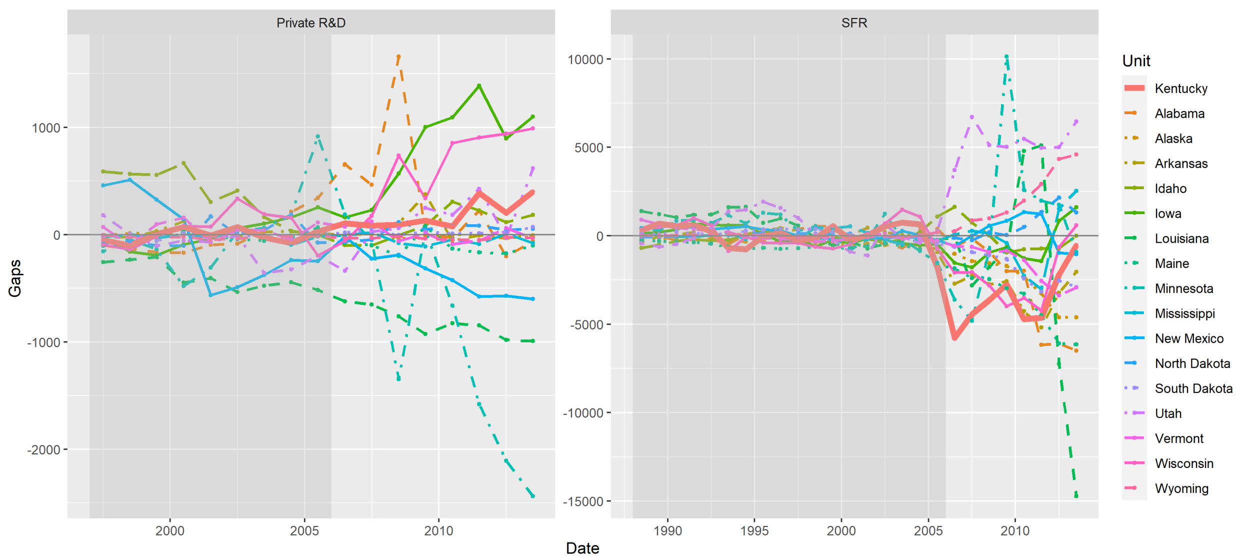


Note: Placebo test using Model A2. For information on the model, see Appendix A. For information on how to read the figure, see Figure 7.

Figure A.7: Model A4 Synthetic Control

Note: Results of Model A4. For information on the model, see Appendix A. For information on how to read the figures, see Figure 2 and Figure 4.

Figure A.8: Model A4 Placebo Test



Note: Placebo test using Model A4. For information on the model, see Appendix A. For information on how to read the figures, see Figure 3 and Figure 5.