

Vogt, William B.; Bhattacharya, Jayanta

Working Paper

Could we Tell if Health Insurance Mandates Cause Unemployment? A Note on the Literature

Working Paper, No. 167

Provided in Cooperation with:

George J. Stigler Center for the Study of the Economy and the State, The University of Chicago
Booth School of Business

Suggested Citation: Vogt, William B.; Bhattacharya, Jayanta (2001) : Could we Tell if Health Insurance Mandates Cause Unemployment? A Note on the Literature, Working Paper, No. 167, The University of Chicago, George J. Stigler Center for the Study of the Economy and the State, Chicago, IL

This Version is available at:

<https://hdl.handle.net/10419/262569>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Could we tell if health insurance mandates cause unemployment?

A note on the literature.

Jay Bhattacharya

RAND

and

William Vogt

Carnegie Mellon University

& NBER

March 16, 2000

“Theories that are right only fifty percent of the time are less economical than coin flipping.”

George Stigler, *The Theory of Price*

Introduction

Since the 1970s and at an increasing pace in the 1990s, states have adopted a wide variety of mandates designed to expand the set of medical care activities covered by health insurance. Among these mandates are requirements to cover maternity and mental health as well as requirements to cover the services of acupuncturists and nurse practitioners. Since in the United States, health insurance is typically tied tightly to the workplace, it is reasonable to expect that these mandates to have effects on employment and wages. Because the Employee Retirement Income Security Act (ERISA) effectively exempts large employers from state mandates, their effects should be greatest among small employers. One might expect that mandates raise the fixed costs of employment and reduce both wages and employment.

There have been several studies by economists, such as Gruber (1992,4) and Kaestner and Simon (2000), which evaluate the labor market effects of these policies. These studies tend to reach the same two conclusions: that mandates significantly reduce wages, but not employment. Using the variation in the timing of adoption of these mandates by states as a “natural experiment,” the studies rely on individual-level data from the Current Population Survey (CPS) to estimate the wage and employment effects of mandates. The studies typically find statistically insignificant effects on employment.

The main goal of this note is to evaluate the power of these natural experiments. If the employment effects are small in size relative to the “noise” in the data (but not small in size substantively), there is a risk of falsely accepting the hypothesis that there are no employment effects. There are two specific aims for our study:

1. Replicate an older paper (Gruber, 1994) and a newer paper (Kaestner and Simon, 2000) from the natural experiments literature on the employment effects of mandates
2. Calculate the power of these experiments to detect changes in employment due to the mandates

The papers we analyze find an absence of evidence for an employment effect of state health insurance mandates. Glibly, our research purpose is to discover whether “absence of evidence is evidence of absence” in this case.

Background

State Mandates

Among the first state mandates passed in the 1970s were requirements that health insurance cover maternity benefits. Most of the rhetoric surrounding the passage of these mandates centered on the unfairness of not covering a costly medical service used by women, who were entering the workplace at an increasing rate. After a lull in the early

1980s, the passage of state mandates picked up as mandates for mammograms, breast reconstruction after mastectomy, minimum inpatient stays after delivery became common, as did requirements relating to coverage of mental health, drug abuse and alcoholism treatment, and prostate cancer screening.

The appendix includes a table showing the year in which each state passed each specific mandate. As is evident from this table, the mandates are myriad in their nature, encompassing requirements for coverage of different types of medical care, different types of non-traditional providers, and different categories of dependents.¹

As a result of a series of court decisions over the 70s and 80s interpreting ERISA, self-insured employers (mostly large employers) became exempt from state health insurance mandates.

Economics of State Mandates:

Because private health insurance in the U.S. is provided primarily by employers, the economics of state mandates starts with most naturally with the theory of labor demand. Health insurance is just one component of a portfolio of benefits (which also include wages, retirement plans, and so on) that workers receive in exchange for their labor. The simple neo-classical model of the labor market predicts that the total compensation that workers will receive will equal their marginal product. Since the latter is unchanged by the state mandates, total compensation also should not change. This implies that for those workers who were receiving, prior to the passage of the law, non-compliant benefits, mandates will induce a reshuffling within the compensation portfolio without changing its total value. The mandates require the value of health insurance provided to increase, so it must necessarily decrease the value of the other components of the portfolio, though there is no theoretical prediction regarding which of these other components will decrease.

The theory also predicts another potential response by employers in response to mandates—eliminate the provision of health insurance from the compensation portfolio altogether. Indeed, there is considerable empirical evidence that health insurance coverage decreases when states pass mandates (see Simon 2000, Sloan and Connover 1997, and Sloan et al 2000). The literal effect of mandates is to ban certain types of contracts; in particular they ban health insurance contracts that do not cover the services or providers required by the mandate. By banning the low cost health insurance contracts (the ones that do not cover the mandated services), the plans make health insurance too expensive to provide for some employers and/or employees. However, if this is indeed what employers do, then the market will force them to increase the other components of the compensation portfolio as a substitute for the decrease in net compensation from cutting health insurance; wages, for example, may increase though total compensation stays the same.

¹ Our source for this Appendix table is the Blue Cross/Blue Shield Association. The table does not include all Federal mandates, which is sometimes important in interpreting the table. For example, the federal passage of mandated maternity care in 1978 stemmed the passage of similar state laws after that year. It also excludes mandates such as the one passed by Hawaii's Prepaid Health Care act of 1974, which mandates employer contributions toward health insurance premiums of their employees.

Unlike the wage effects, theoretical predictions regarding the employment effects of mandates are unequivocal—mandates cannot increase employment, and if there is any downward wage rigidity in the economy, mandates will decrease employment in the long run (see Zax et al. 1993, Woodbury and Hogan 1992, Morrissey 1991, Mitchell 1990, Flynn et al. 1997, and Summers 1989). While the analytic modeling on this point can be complicated, the logic here is simple; mandates can make certain types of labor, especially those with a marginal connection to the workforce, too expensive to hire.

To summarize, mandates can only change components of the compensation portfolio provided to workers, and cannot cause the total value of that portfolio to increase. To the extent that employers continue to provide health insurance, wages should tend to decrease with mandates, but if health insurance coverage is cut, wages can actually increase. Mandates raise the costs of hiring, leading to unemployment for workers who, after the passage of mandates, are too expensive to hire. Since large firms are by and large exempted from state mandates by ERISA, one should naturally expect the greatest share of economic effects to fall on small and medium sized firms.

Literature Review:

The two principal empirical papers in the literature on the labor market effects of state mandates are Gruber (1992,4) and Kaestner and Simon (2000). The former paper focuses on a particular type of state mandate—the requirement by states that insurance plans cover maternity care. In the 1970s, 12 states adopted such laws, and then in 1978 the federal government implemented the mandate nation-wide. At the time of passage of these laws, ERISA had not yet been interpreted as exempting self-insured plans from the mandates. Using these adoptions as “natural experiments,” Gruber finds that young working women of child-bearing age bear the costs of this mandate in reduced wages (relative to men and older working women), but finds no statistically significant evidence of decreased employment. Kaestner and Simon, rather than focusing just on mandates requiring maternity care, use all state mandates that have been passed in their analysis. Like Gruber, they find wage effects but insignificant employment effects.

Gruber(1992,4)

This paper examines the effect on labor markets of laws mandating that maternity benefits be a part of health insurance benefits.

The study uses a difference in differences in differences (DDD) approach. People are divided into treatment and control groups. Using the CPS, individuals along with their characteristics and state of residence are identified. This is done for several years: two years before and two years after a policy change. A treatment group is constructed consisting of married women between twenty and forty years of age (people likely to be most affected by mandating maternity benefits). The control group for this treatment group consists of single men between twenty and forty and all people older than forty. States are divided into experimental and non-experimental groups. In experimental states, at some known time a mandatory maternity benefits law comes into effect while in non-experimental states there was no change (either they had the law throughout or they did not have the law throughout).

The labor market outcome of interest (wages, hours, employment) is placed on the left hand side of an individual-level regression. On the right hand side are individual characteristics, state dummies, a dummy for membership of the state in the “experimental group”, time dummies, a dummy for “after” the policy takes effect², and a dummy variable for presence in the “Treatment Group.” All two-way interactions between “after”, “experimental”, and “treatment” are included as is the three-way interaction between “after”, “experimental”, and “treatment”. This three-way interaction is interpreted as embodying the effect of the law.

Gruber (1994) and we (in replication) analyze two different experiments. First, in New York, New Jersey, and Illinois mandatory maternity benefits laws came in to effect in the latter half of 1976. People in these experimental states are compared with people in the non-experimental states of Ohio, Indiana, Massachusetts, Connecticut, and North Carolina.³ The period before the policy change is implemented with the May CPS for 1974 and 1975. The period after is implemented with the May CPS for 1977 and 1978. Second, a federal mandated maternity benefits law came into effect in late 1978 (but its effects were delayed for most businesses until mid 1979). For this experiment, the twelve experimental states are those that did not have a mandated maternity benefit law in the two years before the federal law took place, while the non-experimental group consisted of the twenty-eight states that had a mandated maternity benefits law for all of the two years before the federal law took effect. The before period is implemented by the 1977 and 1978 May CPS, while the after period is implemented by the 1980 and 1981 May CPS.

Kaestner and Simon (2000)

Kaestner and Simon (2000) pursue a difference-in-differences (DD) approach to estimating the effect on the labor market of various state regulations on the small-group health insurance market. We do not have a current copy of the paper, as it was under revision as we were preparing this note.

Kaestner and Simon pursue a straightforward application of DD to their problem. They put labor market outcomes of interest on the left hand side of individual level regressions based on the March CPS over the period 1988-97. On the right hand side appear dummies for state, time, and the existence of various small group market reforms as well as individual characteristics and state time trends. They group the reform packages into no reform, partial reform, and full reform. The main difference between partial and full reform appears to be the inclusion of price regulation in the full reform package.

² Of course, “after” is collinear with the time dummies and “experimental” is collinear with the state dummies, so a time dummy and a state dummy (actually two of each to avoid collinearity with the intercept) are dropped to guarantee identification. The various interactions of these variables are the ones relevant for identification.

³ These states are all separately identifiable in the May CPS from those years, but not in the March CPS. This is the main reason to use the May CPS. Since Gruber (1994) examines only the labor market effects of these mandates, he does not need the information on health insurance coverage that is provided in the March CPS.

The coefficients on the full and partial reform dummies is interpreted as the effect of the reforms.⁴ One of their findings is that the effect of full reform on employment is insignificant.⁵

Other Literature

There is a large literature on state mandates that require employers to provide insurance. For example, Thurston (1997) and Dick (1994) offer analyses of the labor market effects of Hawaii's mandated provision law. Gruber and Hanratty (1993) analyze the phased adoption of the Canadian national health care plan, and perversely find that it led to increased employment in Canada. O'Neill and O'Neill (1993), Klerman and Goldman (1994), and CBO (1994) derive conflicting estimates of job loss from adopting the failed Clinton Health Care plan, though all predict job losses ranging from 100,000 to 1,000,000 workers. Unlike Gruber (1992,4) and Kaestner and Simon (2000) these studies do not use micro-level data to derive their estimates.

Methods

In this section, we discuss the Current Population Survey (CPS)—the main data set we use to conduct our research. This is the same data set used in the principal empirical papers in the literature on state mandates. Finally, we discuss our analytic strategy to address each specific aim.

Data

The Bureau of Labor Statistics collects the CPS monthly for the purpose of tracking employment and wage trends. The official government statistics on the unemployment rate are based upon the CPS. It is a large scale data set with over 400,000 respondents each month, who are representative of the nation as a whole. Every March, the CPS also asks its respondents about their health care coverage, including the source by which the respondents receive the insurance. Since state identifiers are available in the public release version of the CPS, we link the data in the Appendix to the multiple March CPS samples. This linked data set is similar to the one used by the main papers in the state mandates literature. In addition, we use multiple years of the May CPS to replicate part of the Gruber (1994) study.

In our replication of Gruber's work, we use the information on state mandates contained in his papers. In our replication of Kaestner and Simon, we use data from a Blue Cross and Blue Shield Association database of state health insurance mandates, described in their paper. However, our database is a slightly updated version of the one they use.

⁴ Kaestner and Simon also include other variables describing the various regulations, but we focus on the full and partial reform dummies.

⁵ Their measure of employment is the probability of being employed by a large firm --- presumably they are thinking that if the small group reforms have an employment effect, that this effect will operate by pushing people from small employers to large employers, rather than operating by pushing people from small employers to unemployed or out of the labor force.

Replicating Gruber (1994) and Kaestner and Simon (2001)

We replicate Gruber's individual-level regressions to estimate the change in employment probability from the mandates in both the state and federal experiments. Like him, we find no statistically significant employment effect of mandatory maternity benefits in both experiments. We use exactly the same specification of explanatory variables, and the same selection of people for inclusion in the study that Gruber used.⁶ The main difference Gruber's specification and ours is that we use the linear probability model (with heteroskedasticity-robust standard errors) rather than the probit model, since its coefficients are more readily interpretable.⁷ In addition to this replication, we calculate power functions and inverse power functions to estimate the how statistically important are these null findings.

Both because the law changes occurred at the state level and because we think it likely that the error terms in the equations will be correlated among residents of the same state, we think that the standard errors (although not necessarily the coefficients) from the individual level analysis will be biased and probably understated. A conservative way of dealing with this problem is to aggregate the estimating equation up to the state level. We aggregate not quite all the way to the state level --- we have 2 cells for each state-year: the treatment cell and the control cell. The left and right hand side variables from the individual level regression are averaged up to state-year-treatment cells and the regression is re-run again with heteroskedasticity-robust standard errors.

To analyze Kaestner and Simon (2000), we simply use their reported results in table six of the version we have. Since we do not aggregate up to the state level, the standard errors we use for our analysis may be biased and again we think likely they are too small.

Inverse Power Function and Power Function

The above two papers contain negative results --- they fail to reject the hypothesis that insurance market regulations have no effect on employment. We are interested in answering the question: "Do these results provide evidence that there was no effect, or do they merely show that the methods and data used cannot reveal any effect which was present?"

To do this we must do an analysis of the power of the tests. We employ two tools to analyze power.

Power Function

The traditional method for power analysis is the asymptotic power function. This is described, for example, in Davidson and MacKinnon (1993, section 12.3). The power

⁶ The equation is estimated using CPS data (March for the federal experiment and May for the state experiment) from the relevant years on members of the treatment group (married females aged between twenty and forty years, inclusive) and the control group (unmarried males aged between twenty and forty years inclusive and all persons aged between forty-one and sixty-four years, inclusive).

⁷ We also estimated probit models, and the results were not different in any important way. We present the linear probability model only for its ease of interpretation.

function (as we use it) plots the probability of rejecting the null hypothesis given the true value of a parameter. For example, Figure 1 is a power function for the effect of mandatory maternity benefits on employment. One point on this function is true effect of -2 and probability of 54.7% . This point says that were the true effect of mandatory maternity benefits to reduce employment by two percentage points, the probability of rejecting the null hypothesis of zero employment effect would be 54.7% . So, a “perfect” power function would be equal to the size of the test (0.05 throughout our note) at zero effect and 1 elsewhere.

Inverse Power Function

Andrews (1989) provides an elegant way of summarizing the information available in a power function. Hypothesis tests and confidence intervals are normally constructed by first choosing a size (probability of Type I error) and then constructing the hypothesis test and/or confidence interval so as to have no greater than the specified size. Andrews suggests constructing a summary measure of power by first choosing a level of power (one minus the probability of Type II error) and constructing the interval of potential true values of the relevant parameter against which a test has the chosen power or less.

He suggests the use of two levels of power: 50% and 1 -size. Consider the inverse power function using power equal to 50% . This yields the set of potential true values of a parameter against which a test has less than a 50% chance of rejecting the null. This is the set of potential true values of the parameter for which a hypothesis test is *no more powerful than flipping a coin*. Andrews calls this interval the region of low power, and it is the region in which the test is essentially useless to detect deviations from the null hypothesis. If there are economically significant values of the parameter in this region, we can conclude that the test is not useful to rule out economically significant effects.

The inverse power function using power equal to 1 -size (or 95% throughout our note), gives the interval of potentially true parameter values for which an acceptance of the hypothesis gives less certainty that the alternative is false than a rejection would give that the null is false. This interval is nice in that it is in some sense symmetric. If you want to be equally sure that an acceptance means the alternative is false as you are that a rejection means that the null is false, you must think only of alternatives outside the 95% inverse power function. Again, if there are economically significant values of the parameter in the 95% inverse power function, then we are not highly confident that the test has ruled out economically significant effects.

Finally, calculating the inverse power function is easy. The inverse power function for a single parameter at the 50% level of power (for a two tailed test of $H_0 : q = q_0$) is:

$(q_0 - 1.96\hat{S}_q, q_0 + 1.96\hat{S}_q)$. Similarly, for the 95% level of power:

$(q_0 - 3.605\hat{S}_q, q_0 + 3.605\hat{S}_q)$.

Results

Recall that Gruber analyzed two experiments: the federal experiment of 1978 and the NY-NJ-IL experiment of 1976. Recall further that, for each of these two experiments, we

replicated Gruber's individual-level regression and also performed a more aggregate state-level regression.

Gruber(1992/4)

Federal Experiment

Recall that the federal experiment is based on the change in mandated maternity benefits caused by the federal government's 1978 maternity benefits mandate. Gruber tests the effect of this law using a DDD probit regression. We attempt to "replicate" using a DDD linear probability model.

Gruber did not report a probability derivative for the federal experiment, so that it is not possible directly to verify the success of our replication quantitatively. However, qualitatively, we obtained similar results at both the individual and the state levels. The coefficient on the triple interaction is much smaller than its standard error and is therefore not significant at conventional levels. Furthermore, in the individual regression, the point estimate of the effect is also quite small. Looking at table 1, one can see our point estimates of the employment effect of mandated maternity benefits are an increase of 0.02 percentage points and a decrease of 0.61 percentage points at the individual and state levels, respectively. The standard errors associated with these two estimates are 0.96 and 1.40. This yields t statistics of 0.02 and -0.44.

However, it would be a grave error to conclude that there is no effect of these mandates on employment, for the DDD test is not at all powerful. First, consider that the 95% confidence interval for the individual level analysis includes an almost 2 percentage point decline in employment, and the 95% confidence interval for the state level analysis includes a 3 percentage point decline in employment.

Table 1: Gruber, federal experiment replication

Probability Derivatives		
Linear Probability Model		
Gruber	Individual	State
	0.02% (0.96)	-0.61% (1.40)
Inverse Power Functions		
Linear Probability Model		
Level	Individual	State
50%	(-1.89,1.89)	(-2.73,2.73)
95%	(-3.47,3.47)	(-5.03,5.03)

The table also contains inverse power functions for this test. For the individual analysis at the 50% level in table 1, we have calculated the inverse power function as $(-1.89, 1.89)$. This means that, if the true parameter value were to be between -1.89 and 1.89 , then the DDD test would fail to reject the null hypothesis of a zero effect at least 50% of the time. The inverse power function for the 50% level represents the set of alternatives against which a coin flip is more powerful than the DDD test. We (following Andrews) interpret the 50% level's inverse power function as the set of alternatives against which DDD is essentially useless. So, the inverse power functions in table 1 reveal that the DDD test is essentially useless against the alternative hypothesis of a 1.89 percentage point decline in the case of the individual level analysis and a 2.73 percentage point decline in the state level analysis.

For alternative hypotheses outside the 95% level's inverse power function, an acceptance in the DDD test is as strong evidence against the alternative as rejection of the null would be against the null. So, if we want to be as sure about acceptance as we are about rejection, we must consider an alternative of an employment decline of more than 3.47 percentage points in the individual level analysis and 5.03 percentage points in the state level analysis. So, based on the DDD test's acceptance, we can be quite confident that the employment effect of mandated maternity benefits was no larger than a reduction of 3.47 percentage points or 5.03 percentage points, depending on which level of aggregation you prefer.

Finally, in figures one and two, we present conventional power graphs based on the non-central chi-squared distribution. These graphs contain roughly the same information presented in table 1.

Figure 1: Gruber, Individual, Federal

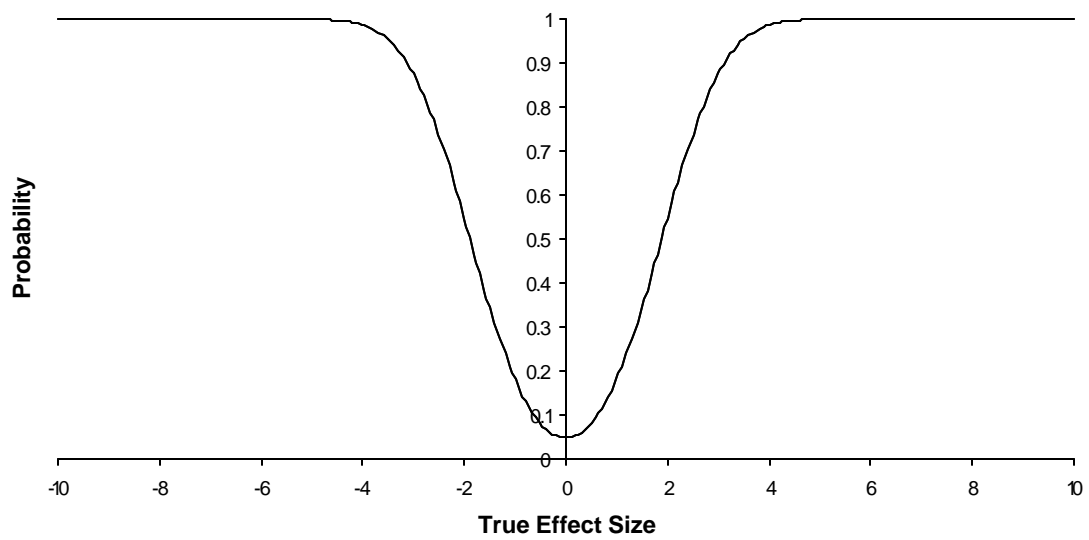
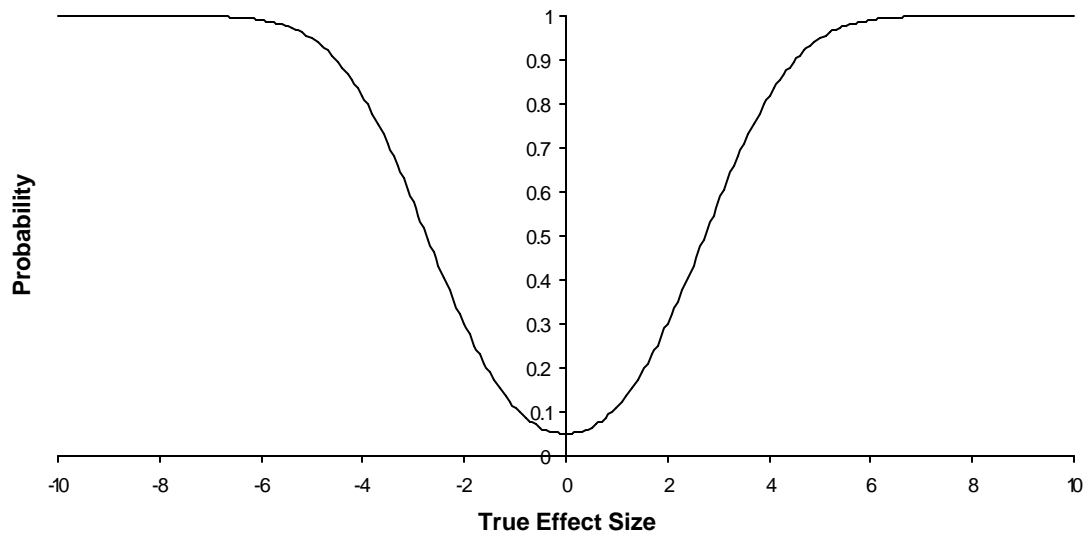


Figure 2: Gruber, State, Federal



State Experiment

Recall that the state experiment is based upon changes in state mandated maternity benefits in three states: New York, New Jersey, and Illinois. These three states passed mandated maternity benefits laws which went into effect in the second half of 1976. The control states in this analysis are Ohio, Indiana, Connecticut, Massachusetts, and North Carolina.

Gruber ran a probit regression at the individual level. He measured the effect size using a triply interacted dummy variable corresponding to his DDD approach. For this analysis, he calculates a probability derivative, which appears in table 4 and is -1.6 percentage points. In table 2, we show the results of our linear probability model, and it is clear that our individual analysis yields substantially the same results as Gruber's probit. Furthermore, the t-statistic in Gruber's probit was 0.98, and ours is 1.39. So, our analysis yields a similar point estimate for the probability derivative and also yields the same result in a conventional hypothesis test: we would both fail to reject the null hypothesis of zero employment effect.

It is not possible using the information reported in Gruber's paper to construct a proper inverse power function or power function for his probability derivative. We may proceed in a back-of-the-envelope way however. If we assume that the ratio of the probability derivative to its standard error is roughly the same as the ratio of the associated probit coefficient to its standard error,⁸ we would have a standard error for the probability derivative of 1.6. So, a 95% confidence interval for the effect size would be about $(-4.7, 1.5)$. Thus, the size of the employment effect could be as large as four and one half percentage points. A 50% inverse power function for the effect size would be $(-3.1, 3.1)$, so that the DDD test would be useless against an alternative as big as a three percentage point drop in employment. The 95% inverse power function for the effect size would be $(-5.8, 5.8)$, so that based on the acceptance by the DDD test, we would not be very confident that an effect size as big as five and a half percentage points was ruled out by the DDD test's failure to reject the null.

⁸ The probability derivative in a probit is $f(X\mathbf{b})\mathbf{b}_i$. Our back-of-the-envelope calculation amounts to ignoring the estimated quantity $f(X\hat{\mathbf{b}})$ as a source of variation in the probability derivative.

Table 2: Gruber, state experiment replication

Probability Derivatives		
Linear Probability Model		
Gruber	Individual	State
-1.6%	-1.56% (1.12)	-3.46% (1.76)
Inverse Power Functions		
Linear Probability Model		
Level	Individual	State
50%	(-2.20,2.20)	(-3.45,3.45)
95%	(-4.05,4.05)	(-6.35,6.35)

The results from our linear probability model are essentially similar. Using the individual level of analysis, the DDD test is useless against alternative hypotheses with employment drops of as much as 2.2 percentage points and does not provide strong evidence against the alternative until the alternative is a bigger than 4 percentage point drop in employment. Using the state level of analysis, the DDD test is useless against alternative hypotheses with employment drops of up to 3.45 percentage points, and the failure to reject with the DDD test does not provide strong evidence against alternatives until they call for more than 6 percentage point drops in employment.

Conclusion

The DDD test used in Gruber(1992,4) is not able to distinguish the null hypothesis of zero employment effect from alternative hypotheses of quite large employment effects. Even using the very liberal individual specification, the DDD test is no more powerful than a coin flip against alternative hypotheses calling for drops in employment of almost two percentage points. The test looks even worse under the more conservative state level analysis and worse yet when one demands the same level of confidence in an acceptance that one demands in a rejection. Using the state level analysis, acceptance of the null hypothesis does not provide convincing evidence against alternative hypotheses of five or six percentage point declines in employment, depending on the experiment. Absence of evidence is most assuredly not evidence of absence in this application.

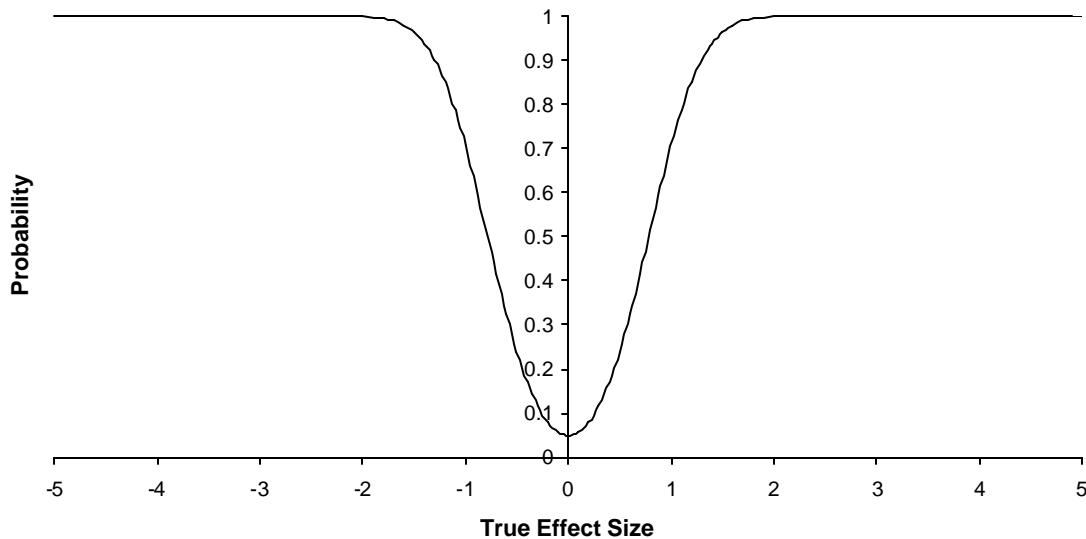
Kaestner and Simon (2000)

The employment results in Kaestner and Simon (2000) are based upon a fixed effects regression using the probability of being employed by a large firm as the dependent variable. According to the results on page 39 in table 6 of their work, the probability

derivative for “full reform” on employment in a large firm is 0.1 percentage points with a standard error of 0.4. The probability derivative for “partial reform” is -0.8 with a standard error of 0.4.

Since the full reform coefficient is insignificant at conventional levels, we perform our analysis on that. First, a 95% confidence interval for the effect of full reform on employment in a large employer is $(-0.7, 0.9)$. An inverse power function at the 50% level is $(-0.8, 0.8)$. So, the test is no better than flipping a coin as long as the true effect of full reform on employment in large firms is no greater than to increase it by 0.8 percentage points. An inverse power function at the 95% level is $(-1.44, 1.44)$. Thus, for employment effects as large as 1.44 percentage points, the failure to reject the null hypothesis is not as convincing with respect to the truth of the alternative as a rejection would be to the falsity of the null. Figure 5 presents this information in the form of a power graph.

Figure 5: K&S Full Reform



We also ran a specification similar to Kaestner and Simon’s. We ran a linear probability model on the 1989-98 March CPS selecting on employed people aged 18-54, non-self-employed, non-government-employed, and working in a business with less than 100 employees and holding health insurance in their own name ($N=85,637$).⁹ For our right-hand-side variables, we imitated the specification on page 39 in table 6 of their work --- we included dummies for full and partial reform, dummies for total mandates 4-6, 7-9, 10+, dummies for age categories, sex, race, schooling, marital status, number of children under 6 and under 18, and a full set of state and time dummies as well as state-specific time trends. All our standard errors are heteroskedasticity consistent.

⁹ This N does not match the N reported in Kaestner and Simon; although, we performed the selections suggested in their paper and we examined the source code they provided to us. There are other differences between our specification and theirs. We use an updated version of the Blue Cross and Blue Shield Association database. Also, we use a linear probability model, rather than the probit they employ.

In addition, we ran another specification in which we aggregated both the left and right hand side variables in the individual regression up to the state level. Again, we ran the regression with robust standard errors. Since we are interested in the effects of partial and full reform, we report those coefficients and related statistics in Table 3.

Our results are quite similar to those of Kaestner and Simon. In the individual level regressions, we find that the effect of partial reform on the probability of employment in large firms is to decrease it (significantly) by about 2 percent. Kaestner and Simon found a decrease of about 1 percent. The effect on employment in large firms of partial reform we estimate to be an (statistically insignificant) increase of about 0.2% compared to Kaestner and Simon's 0.1%. The state level regression was also similar, yielding statistically insignificant changes for each of the reforms. Partial reform decreased employment among large firms by 1.2%; whereas, full reform increased it by about 0.7%.

Table 3: Kaestner and Simon replication

Probability Derivatives		
	Linear Probability Model	
	Individual	State
Full	0.23% (0.94)	0.72% (0.81)
Partial	-2.1% (0.9)	-1.2% (0.8)
Inverse Power Functions, Partial Reform		
	Linear Probability Model	
Level	Individual	State
50%	(-1.76,1.76)	(-1.57,1.57)
95%	(-3.24,3.24)	(-2.88,2.88)
Inverse Power Functions, Full Reform		
	Linear Probability Model	
Level	Individual	State
50%	(-1.84,1.84)	(-1.59,1.59)
95%	(-3.39,3.39)	(-2.92,2.92)

Figure 5: K&S, Individual, Partial

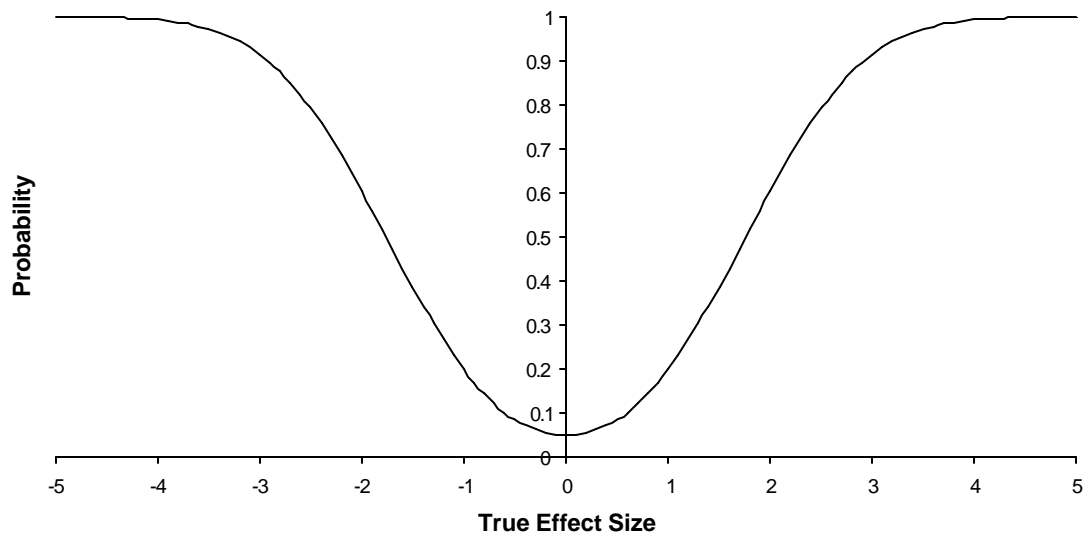


Figure 6: K&S, Individual, Partial

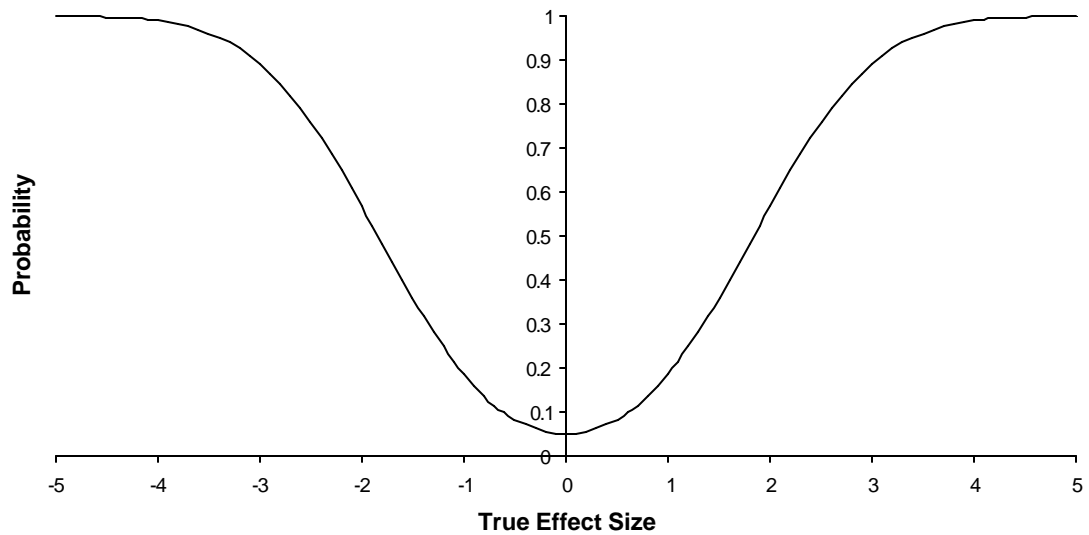


Figure 7: K&S, State, Partial

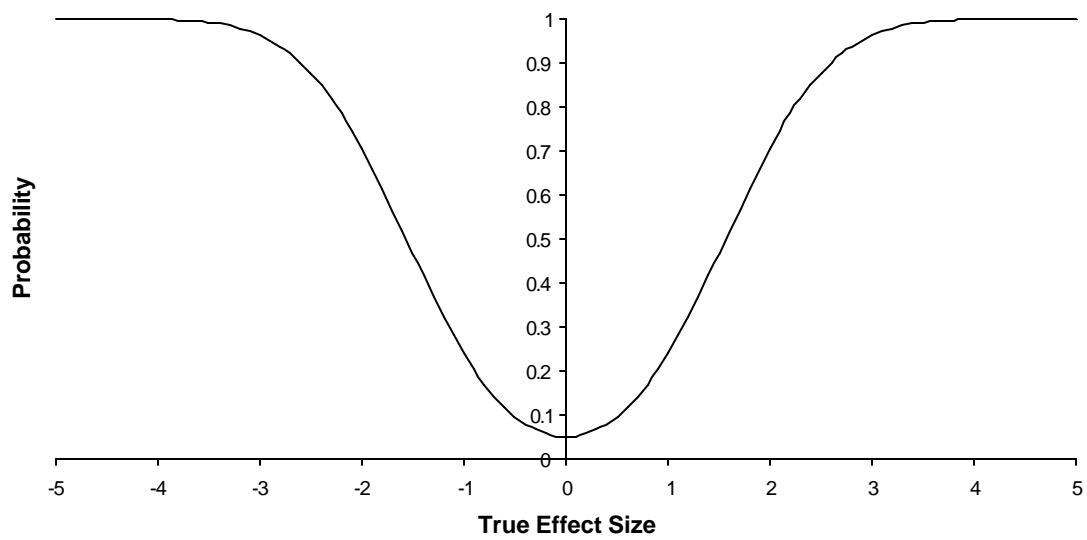
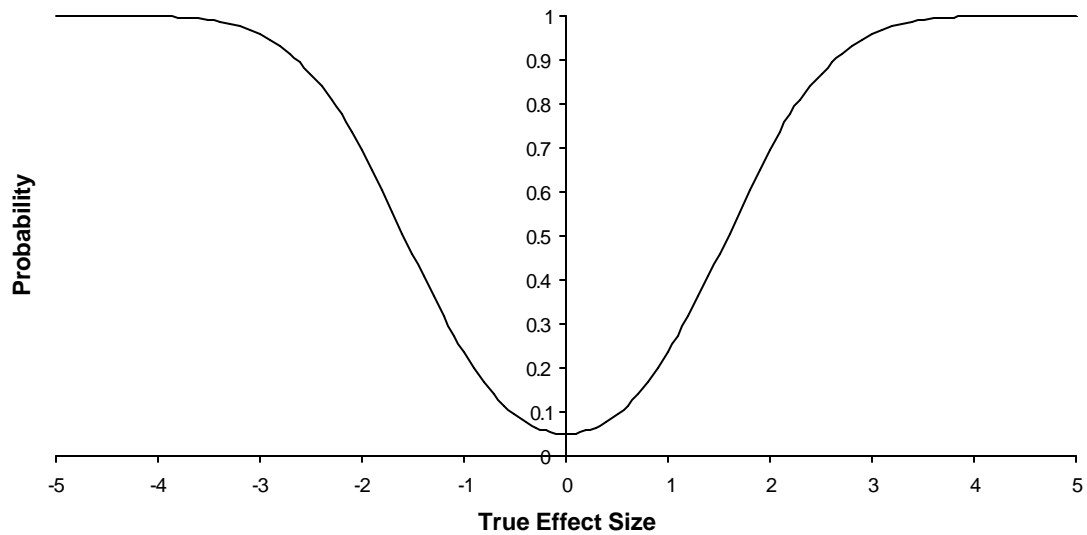


Figure 8: K&S, State, Full



As with Gruber's work discussed above, it would be a grave error to conclude from the insignificance of these coefficients that there is no effect of the laws on employment, however. For each of the estimators of full reform's effect, an employment increase of 2% in the large firms is within the 95% confidence interval of the estimators. For the partial reform estimator at the state level, an employment increase of 1% in the large firms is within the 95% confidence interval.

An analysis of the power of these tests reveals that they are poorly suited to detect empirically meaningful changes due to these reforms. Power functions for the tests of various coefficients equaling zero are presented in Figure 5 through Figure 8, and the inverse power functions appear in Table 3. The 50% inverse power functions for all of the reforms in all of the regressions contain employment effects as large as 1.5% in either direction. That means that, for employment effects as large as 1.5%, running the usual t-test in these regressions is no more informative about the alternative hypothesis than is flipping a coin. As the 95% inverse power functions reveal, in order for this test to be powerful, the alternative against which we are testing must be in the neighborhood of a 3% employment change. Thus, the insignificant findings are strong evidence that the true effect of the reforms is smaller than a 3% employment change, but they are not strong evidence that the effect of the reforms is smaller than, say, 2.5%.

As with Gruber's DDD analysis, the failure to reject the null hypothesis of no employment effect here is not at all strong evidence of the truth of the null hypothesis. We think reasonable people will agree that an employment effect of 1.5 or 2.5 percentage points is a large effect. Since the test cannot distinguish the null from these large effects, absence of evidence is again not evidence of absence.

Conclusion

It is not justified to conclude from the results of either of these papers that there is no employment effect of mandated health benefits. Neither paper's results can rule out large employment effects from the state health insurance mandates. That is, even if the true effect of the analyzed mandates were to be quite large, neither of the papers analyzed presents a methodology which would be capable of detecting this fact. In fact, employment reductions in the neighborhood of three percent are consistent with null results in these paper's tests. The proper conclusion to come to based on the evidence of the two papers is "It is not possible using these methods to determine whether or not health insurance mandates decrease employment." Possibly the effects are zero; possibly the effects are large and negative; possibly the effects are large and positive.

References

- Congressional Budget Office (1994) *An Analysis of the Administration's Health Proposal*. Washington DC: Congressional Budget Office.
- Dick A (1994) "Will Employer Mandates Really Work? Another Look at Hawaii," DataWatch Update, *Health Affairs* 1(Spring):343-49.
- Flynn P, Wade M, and Holahan J (1997) "State Health Reform: Effects on Labor Markets and Economic Activity," *Journal of Policy Analysis and Management*, 16(2):219-36.
- Gruber J (1994) "The Incidence of Mandated Maternity Benefits," *American Economic Review* 84(3):622-41.
- Gruber J and Hanratty M (1993) "The Labor Market Effects of Introducing National Health Insurance: Evidence from Canada," NBER working paper No. 4589.
- Kaestner R and Simon KI. (2000) "The Labor Market Consequences of State Health Insurance Regulation," mimeo, Michigan State University.
- Klerman JA and Goldman DP (1994) "Job Loss Due to Health Insurance Mandates," *JAMA* 272(7):552-6.
- Mitchell OS (1990) "The Benefits of Mandating Benefits Packages," *Research in Labor Economics* 11, pp. 297-320.
- Morrissey MA (1991) "Mandated Benefits and Compensating Differentials," paper prepared for the American Enterprise Institute Conference, Washington DC, October 3-4.
- O'Neill J and O'Neill D (1993) *The Impact of a Health Insurance Mandate of Labor Costs and Employment: Empirical Evidence*. Washington DC: Employment Policies Institute.
- Simon KI (2000) "Did Small Group Reform Work? Evidence from the March CPS 1992-1997." revise and resubmit at *American Economic Review*.
- Sloan FA, Conover CJ, and Hall MA (2000) "State Strategies to Reduce the Growing Numbers of People without Health Insurance," *Regulation* 22(3):24-31.
- Sloan FA and Conover CJ (1998) "Effects of State Reforms on Health Insurance Coverage of Adults," *Inquiry* 35:280.

Summers LH (1989) “What Can Economics Contribute to Social Policy? Some Simple Economics of Mandated Benefits,” *American Economics Review Papers and Proceedings* 79(2):177-83.

Thurston NK (1997) “Labor Market Effects of Hawaii’s Mandatory Employer-Provided Health Insurance,” *Industrial and Labor Relations Review* 51(1):117-135.

Woodbury SA and Hogan AJ (1992) “Labor Market Impacts of Policies to Expand Access to Health Care,” in JH Gooderis and AJ Hogan (eds.) *Improving Access to Health Care: What Can the States Do?* Kalamazoo, MI: WE Upjohn Institute for Employment Research.

Zax JH, Mocan N, Pearson D, and O’Connell J (1993) “The Economic Effects of Colorado Care,” report for the Office of the Governor, State of Colorado.