

Using Placebo Laws to Test “More Guns, Less Crime”: A Note

Eric Helland

George J. Stigler Center for the Study of the Economy and the State
Graduate School of Business
University of Chicago
and
Claremont McKenna College
1101 East 58th Street
Chicago IL 60637
ehelland@gsb.uchicago.edu

Alexander Tabarrok

Department of Economics, MSN 1D3
George Mason University
Fairfax, VA, 22030
Tabarrok@gmu.edu

Abstract

A boomlet has occurred in recent years in the use of quasi-natural experiments to answer important questions of public policy. The intuitive power of this approach, however, has sometimes diverted attention from the statistical assumptions that must be made, particularly regarding standard errors (Bertrand, Duflo and Mullainathan 2002, Donald and Lang 2001). Failing to take into account serial correlation and grouped data can dramatically reduce standard errors suggesting greater certainty in effects than is actually the case. We reexamine Mustard and Lott’s important and controversial study on the affect of “shall-issue” gun laws on crime using an empirical standard error function randomly generated from “placebo” laws. We find that in some specifications the effect of shall-issue laws on specific types of crimes is much less well-estimated than the Mustard and Lott (1997) and Lott (2000) results suggest (i.e. placebo shall-issue laws produce estimated real effects at greater rates than suggested by the standard errors in the original studies.) We also find, however, that the cross equation restrictions implied by the Lott-Mustard theory are strongly supported.

The authors wish to thank John Lott and Ian Ayers for generously sharing their data. We have also received helpful comments from John Donohue, John Lott and Florenz Plassmann. This work was undertaken while Helland was a visitor at the Stigler Center. He wishes to thank Sam Peltzman, the Olin Foundation and University of Chicago for their generous support. Tabarrok wishes to thank the Mercatus Center and the Center for Study of Public Choice for their generous support.

I. Introduction

Quasi-natural experiments, usually estimated with panel data and a difference-in-difference model (henceforth DD studies); have produced important results in the economics of immigration (Card 1990), abortion (Gruber, Levine and Staiger 1999), and crime (Mustard and Lott 1997, Lott 2000) as well as in many other areas.¹ The intuitive power of this approach, however, has sometimes diverted attention from the careful analysis of statistical assumptions particularly regarding standard errors (Bertrand, Duflo and Mullainathan 2002, Donald and Lang 2001). Failing to take into account serial correlation and grouped data can dramatically reduce standard errors thereby suggesting greater certainty in effects than is actually the case.

In a typical difference-in-difference study, individual outcomes are regressed on at least one variable that applies to all individuals in a group – for example, the wage rates of individuals living across the U.S. might be regressed on dependent variables such as state policies. Since each individual is treated as independent even though all individuals in a given state are potentially subject to common state-level shocks, standard errors can be biased downward. Although the basic point is familiar from Moulton (1990) the idea has not always been incorporated in difference-in-difference studies. Relatedly, Donald and Lang (2001) show that when using the correct asymptotics for difference-in-difference studies the t-statistic is not normally distributed.

Serial correlation is another familiar problem that is not always properly accounted for in DD studies (Bertrand, Duflo, and Mullainathan 2002 (BDM)). Positive serial correlation in the error term will cause standard errors to be understated and will do so to greater extent when the independent variable is also serially correlated. Since the variable of interest in a DD study is typically a policy variable represented by a dummy set to zero up to some point in time and 1 thereafter the relevant independent variable in DD studies is very serially correlated. State laws, for example, are not typically reevaluated every year so knowing that the law is in effect today tells us that the law is very likely to be in effect tomorrow and this dependence needs to be taken into account when calculating standard errors if, as is common, the error term also exhibits serial correlation.

¹ For a review see Meyer (1995) and Rosenzweig and Wolpin (2000).

2. The Placebo Law Technique

It is possible to fix both of these problems by making assumptions about the true error distribution and “correcting” the OLS standard errors in light of these assumptions. An alternative and generally superior approach is to create standard errors and critical values using an estimate of the true error distribution bootstrapped from the data (the technique is extensively discussed in Bertrand, Duflo and Mullainathan 2002; an earlier less extensive use is Bound, Jaeger and Baker 1995). Consider a typical DD study that estimates the effect of a state law using panel data from across the US states. For example,

$$y_{it} = \mathbf{a} \text{year}_t + \mathbf{b} x_{it} + \mathbf{d} T_{it} + \mathbf{f}_i + \mathbf{e}_{it} \quad (1)$$

where y_{it} is the outcome variable in question for county (or state) i in year t , year_t are year specific intercepts, x_{it} are control variables, T_{it} is a indicator variable equal to 1 if the law in question applies to observation i in period t , \mathbf{f}_i are county specific intercepts, and \mathbf{e}_{it} is the standard error.

To estimate the true error distribution randomly choose a time and a state and create a variable that indicates that at that time and in that state a policy intervention occurred. That is replace the actual law indicator T_{it} with P_{it} and estimate

$$y_{it} = \mathbf{a} \text{year}_t + \mathbf{b} x_{it} + \mathbf{d} P_{it} + \mathbf{f}_i + \mathbf{e}_{it} \quad (2)$$

Store the coefficients on the relevant variables. Repeat the procedure sufficiently many times to generate a probability distribution function for the relevant coefficients. The empirical PDF, called the placebo PDF, can then be used to identify the correct standard errors and critical values for the model estimated on the true data. Since the placebo PDF is generated from the actual data it correctly reflects the presence of grouped data, serial correlation, non-normality or other irregularities that would otherwise confound appropriate statistical inference.

The placebo PDF is so-called because of analogy to clinical trials for new pharmaceuticals. In a randomized clinical trial the effects of a new drug are not tested against a theoretical null of zero but against the effect of a placebo. The idea is similar

here except that our primary interest is not in comparing the size of a coefficient but in properly estimating the uncertainty with which the coefficient is measured.

3. The More Guns Hypothesis

We use this method to reexamine one of the most important and controversial policy analyses of recent times, Lott and Mustard's (1997) paper on shall-issue gun laws (expanded in Lott's book *More Guns, Less Crime* 2nd ed. 2000). "Shall-issue" laws essentially allow any citizen to obtain a permit for a concealed weapon, much like obtaining a driver's license. Lott and Mustard estimate the impact of "shall-issue" laws with the following specification (among others):

$$y_{it} = \mathbf{a} \text{ year}_t + \mathbf{b} x_{it} + \mathbf{d} \text{ shall}_{it} + \mathbf{f}_i + \mathbf{e}_{it}, \quad (3)$$

where (3) differs from (1) only in that y_{it} is explicitly the natural log of the relevant crime category per 100,000 people in county i in year t , shall_{it} is a dummy variable equal to one if county i was covered by a "shall-issue" law in period t , and \mathbf{e}_{it} is a robust (White) standard error weighted by county population.

There is considerable evidence that county level crime data is subject to "shocks" that are common to a group at a particular point time and that extend through time.² Such shocks would confound conventional standard error corrections and, if they exist may cause serious downward bias in the standard errors.

Lott and Mustard find that murder, rape and aggravated assault rates fall dramatically after the passage of shall-issue laws while property crime, auto theft, burglary and larceny increase. Lott and Mustard argue that the fall in crime occurs because shall-issue laws increase the proportion of potential crime victims who are armed and that the prospect of meeting an armed victim deters criminals.

Lott's work on guns and crime has been enormously influential. Lott's work, for example, was cited by eighteen state attorneys general in a letter to Attorney General John Ashcroft (July 8, 2002) in support of Ashcroft's decision to interpret the second

² For example, Grogger and Willis (2000) look at the impact of crack cocaine on local crime rates. Mustard (2002) examines the importance of omitted variable bias in estimating models using county level crime data and Maltz and Targonski (2002) examine how the FBI's data imputation method may create local shocks even when none actually exist.

amendment as protecting an individual right. Lott's work has also been cited in Congressional testimony and in testimony before a number of state legislatures (Ayres and Donohue 2002). Testing Lott's findings for robustness, therefore, is of great importance. We do not attempt, however, to reevaluate every aspect of Lott's data and statistical model. Other papers, in particular a lengthy study by Ayres and Donohue (2002) reestimate the model using different specifications, functional forms etc.³ In this note, we focus attention on the uncertainty surrounding the estimated effects. Our goal is to explain, illustrate and apply the "placebo law" technique to an issue of importance.

4. Results

The first section of Table 1 presents the results of the Lott and Mustard model where shall issue laws are modeled as a dummy variable taking on the value of 0 prior to the law and 1 after.⁴ A number of other variables including county and state fixed effects, arrest rates, population controls, measures of income etc. are also included in the model but are not shown in the table. We randomly generated 1000 placebo laws, ran regressions on each one of these generated datasets and computed beta coefficients. The resulting distribution of coefficients is the distribution under the null hypothesis of no effect.⁵ Thus, we then computed the standard deviation and the 97.5% and 2.5% centiles to find the standard error and critical values under the null hypothesis. As the table indicates the standard errors computed from the placebo distribution are typically 3 to 4 times larger than the original standard errors. The corrected 95 percent failure to reject region (i.e. the region bounded by the critical values) follow below the standard errors and in every case the estimated coefficient is well within the failure to reject region. Note that the fail to reject region is computed directly from the placebo PDF rather than

³ Ayers and Donohue (2002), for example, argue that the model should include a state specific time trend while Lott et al. (2002) argue that it should not. We follow Lott's model not because we necessarily agree that it is the most appropriate but because we wish to focus attention on the use of placebo laws to better estimate standard errors.

⁴ Results are not identical to those in Lott and Mustard because we use an expanded dataset constructed by Ayres and Donahue which is very similar to that of Lott and Mustard but with the "correction" of some minor errors and an expanded number of years. Correction is placed in quotes because the differences are debatable see Lott, Plassmann and Whitley (2002)

⁵ As should be expected given that these are placebo laws we found the mean estimated effects to be equal to zero within two significant digits, i.e. a typical mean was 0.00x.

based upon the corrected standard error. Thus the critical values do not rely on a normality assumption (although in this case the differences are not substantial).

Much of the criticism of Lott and Mustards' work has focused on the point estimates of "shall issue" laws on crime rates. Different sub-samples or aggregations of the data (for example, Black and Nagin (1998), Ayers and Donohue (2002), Olson and Maltz (2001)), different time periods (Ayers and Donohue (2002)) or different functional form (Dezhbakhsh and Rubin (1998), Ayers and Donohue (1999), Moody (2001); Plassmann and Tideman (2001), Duggan (2001)) have been tested but relatively little attention has focused on the standard errors. The failure to reject regions in Table 1 highlight the problem with this approach – essentially all the "revised" figures are within the critical values making these results nugatory! In other words, conventional statistical inference finds these estimates indistinguishable.

For comparison purposes, we have also included standard errors clustered on states, the usual prescription for Moulton type problems. Although the clustered standard errors are closer to the placebo distribution than the robust standard errors they are still a significant underestimate (by some 20-50 percent) – which is not surprising since clustering will not solve autocorrelation problems.⁶

Rather than focusing on the correct standard errors we can also calculate how often the standard procedures would have falsely rejected the null. In 1000 trials using placebo laws we found 282 out of 1000 instances in which the null hypothesis of no effect on the murder rate, for example, would have been rejected at the greater than 5% level! Of these 180 trials would have produced a coefficient ranging from 4.3 percent to 19.9 percent and averaging 9.1 percent. The remaining 102 cases where the null would have been falsely rejected had coefficients on murder ranging from -4.6 to -17.4 percent and averaging -9.0 percent. Thus, it is clear that the standard procedure can very easily lead to false inferences when all of the assumptions of the standard model are not met.

The data covers 1977-1992 and 10 shall-issue laws were passed post-1984. Since the exact process that generated the shall-issue laws is unknown it's an open question as to the best range from which to draw placebo laws. The first set of corrected standard

⁶ Ayers and Donohue, 2002 discuss the impact of clustering but do not focus on it and do not correct for serial correlation. Moody (2001) explicitly attempts to deal with serial correlation but uses parametric techniques that Bertrand et al. (2002) find are less accurate than the placebo technique.

errors comes from 10 randomly chosen states in random years post-1984. To check for robustness we look at other possible ranges to draw placebo laws from. In the second set of results we create n placebo laws for each post-1984 year where n is the number of states that actually passed a shall-issue law in that year. In 1989, for example, we create 3 placebo laws because in that year 3 states passed a shall-issue law. (Once a state has been drawn in some year it is removed from the list of possible states to be drawn from in later years.) We repeat this process 1000 times and compute the placebo PDF. The results are very similar to those found earlier. As before, in no case does the coefficient value fall outside the 95% critical values (not shown because very similar to that found earlier).

Since Lott and Mustard's original paper an additional 5 years of data have become available and during those five years (1992-97) 14 states adopted shall-issue laws.⁷ Does the additional data reduce the true uncertainty enough to draw firm conclusions? Not really. As before, the standard errors are much larger when a correct accounting is made of serial correlation and other factors (although as expected the standard errors do shrink with more data). None of the negative coefficients are statistically significant at the 5% level once the standard errors have been corrected. On the other hand, the positive coefficients on property crime, auto theft and larceny all remain statistically significant at the 5% level even after correction. One interpretation is that we are more certain that shall-issue laws *raise* non-violent crime rates than we are that shall-issue laws lower murder, rape aggravated assault and robbery. This interpretation would be naïve, however, because it ignores the cross-equation restrictions implicit in the Lott model. Incorporating these restrictions casts the shall-issue laws in more favorable light.

Lott and Mustard also estimate their model using before and after trends. Passage of a shall-issue law will not cause the population of potential victims to arm themselves overnight. An immediate before and after comparison, therefore, could easily miss important changes in the trend rates of crime. The final section of Table 1 presents the trend results from the original model. The coefficient on murder, for example, indicates

⁷ Since 1997 three additional states, Michigan, New Mexico, and Colorado have passed shall issue laws although the New Mexico Supreme Court prevented the law's implementation in that state.

that after shall-issue were passed the trend rate for murders fell by 4.3% compared to before the passage of the shall-issue law. As before we simulate this model 1000 times and generate a placebo PDF used to generate true standard errors and critical values. The standard errors are considerable larger than the original model suggests, up to twice as large in some cases. It is apparent that use of trends usefully absorbs some of the serial correlation discussed earlier so the relative increase in standard errors is not as large as with the dummy model. Importantly, the coefficient on murder remains outside the failure to reject region. Thus, even with the revised standard errors the trend model indicates that shall-issue laws cause a large and significant drop in the murder trend rate.

5. Cross-Equation Restrictions

The Lott-Mustard theory of shall-issue laws is that they deter crime by making it more likely for a criminal to encounter an armed victim. On this theory, shall-issue laws should lower crimes against persons but should have no direct effect on crimes against property such as property crime, auto theft, burglary or larceny. Indirectly shall-issue laws could increase crimes against property if there is a substitution effect away from crimes against persons (i.e. if criminals take more care to avoid victims when victims have a higher probability of being armed but continue to commit crimes.) From this perspective the negative coefficients on crimes against persons and the positive coefficients on crimes against property are potentially very informative. How likely is it that such a pattern would be produced by chance?

We can estimate the probability that a pattern is produced by chance by counting the number of times the pattern appears in data generated from randomly assigned placebo laws. In the 1977-1992 data generated by randomly choosing placebo states post-1985 we find only a .2% probability (i.e. 2 out of 1000) of finding negative coefficients on crimes against persons and positive coefficients on crimes against property. Note that this is a very strong test since we look only for the negative/positive pattern. For example, in the placebo data there are no examples of the pattern occurring when the coefficient on murder is as large or larger than -7.3% (i.e. we estimate that the probability of this occurring by chance is less than 1 in 1000). We find that the

probability of the pattern occurring by chance is similarly low using the set of placebo laws generated from the same years and random states. Using the longer dataset we find only 7 out of 1000 randomly generated cases fitting the negative/positive pattern. Only 4 fit the pattern with a coefficient on murder less than or equal to -7.8 and zero fit the pattern with coefficients on murder less than or equal to -7.8 and on property greater than or equal to 7.6%.

6. Conclusion

Accounting for serial correlation, grouped data and potential non-normalities can often be difficult especially when the exact form of the problem is not known. An alternative approach is to estimate standard errors from an empirical PDF generated via placebo laws. Using this approach we reexamined Lott and Mustard's important and controversial study on shall-issue laws. We find that that corrected standard errors are dramatically larger than the reported standard errors. Using a dummy model to estimate the effects we can rarely reject the null hypothesis of zero effect once the correct standard errors are used. In this sense, the true uncertainty about the effect of shall-issue laws is larger than has previously been estimated. Standard errors are also larger when using a trend model but in this specification the data continue to reject a null-hypothesis of no effect on the trend murder rate. The Lott-Mustard theory, moreover, contains a little remarked upon cross equation restriction. Shall-issue laws should reduce crimes against persons but increase crimes against property. Using the placebo approach we estimate the probability that these findings could occur by chance. Although individual coefficients are difficult to pin down the negative/positive pattern of results is very rare in the placebo data. Surprisingly, therefore, we conclude, that there is considerable support for the hypothesis that shall-issue laws cause criminals to substitute away from crimes against persons and towards crimes against property.

Table 1: The Estimated Impact of Shall Issue Laws on Crime, County Data

	Violent Crime	Murder	Rape	Aggravated Assault	Robbery	Property Crime	Auto Theft	Burglary	Larceny
Dummy Model (1977-1992)									
Coefficient	-3.5%*	-7.3%*	-4.8%*	-5.3%*	-0.1%	5.2%*	8.9%*	2.3%*	5.9%*
Original Standard Errors (robust)	1.2%	2.5%	1.5%	1.6%	1.9%	1.1%	2.0%	1.1%	1.9%
Placebo Standard Errors (post 85, random states)	4.9%	6.4%	5.6%	6.6%	7.5%	5.1%	6.5%	5.7%	5.7%
Critical Values/Failure to Reject Region	-9.2—10.7	-13.5—12.5	-12.3—11.0	-11.7—14.3	-15.4—14.2	-10.2—9.8	-12.0—13.9	-11.9—9.6	-11.1—10.4
Standard Errors Clustered on State (robust)	3.5%	5.5%	4.3%	4.8%	4.6%	2.6%	4.4%	3.5%	2.9%
Placebo Standard Errors (same years, random states)	4.6%	6.2%	5.1%	6.0%	6.7%	4.5%	6.1%	4.9%	5.3%
Dummy Model (1977-1997)									
Coefficient	0.2%	-7.8%*	-2.9%*	-0.1%	-0.4%	7.6%*	10.8%*	1.5%	9.6%*
Standard Errors (robust)	1.1%	1.7%	1.1%	1.3%	1.3%	0.8%	1.5%	0.9%	1.2%
Placebo Standard Errors (post 85, random states)	3.7%	5.4%	4.0%	5.2%	5.0%	3.6%	5.2%	3.9%	4.2%
Critical Values/Failure to Reject Region	-6.6—7.7	-11.0—10.2	-8.0—7.7	-9.8—10.1	-9.8—9.2	-7.4—6.5	-9.9—10.7	-7.8—7.4	-8.5—8.1
Trend Model									
Coefficient	-0.4%	-4.7%*	-1.7%*	0.5%	-1.9%**	0.1%	0.1%	-0.4%	0.8%
Original Standard Errors (robust)	0.5%	1.1%	0.6%	0.7%	0.8%	0.7%	0.9%	0.5%	1.4%
Placebo Standard Errors (post 85, random states)	1.1%	1.6%	1.0%	1.4%	1.4%	.9%	1.4%	1.0%	1.1%
Critical Values/Failure to Reject Region	-1.9—2.4	-2.9—3.1	-2.1—2.2	-2.5—2.9	-2.8—2.8	-1.7—1.8	-2.8—2.9	-2.1—2.1	-1.9—2.1

References

- Ayers, Ian and John Donohue (1999) "Nondiscretionary Concealed Weapons Law: A Case Study of Statistics, Standards of Proof and Public Policy," *American Law and Economic Review* 1:436-
- Ayers, Ian and John Donohue (2002) "Shooting Down the More Guns, Less Crime Hypothesis," Forthcoming *Stanford Law Review*.
- Black, Dan and Daniel Nagin (1998) "Do Right-To-Carry Laws Deter Violent Crime?" *Journal of Legal Studies*, 27:209-219.
- Bertrand, Marianne, Esther Duflo and Sendhil Mullainathan (2002) "How Much Should We Trust Difference-In-Difference Estimates?" NBER Working Paper 8841.
- Bound, John, David A. Jaeger and Regina M. Baker (1995) "Problems with Instrumental Variables Estimation When the Correlation Between the Instrument and the Endogenous Explanatory Variable is Weak." *Journal of the American Statistical Association* 90 (430): 443-450.
- Card, David (1990) "The Impact of the Mariel Boatlift on the Miami Labor Market," *Industrial and Labor Relations Review*, 43:245-257.
- Dezhbakhsh, Hashem and Paul H. Rubin. (1998) "The Economics of Gun Control Lives Saved or Lives Lost? The Effects of Concealed-Handgun Laws on Crime" *The American Economic Review*, 88(2):468-474
- Donald Stephen and Kevin Lang (2001) "Inference with Difference in Difference and Other Panel Data," Working paper.
- Duggan, Mark (2001) "More Guns, More Crime." *Journal of Political Economy* 109(5):1086-1114
- Grogger, Jeff and Michael Willis (2000) "The Emergence of Crack Cocaine and the Rise of Urban Crime Rates." *Review of Economics and Statistics* 82(4):519-529.
- Gruber, Jonathan, Phillip Levine and Doug Staiger (1999) "Abortion Legalization and Child Living Circumstances: Who is the "Marginal Child?" *Quarterly Journal of Economics*, 114(1): 263-292.
- Lott, John and David Mustard (1997) "Crime, Deterrence, and Right-To-Carry Concealed Handguns," *Journal of Legal Studies*, 26:1-68.
- Lott, John (2000) *More Guns, Less Crime*, University of Chicago Press.

Lott, John, Florenz Plassmann and John Whitley (2002) "Confirming More Guns, Less Crime." Forthcoming *Stanford Law Review*.

Maltz, Michael and Joseph Targonski (2002) "A note on the use of county level UCR data." *Journal of Quantitative Criminology* 18(3):297-318.

Meyer, Bruce (1995) "Natural and Quasi-Experiments in Economics," *Journal of Business and Economic Statistics* 13(2):151-161

Moody, Carlisle (2001) "Testing for the Effects of Concealed Weapons Laws: Specification Errors and Robustness," *Journal of Law and Economics*, XLIV:799-813.

Moulton, Brent (1990) "An Illustration of the Pitfall in Estimating the Effects of Aggregate Variables on Micro Units," *Review of Economics and Statistics*, 72:334-8.

Mustard, David (2002) "Re-examining Criminal Behavior: The Importance of Omitted Variable Bias" forthcoming *Review of Economics and Statistics*.

Plassmann, Florenz and T. Nicolaus Tideman (2001) Does the Right to Carry Concealed Handguns Deter Countable Crimes? Only a Count Analysis Can Say." *Journal of Law and Economics*, XLIV: 771-798\

Rosenweig, M. and K. Wolpin (1980) "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment," *Econometrica*, 48:227-240.