NBER WORKING PAPER SERIES

THE IMPACT OF RIGHT TO CARRY LAWS AND THE NRC REPORT: THE LATEST LESSONS FOR THE EMPIRICAL EVALUATION OF LAW AND POLICY

Abhay Aneja
John J. Donohue III
Alexandria Zhang

Working Paper 18294
http://www.nber.org/papers/w18294

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2012

The authors wish to thank David Autor, Alan Auerbach, Phil Cook, Peter Siegelman, and an anonymous referee for helpful comments, Todd Elder for assistance in understanding the technique of testing for omitted variable bias, Akshay Rao, Vikram Rao, and Andrew Baker for outstanding research assistance, and Stanford Law School and Yale Law School for financial support. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Impact of Right to Carry Laws and the NRC Report: The Latest Lessons for the Empirical Evaluation of Law and Policy
Abhay Aneja, John J. Donohue III, and Alexandria Zhang
NBER Working Paper No. 18294
August 2012, Revised January 2014
JEL No. K0

ABSTRACT

For over a decade, there has been a spirited academic debate over the impact on crime of laws that grant citizens the presumptive right to carry concealed handguns in public – so-called right-to-carry (RTC) laws. In 2004, the National Research Council (NRC) offered a critical evaluation of the "More Guns, Less Crime" hypothesis using county-level crime data for the period 1977-2000. 17 of the 18 NRC panel members essentially concluded that the existing research was inadequate to conclude that RTC laws increased or decreased crime. One member of the panel thought the NRC's panel data regressions showed that RTC laws decreased murder, but the other 17 responded by saying that "the scientific evidence does not support" that position.

We evaluate the NRC evidence, and improve and expand on the report's county data analysis by analyzing an additional six years of county data as well as state panel data for the period 1977-2010. We also present evidence using both a more plausible version of the Lott and Mustard specification, as well as our own preferred specification (which, unlike the Lott and Mustard model presented in the NRC report, does control for rates of incarceration and police). While we have considerable sympathy with the NRC's majority view about the difficulty of drawing conclusions from simple panel data models and re-affirm its finding that the conclusion of the dissenting panel member that RTC laws reduce murder has no statistical support. We disagree with the NRC report's judgment on one methodological point: while the NRC report states that cluster adjustments to correct for serial correlation are not needed in these panel data regressions, our randomization tests show that without such adjustments the Type 1 error soars to 21 - 70 percent.

Our paper highlights some important questions to consider when using panel data methods to resolve questions of law and policy effectiveness. We buttress the NRC's cautious conclusion regarding the effects of RTC laws by showing how sensitive the estimated impact of RTC laws is to different data periods, the use of state versus county data, particular specifications, and the decision to control for state trends. Overall, the most consistent, albeit not uniform, finding to emerge from both the state and county panel data models conducted over the entire period with and without state trends and using three different specifications is that aggravated assault rises when RTC laws are adopted. If one narrows the focus to the most complete data (state data over the entire 1977-2010 period) or the period from 1999-2010 (thereby removing the confounding influence of the crack cocaine epidemic) and looks at the dummy and spline models using our preferred specification, then there is always evidence within the four estimates for each of the seven crime categories that RTC laws are associated with higher rates of crime. In six of the

seven crime categories, the finding that RTC laws increase crime is statistically significant at the .05 level, and for robbery, it is statistically significant at the .10 level. It will be worth exploring whether other methodological approaches and/or additional years of data will confirm the results of this panel-data analysis.

Abhay Aneja
Stanford Law School
559 Nathan Abbott Way
Stanford, CA 94305
aaneja@stanford.edu

John J. Donohue III
Stanford Law School
Crown Quadrangle
559 Nathan Abbott Way
Stanford, CA 94305 and
NBER
donohue@law.stanford.edu

Alexandria Zhang
Johns Hopkins University
Baltimore, Maryland
azhang4@jhu.edu

I. Introduction

The debate on the impact of “shall-issue” or “right-to-carry” (RTC) concealed handgun laws on crime—which has now raged on for over a decade—is a prime example of the many difficulties and pitfalls that await those who try to use observational data to estimate the effects of controversial laws.[2] John Lott and David Mustard initiated the "More Guns, Less Crime" discussion with their widely cited 1997 paper arguing that the adoption of RTC laws has played a major role in reducing violent crime. However, as Ayres and Donohue (2003a) note, Lott and Mustard’s period of analysis ended just before the extraordinary crime drop of the 1990s. They concluded that extending Lott and Mustard’s dataset beyond 1992 undermined the “More Guns, Less Crime” (MGLC) hypothesis. Other studies have raised further doubts about the claimed benefits of RTC laws (for example, see Black and Nagin, 1997 and Ludwig, 1998).

But even as the empirical support for the Lott and Mustard thesis was weakening, its political impact was growing. Legislators continued to cite this work in support of their votes on behalf of RTC laws, and the “More Guns, Less Crime” claim has been invoked often in support of ensuring a personal right to have handguns under the Second Amendment. In the face of this scholarly and political ferment, in 2003, the National Research Council (NRC) convened a committee of top experts in criminology, statistics, and economics to evaluate the existing data in hopes of reconciling the various methodologies and findings concerning the relationship between firearms and violence, of which the impact of RTC laws was a single, but important, issue. With so much talent on board, it seemed reasonable to expect that the committee would reach a decisive conclusion on this topic and put the debate to rest.

[2] The term “RTC laws” is used interchangeably with “shall-issue laws” in the guns and crime literature.

The bulk of the NRC report on firearms, which was finally issued in 2004, was uncontroversial. The chapter on RTC laws was anything but. Citing the extreme sensitivity of point estimates to various panel data model specifications, the NRC report failed to narrow the domain of uncertainty about the effects of RTC laws. Indeed, it may have broadened it. However, while the NRC report concluded there was no reliable statistical support for the “More Guns, Less Crime” hypothesis, the vote was not unanimous. One dissenting committee member argued that the committee's own estimates revealed that RTC laws did in fact reduce the rate of murder. Conversely, a different member went even further than the majority’s opinion by doubting that *any* econometric evaluation could illuminate the impact of RTC laws.

Given the prestige of the committee and the conflicting assessments of both the substantive issue of RTC laws' impact and the suitability of empirical methods for evaluating such laws, a reassessment of the NRC’s report would be useful for researchers seeking to estimate the impact of other legal and policy interventions. Our systematic review of the NRC's evidence—its approach and findings—also provides important lessons on the perils of using traditional observational methods to elucidate the impact of legislation. To be clear, our intent is not to provide what the NRC panel could not—that is, the final word on how RTC laws impact crime. Rather, we show how fragile panel data evidence can be, and how a number of issues must be carefully considered when relying on these methods to study politically and socially explosive topics with direct policy implications.

The outline of this paper is as follows. Section II offers background on the debate over RTC laws, and Section III describes relevant aspects of the NRC report in depth. Section IV discusses how the NRC majority presented some panel data models based on the Lott and Mustard specification in support of the conclusion that one could not reach a definitive conclusion about the impact of RTC laws. While this conclusion was correct, the models contained an array of errors that opened the door for the Wilson dissent to argue that RTC laws reduce murder. We discuss these errors in depth and show that Wilson would have been unable to make his dissent if the errors in the presented models (and standard error calculations) had been corrected. Sections V and VI explore two key econometric issues in evaluating RTC laws—whether to control for state-specific trends (which the NRC panel did not address) and whether to adjust standard errors to account for serial or within-group correlation (we show that the NRC report was in error when it concluded such adjustment was not needed). Section VII extends the analysis through 2006, and Section VIII offers improvements to the NRC model by revising the regression specification in accordance with past research on crime. Section IX discusses the issue of whether the impact of RTC laws can be better estimated using county- or state-level data. Section X delves further into the issue of omitted variable bias in assessing the impact of RTC laws, and in particular, how the difficult-to-measure effect of the crack epidemic may influence our estimates. Section XI offers concluding comments on the current state of the research on RTC laws, the difficulties in ascertaining the causal effects of legal interventions, and the dangers that exist when policy-makers can simply pick their preferred study from among a wide array of conflicting estimates.

II. Background on the Debate

In a widely-discussed 1997 paper, "Crime, Deterrence, and Right-to-Carry Concealed Handguns," John Lott and David Mustard (1997) argued, based on a panel-data analysis, that right-to-carry laws were the primary driving force behind falling rates of violent crime. Lott and Mustard used county-level crime data (including county and year fixed effects, as well as a set of control variables) to estimate the impact of RTC laws on crime rates over the time period 1977-1992. In essence, Lott and Mustard's empirical approach was designed to identify the effect of RTC laws on crime in the ten states that adopted them during this time period. Using a standard difference-in-difference model, the change in crime in the ten RTC states is compared with the change in crime in non-RTC states. The implicit assumption is that the controls included in the regression will explain other movements in crime across states, and the remaining differences in crime levels can be attributed to the presence or absence of the RTC laws.

Lott and Mustard estimated two distinct difference-in-difference-type models to test the impact of RTC laws: a dummy variable model and a trend, or "spline," model.[3] The "dummy model" tests whether the average crime level in the pre-passage period is statistically different from the post-passage crime level (after controlling for other factors). The "spline model" measures whether crime *trends* are altered by the adoption of RTC laws. Lott and Mustard noted that the spline approach would be superior if the intervention caused a reversal in a rising crime rate. Such a reversal could be obscured in a dummy variable model that only estimates the average change in crime between the pre- and post-passage periods. An effective RTC law might show no effect in the dummy model if the rise in the pre-passage crime rate and the fall in the post-passage rate were to leave the average "before" and "after" crime levels the same.

In both regression models, Lott and Mustard included only a single other criminal justice explanatory variable -- county-level arrest rates -- plus controls for county population, population density, income, and thirty-six(!) categories of demographic composition. As we will discuss shortly, we believe that many criminological researchers would be concerned about the absence

[3] In the "dummy model," RTC laws are modeled as a dummy variable which takes on a value of one in the first full year after passage and retains that value thereafter (since no state has repealed its RTC law once adopted). In the "trend model," RTC laws are modeled as a spline variable indicating the number of years post-passage.

of important explanatory factors such as the incarceration rate and the level of police force.

Lott and Mustard's results seemed to support the contention that laws allowing the carry of concealed handguns lead to less crime. Their estimates suggested that murder, rape, aggravated assault, and overall violent crime fell by 4 to 7 percent following the passage of RTC laws. In contrast, property crime rates (auto theft, burglary, and larceny) were estimated to have increased by 2 to 9 percent. Lott and Mustard thus concluded that criminals respond to RTC laws by substituting violent crime with property crime to reduce the risk that they would be shot (since, according to them, victims are more often absent during the commission of a property crime). They also found that the MGLC contention was strengthened by the trend analysis, which ostensibly suggested significant *decreases* in murder, rape, and robbery (but no significant increases in property crime).

From this evidence, Lott and Mustard (1997) concluded that permissive gun-carrying laws deter violent crimes more effectively than any other crime reduction policy: "concealed handguns are the most cost-effective method of reducing crime thus far analyzed by economists, providing a higher return than increased law enforcement or incarceration, other private security devices, or social programs like early education." They went even further by claiming that had remaining non-RTC states enacted such legislation, over 1,400 murders and 4,100 rapes would have been avoided nationwide, and that each new handgun permit would reduce victim losses by up to $5,000.

A. The Far-Reaching Impact of "More Guns, Less Crime"

The first "More Guns, Less Crime" paper and Lott's subsequent research (and pro-gun advocacy) have had a major impact in the policy realm. Over the past decade, politicians as well as interest groups such as the National Rifle Association have continually trumpeted the results of this empirical study to oppose gun control efforts and promote less restrictive gun-carrying laws. Lott relied on his own research to advocate for the passage of state-level concealed-carry gun laws, testifying on the purported safety benefits of RTC laws in front of several state legislatures, including Nebraska, Michigan, Minnesota, Ohio, and Wisconsin (Ayres and Donohue 2003a).

The impact of the Lott-Mustard paper can also be seen at the federal level. In 1997, ex-Senator Larry Craig (R-Idaho) introduced the Personal Safety and Community Protection Act with Lott's research as supporting evidence. This bill was designed to allow state nonresidents with valid handgun permits in their home state to possess concealed firearms (former football athlete Plaxico Burress sought to invoke this defense when he accidentally shot himself in a Manhattan nightclub with a gun for which he had obtained a Florida permit). According to Craig, Lott's work confirmed that positive externalities of gun-carrying would result in two ways: by affording protection for law-abiding citizens during criminal acts, and by deterring potential criminals from ever committing offenses for fear of encountering an armed response.[4] Clearly, Lott's work has provided academic cover for policymakers and advocates seeking to justify the view—on public safety grounds—that the 2nd Amendment conferred a private right to possess handguns.

B. Questioning "More Guns, Less Crime"

Immediately after the publication of the Lott-Mustard paper, scholars started raising serious questions about the theoretical and empirical validity of the "More Guns, Less Crime"

[4] 143 CONG. REC. S5109 (daily ed. May 23, 1997) (statement of Sen. Craig). The bill was again introduced in 2000 by Congressman Cliff Stearns (R-Florida), who also cited Lott's work. 146 CONG. REC. H2658 (daily ed. May 9) 2000) (statement of Rep. Stearns).
Indeed, this proposed legislation, now derisively referred to as "Plaxico's Law," is a perennial favorite of the NRA and frequently introduced by supportive members of Congress (Collins 2009).

hypothesis. For example, Zimring and Hawkins (1997) claimed that the comparison of crime between RTC and non-RTC states is inherently misleading because of factors such as deprivation, drugs, and gang activity, which vary significantly across gun-friendly and non-gun-friendly states (and are often difficult to quantify). To the extent that the relatively better crime performance seen in shall-issue states during the late 1980s and early 1990s was the product of these other factors, researchers may be obtaining biased impact estimates. Underscoring this point, Ayres and Donohue (2003a) pointed out that crime rose across the board from 1985 to 1992, and most dramatically in non-RTC states. Since the Lott and Mustard data set ended in 1992, it could not capture the most dramatic reversal in crime in American history. Figures 1-7 depict the trends of violent and property crimes over the period 1970-2007. For each of the seven crimes, the 50 states (plus the District of Columbia) fall into four groupings: non-RTC states (those states who had not passed RTC laws by 2006), states that adopted RTC laws over the period 1985-1988 ("early adopters"), those that adopted RTC laws over the period 1989-1991 ("mid-adopters"), and those that adopted RTC laws over the period 1994-1996 ("late adopters"). The crime rate shown for each group is a within-group average, weighted by population. The figures corroborate Ayres and Donohue's point: crime rates declined sharply across the board beginning in 1992. In fact, there was a steady *upward* trend in crime rates in the years leading up to 1992, most distinctly for rape and aggravated assault. Moreover, the average crime rates in non-RTC states seemed to have dropped even more drastically than those in RTC states, which suggests that crime-reducing factors other than RTC laws were at work.

Figure 1:

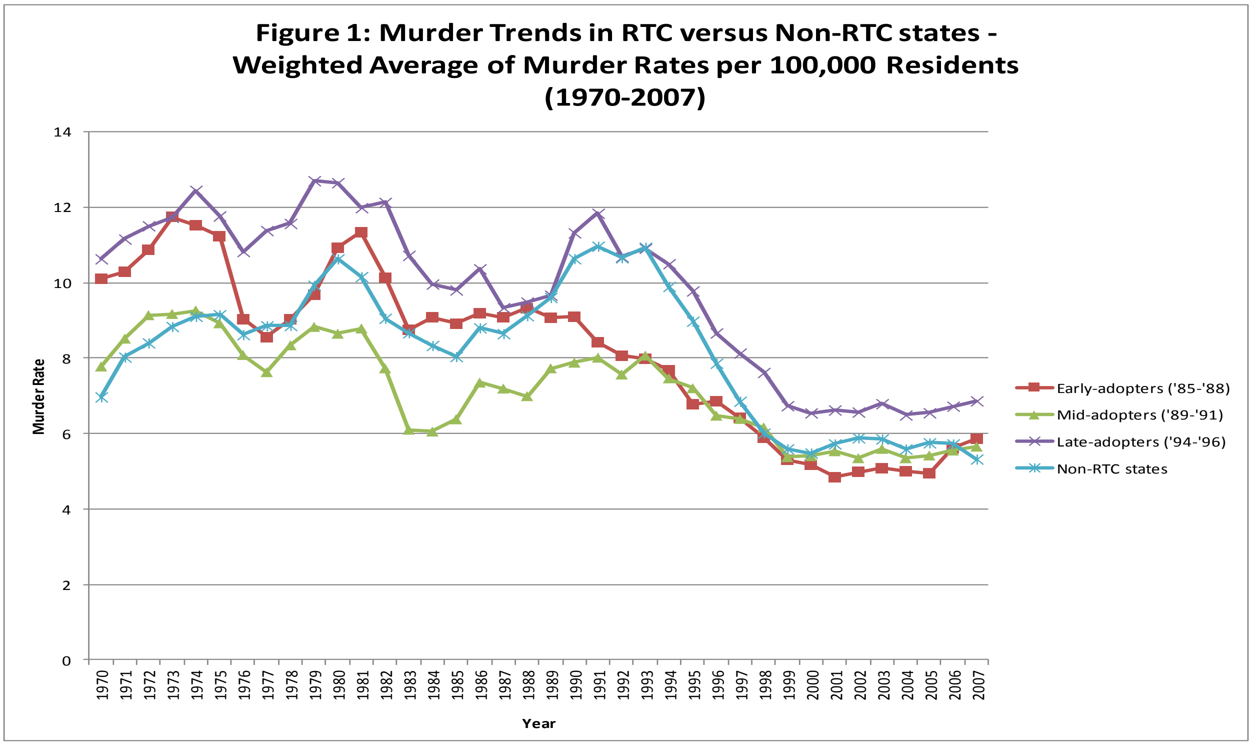


Figure 2:

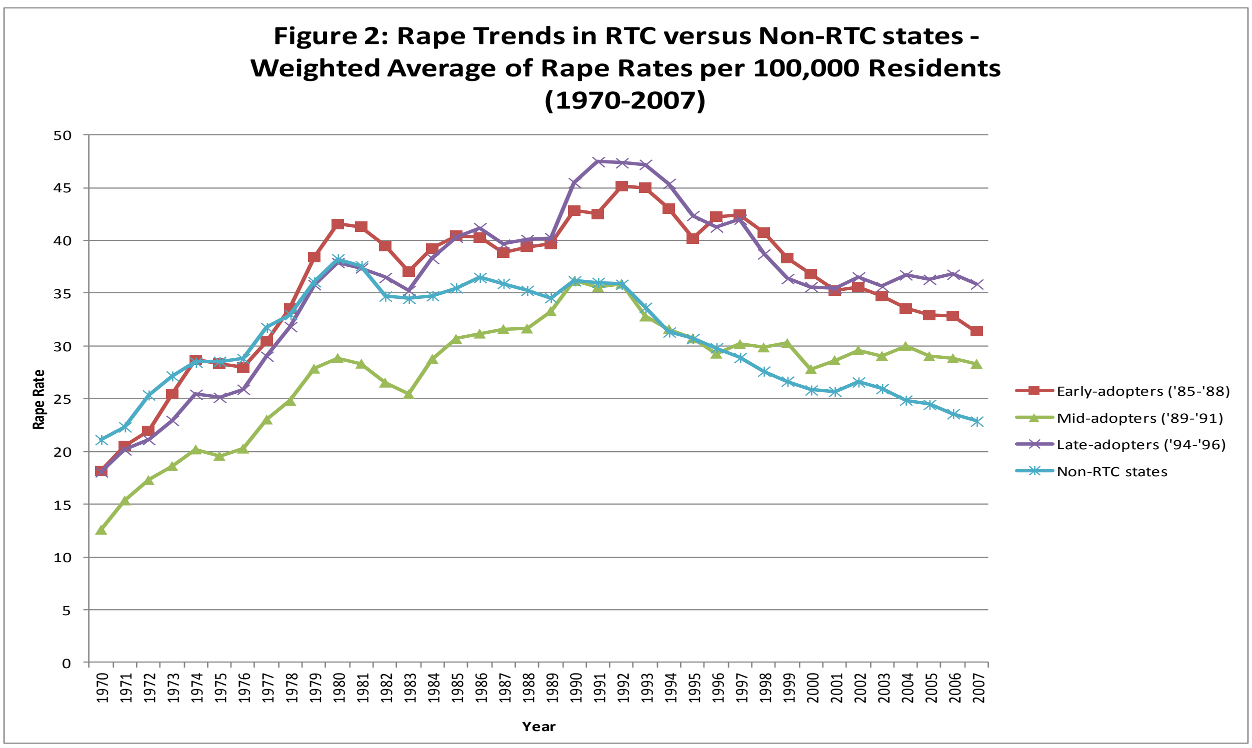


Figure 3:

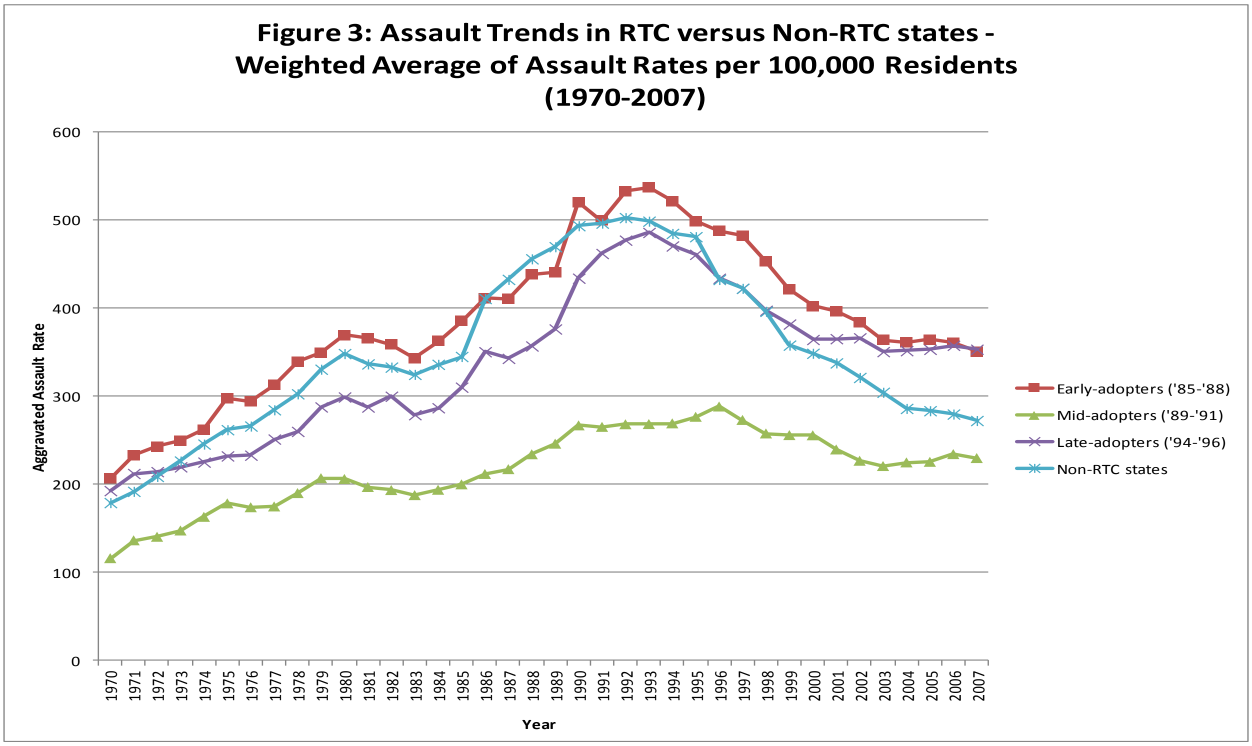


Figure 4:

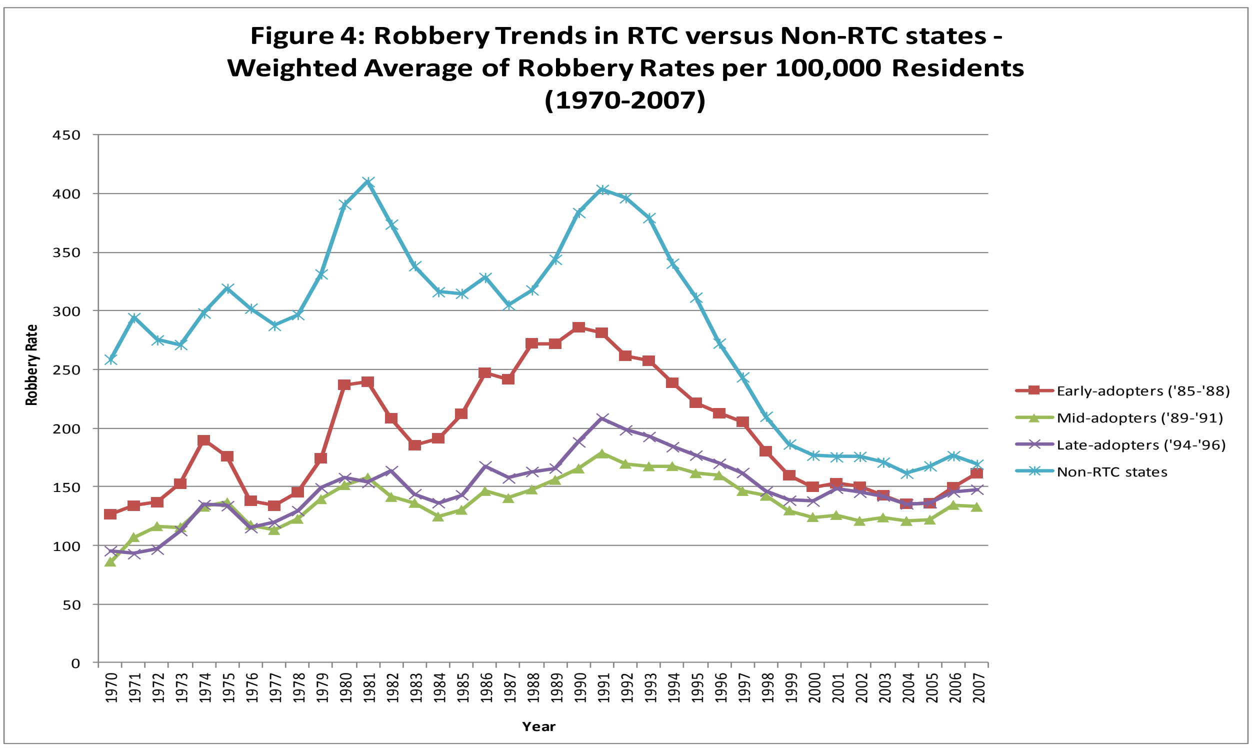


Figure 5:

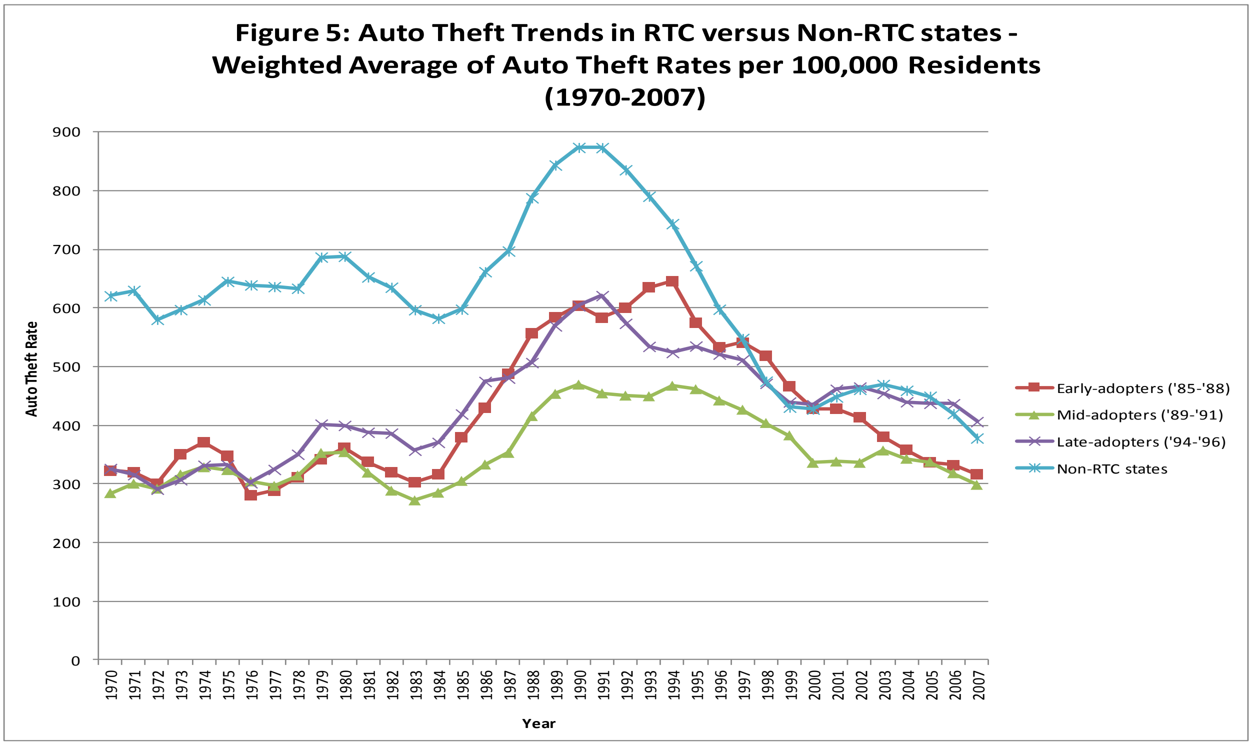


Figure 6:

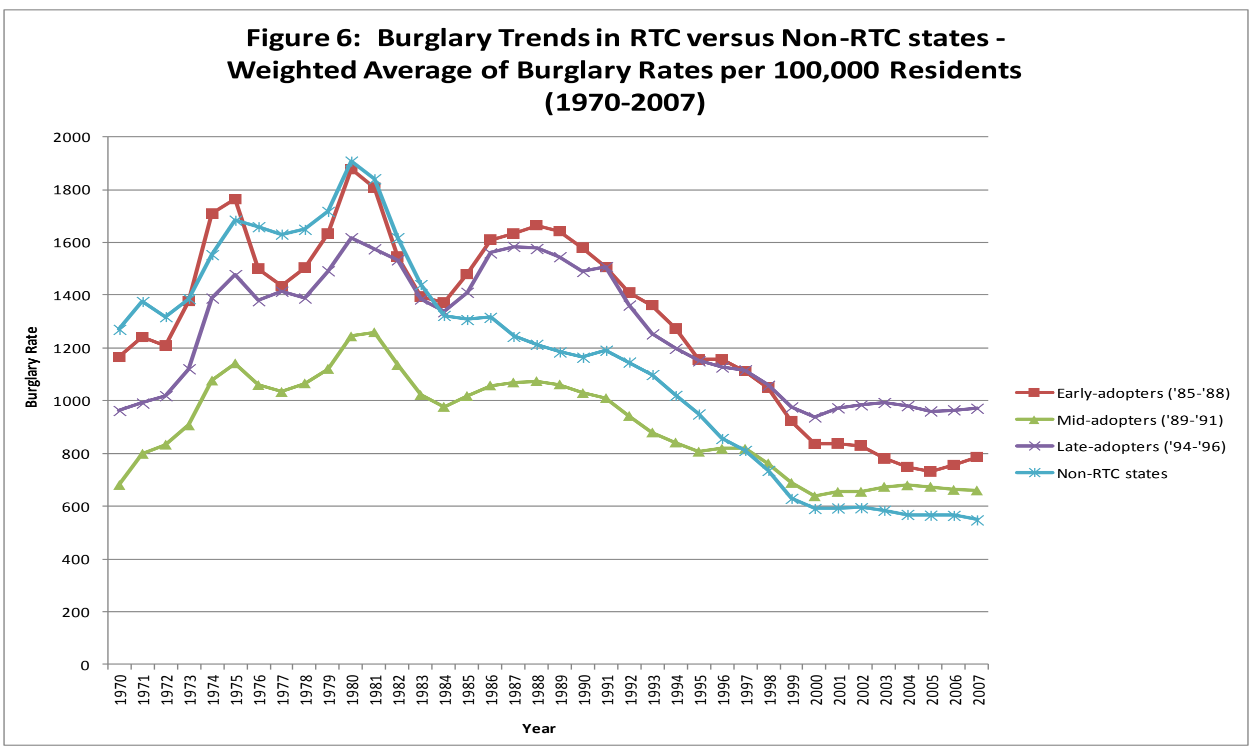
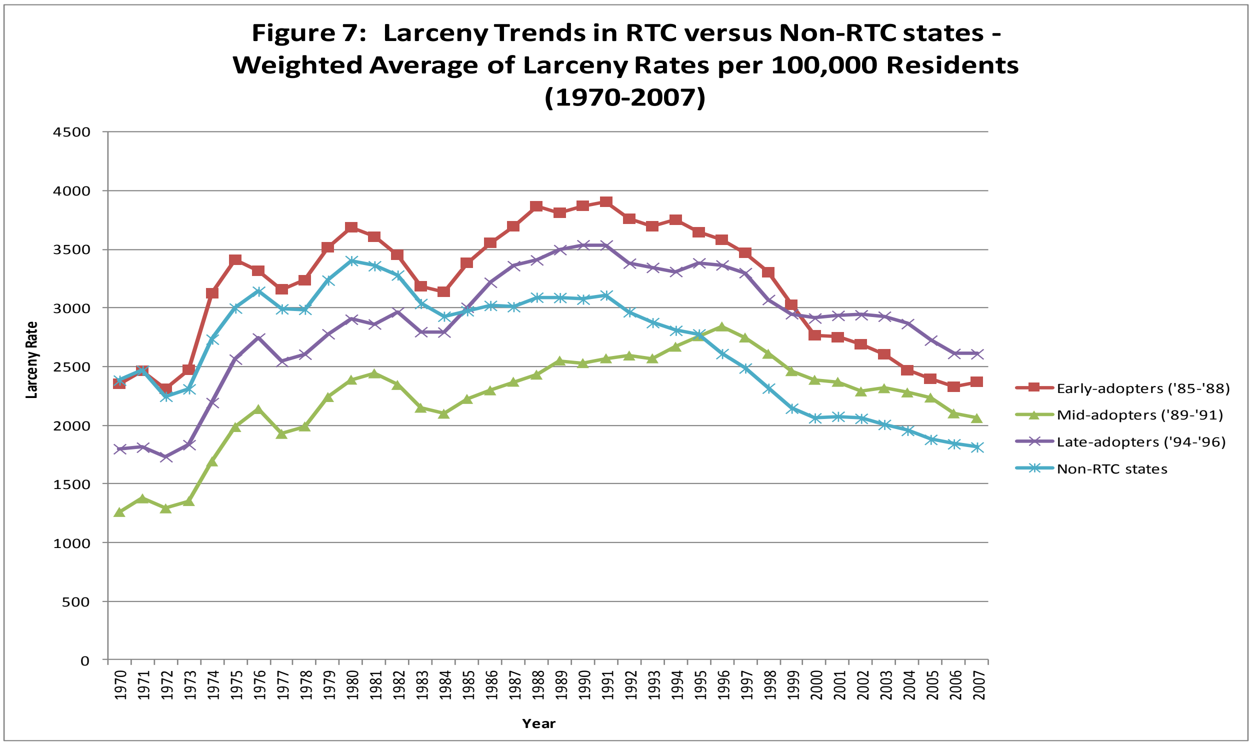


Figure 7:



Ayres and Donohue (2003a) also recommended the use of a more general model, referred to as the "hybrid model," which essentially combined the dummy variable and spline models, to measure the immediate *and* long-run impact of RTC laws on crime. Since the hybrid model nests both the dummy and spline models, one can estimate the hybrid and generate either of the other models as a special case (depending on what the data show). This exercise seemed to weaken the MGLC claim. Their analysis of the county data set from 1977-1997 using the Lott-Mustard specification (revised to measure state-specific effects) indicated that RTC laws across all states *raised* total crime costs by as much as $524 million.

Just as Lott had identified a potential problem with the dummy model (it might understate a true effect if crime followed either a V-shaped or inverted V-shaped pattern), there is a potential problem with models (such as the spline and the hybrid models) that estimate a post-passage linear trend. Early adopters of RTC laws have a far more pronounced impact on the trend estimates of RTC laws than later adopters, since there may only be a few years of post-passage data available for a state that adopts RTC laws close to the end of the data period. If those early adopters were unrepresentative of low crime states, then the final years of the spline estimate would suggest a dramatic drop in crime, not because crime had in fact fallen in adopting states, but because the more representative states had dropped out of the estimate (since there would be no post-passage data after, say, three years for a state that had adopted the RTC law only three years earlier, but there would be such data for Maine, Indiana, and North Dakota, which were the earliest RTC adopters). We recognize that each model has limitations, and present the results of all three in our tables below.[5]

III. Findings of the National Research Council

The sharply conflicting academic assessments of RTC laws specifically and the impact of firearms more generally, not to mention the heightened political salience of gun issues, prompted the National Research Council to impanel a committee of experts to critically review the entire range of research on the relationships between guns and violence. The blue-chip committee, which included prominent scholars such as sociologist Charles Wellford (the committee chair), political scientist James Q. Wilson, and economists Joel Horowitz, Joel Waldfogel, and Steven Levitt, issued its wide ranging report in 2004.

While the members of the panel agreed on the major issues discussed in eight of the nine chapters of the NRC report, the single chapter devoted to exploring the causal effects of RTC

[5]We note that in the latest version of his book, Lott (2010) criticizes the hybrid model, but he fails to appreciate that the problem with the hybrid model –and with the spline model he prefers—is that they both yield estimates that are inappropriately tilted down as the more representative states drop out of the later years, which drive the post-passage trend estimates. An apples to apples comparison that included the identical states to estimate the post-passage trend would not suggest a negative slope. This is clear in Figure 1 and Table 1 of Ayres and Donohue (2003a).

laws on crime proved to be quite contentious. After reviewing the existing (and conflicting) literature and undertaking their own evaluation of Lott's county-level crime data, 17 of the 18 committee members concluded that the data provided no reliable and robust support for the Lott-Mustard contention. In fact, they believed the data could not support any policy-relevant conclusion. In addition, they claimed they could not estimate the true impact of these laws on crime because: (1) the empirical results were imprecise and highly sensitive to changes in model specification, and (2) the estimates were not robust when the data period was extended eight years beyond the original analysis (through 2000), a period during which a large number of states adopted the law.

A. The NRC Presents Two Sets of Estimates of the Impact of RTC Laws

One can get an inkling of the NRC majority's concern about model sensitivity by examining Table 1 below, which reports estimates from the NRC report on the impact of RTC laws on seven crimes. The Table 1b estimates are based on the Lott and Mustard (1997) dummy and spline models using county data for the period 1997-2000 with the full set of Lott and Mustard controls. The Table 1a estimates use the same data but provide a more sparse specification that drops the Lott and Mustard controls and provides estimates with no covariates other than year and county fixed effects. The vastly different results produced by these different models gave the majority considerable pause. For example, if one believed the dummy model in Table 1b, then RTC laws considerably *increased* aggravated assault and robbery, while the spline model in Table 1b suggested RTC laws *decreased* the rate of both of these crimes. Noting that the RTC impact estimates disagreed across their two models (dummy and spline) for six of the seven crime categories, the NRC report concluded that there was no reliable scientific support for the more guns, less crime thesis.

Table 1

Table 1a[6]

Estimated Impact of RTC Laws – Published NRC Estimates – No Controls, All Crimes, County Data, 1977-2000

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-1.95	17.91***	12.34***	19.99***	23.33***	19.06***	22.58***
	(1.48)	(1.39)	(0.90)	(1.21)	(0.85)	(0.61)	(0.59)
Spline Model:	0.12	-2.17***	-0.65***	-0.88***	0.57***	-1.99***	-0.71***
	(0.32)	(0.30)	(0.20)	(0.26)	(0.19)	(0.13)	(0.13)

Table 1b[7]

Estimated Impact of RTC Laws – Published NRC Estimates – Lott-Mustard Controls, All Crimes, County Data 1977-2000

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-8.33***	-0.16	3.05***	3.59***	12.74***	6.19***	12.40***
	(1.05)	(0.83)	(0.80)	(0.90)	(0.78)	(0.57)	(0.55)
Spline Model:	-2.03***	-2.81***	-1.92***	-2.58***	-0.49**	-2.13***	-0.73***
	(0.26)	(0.20)	(0.20)	(0.22)	(0.19)	(0.14)	(0.13)

Interestingly, the conflicting estimates of Table 1 also led to substantial intra-panel dissention, with two members of the Committee writing separately from the NRC's majority evaluation of RTC laws. One sought to refute the majority's skepticism, and one sought to reinforce it. Noted political scientist James Q. Wilson offered the lone dissent to the Committee's report, claiming that Lott and Mustard's "More Guns, Less Crime" finding actually held up under the panel's reanalysis. Specifically, Wilson rejected the majority's interpretation of the

[6] Estimations include year and county fixed effects, and are weighted by county population. Standard errors are in parentheses below estimations. Robust standard errors are not used in the published NRC estimates. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. Throughout this paper, the standard errors appear just below the corresponding parameter estimate.

[7] Estimations include year and county fixed effects, and are weighted by county population. Standard errors are provided beneath point estimates in parentheses. Robust standard errors are not used in the published NRC estimates. The control variables (adopted from the Lott-Mustard model) include: arrest rate, county population, population density, per capita income measures, and 36 demographic composition measures indicating the percentage of the population belonging to a race-age-gender group. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

regression estimates seen in Table 1. Although the majority saw sharp conflicts in the Table 1b results between the dummy and spline models, Wilson was impressed that for one of the seven crimes -- murder -- the dummy and spline models of Table 1b generated estimates that seemingly suggested there were statistically significant drops in crime associated with RTC laws. This agreement in the Table 1b murder estimates led him to heartily endorse the "More Guns, Less Crime" view. Indeed, after dismissing papers that had cast doubt on the MGLC hypothesis (such as Black and Nagin, 1998), on the grounds that they were “controversial,” Wilson concluded: “I find the evidence presented by Lott and his supporters suggests that RTC laws do in fact help drive down the murder rate, though their effect on other crimes is ambiguous” (NRC Report, p. 271.).

The Committee penned a response to Wilson’s dissent (separate from its overall evaluation of RTC legislation), which stressed that the only disagreement between the majority and Wilson (throughout the entire volume on gun issues) concerned the impact of RTC laws on murder. They noted that, while there were a number of negative estimates for murder using the Lott-Mustard approach, there were also several positive estimates that could not be overlooked. In addition, even the results for murder failed to support the MGLC contention when restricting the period of analysis to five years or less after law adoption.[8] The important task was to try to reconcile these contradictions—and the panel majority believed that was not possible using the existing data.

Committee member (and noted econometrician) Joel Horowitz was the ardent skeptic, and not without merit. Horowitz joined the refutation of Wilson but also authored his own

[8] The importance of this restriction on the post-passage data was mentioned earlier: as states dropped out of the post-passage data, the estimated impact of RTC laws became badly biased (since one was no longer deriving the estimated effect from a uniform set of states).

appendix discussing at length the difficulties of measuring the impact of RTC laws on crime using observational rather than experimental data.[9] He began by addressing a number of flaws in the panel-data approach. First, if factors other than the adoption of the RTC law change but are not controlled for in the model, then the resulting estimates would not effectively isolate the impact of the law (we demonstrate the likelihood of this possibility in Section X below). Second, if crime increases before the adoption of the law at the same rate it decreases after adoption, then a measured zero-difference would be misleading. The same problem arises for multiyear averages. Third, the adoption of RTC laws may be a *response* to crime waves. If such an endogeneity issue exists, the difference in crime rates may merely reflect these crime waves rather than the effect of the laws. Lastly, as even Lott (2000) found in his data, RTC states differ noticeably from non-RTC states (e.g., RTC states are mainly Republican and had low but rising rates of crime). It would not be surprising if these distinctive attributes influence the measured effect of RTC laws. In this event, looking at the impact of RTC laws in current RTC states may not be useful for predicting the likely result if these laws were adopted in very different states.

Ideally, states would be randomly selected to adopt RTC laws, thereby eliminating the systematic differences between RTC states and non-RTC states. In the absence of such randomization, researchers introduce controls to try to account for these differences, which generates debate over which set of controls is appropriate. Lott (2000) defended his model by claiming that it included "the most comprehensive set of control variables yet used in a study of crime" (p. 153). But Horowitz was unimpressed by Lott's claim, noting that it is possible to control for too many variables – or too few. He pointed out that Donohue (2003) found a significant relationship between crime and *future* adoption of RTC legislation, suggesting the

[9] While his chapter is directed at the analysis of RTC laws, Horowitz's comments applied to an array of empirical studies of policy that were discussed throughout the entire NRC volume.

likelihood of omitted variable bias and/or the endogenous adoption of the laws. Horowitz concluded by noting that there is no test that can determine the right set of controls: “it is not possible to carry out an empirical test of whether a proposed set of *X* variables is the correct one…it is largely a matter of opinion which set [of controls] to use” (NRC Report, p. 307). Noting the likelihood of misspecification in the evaluation of RTC laws, and that estimates obtained from a misspecified model can be highly misleading, he concluded that there was little hope of reaching a scientifically supported conclusion based on the Lott-Mustard/NRC model.[10]

B. The Serious Need for Reassessment

The story thus far has been discouraging for those hoping for illumination of the impact of legislation through econometric analysis. If the NRC majority is right, then years of observational work by numerous researchers, topped off with a multi-year assessment of the data by a panel of top scholars, were not enough to pin down the actual impact of RTC laws. If Horowitz is right, then the entire effort to estimate the impact of state right-to-carry policies from observational data is doomed. Indeed, there may be simply too much that researchers do not know about the proper structure of econometric models of crime. Notably, however, the majority did not join Horowitz in the broad condemnation of all observational microeconometrics for the study of this topic. Perhaps a model that better accounts for all relevant, exogenous, crime-influencing factors and secular crime trends could properly discern the effects of RTC laws – whether supporting or refuting the Wilson conclusion that RTC laws reduce murder. On the other hand, an examination of additional models might only serve to strengthen the NRC majority conclusion that the models generated estimates that were too

[10] Note that this nihilistic conclusion was very close to that found by a more recent NRC report investigating the deterrent effect of the death penalty. Daniel S. Nagin and John V. Pepper, editors, Deterrence and the Death Penalty (2012). This recent NRC report reviewed 30 years of studies on this deterrence question and found the entire literature to be "uninformative."

variable to provide clear insight into the effect of RTC laws on crime.

IV. Panel Data Estimates in the NRC Report

Previous research on guns and crime has shown how data and methodological flaws can produce inaccurate conclusions. In a follow-up to their initial 2003 *Stanford Law Review* paper, Ayres and Donohue (2003b) demonstrated how coding errors can yield inaccurate and misleading estimates of the effect of RTC laws on crime. Commenting on a study in support of the MGLC premise by Florenz Plassman and John Whitley (2003), Ayres and Donohue (2003b) described numerous coding flaws. After correcting these errors, the evidence supporting the "More Guns, Less Crime" hypothesis evaporated.

A. The NRC's Panel-Data Models

Since the NRC panel based their reported estimates on data provided by John Lott, we thought it prudent to carefully examine the NRC committee's own estimates. With the help of the NRC committee members who provided the NRC 1977-2000 county data set, we were ultimately able to generate the NRC panel data estimates.[11] Once we fully understood the way in which these NRC estimates were generated (shown in Table 1 above), it became clear that the NRC report presented estimates that essentially had three flaws: 1) the specification (used by Lott and Mustard) was problematic in a number of dimensions; 2) the standard errors were incorrect in two ways, both of which made the results appear more significant than they were; and 3) there were some errors in the data, which had been supplied by Lott.

[11] The initial published version of this article -- Aneja, Donohue, and Zhang (2011) -- noted that we had originally failed to replicate the NRC results, with our efforts complicated because the Committee had misplaced the do files that generated the NRC estimates. After publication, we were informed of the precise specification the NRC had employed, which did generate the published NRC estimates (although these estimates are flawed in the manner described in the text).

Given the NRC majority conclusion that the Lott and Mustard thesis was not supported by the data, it was a reasonable choice to simply take the Lott and Mustard data and specifications and adhere to their method of computing standard errors. In essence, the NRC majority was shrewdly saying, "Even if we fully accept everything that Lott and Mustard have argued for, we still find no support for their conclusion." The only problem with the NRC majority approach, though, was that presenting the estimates in Table 1b above opened the door for James Q. Wilson to argue that some support for RTC laws could be gleaned from the ostensibly conflicting evidence.

Wilson's claim, once again, was that Table 1b spoke with clarity, albeit on only one point. He conceded that the Lott and Mustard dummy and spline estimates conflicted for six of the seven crime categories, but since they both showed statistically significant reductions in murder, Wilson claimed that the murder finding was robust and he concluded that RTC laws save lives. The NRC majority responded that Table 1a did not similarly suggest that RTC laws reduced murder but Wilson swatted that response aside by saying that a model with no covariates would not be as persuasive as the Table 1b models with covariates. The NRC majority could have countered Wilson's claim far more effectively if they had simply shown that the Lott and Mustard model was highly assailable and greatly underestimated its standard errors. Indeed, nothing would have been left standing for Wilson to construct a positive story of RTC laws if the NRC majority had simply calculated the correct standard errors for the Table 1b models, since doing so would have eliminated any claim that the RTC laws generated a statistically significant reduction in murder or any other crime.

B. Problems with the Lott and Mustard Models and Data Published in the NRC Report

Our goal in this section is to improve on the estimates presented in the NRC report (Table

1 above) by correcting what we consider to be clear errors in the Lott and Mustard specification, data, and standard errors. Thus, we began by constructing our own county-level data set, which we will refer to as the "updated 2013 data set." We create the same variables found in Lott's data—crime rates, demographic composition, arrest rates, income, population, and population density—and extend our new set to 2006 (the NRC data ended in 2000).[12] This data extension will also provide us an opportunity to explore how the NRC's results are affected when using more current data. As we will see in Section VII, the additional years of data will also enable us to estimate the effect of six additional state adoptions of RTC laws not present in the NRC analysis: Michigan (2001), Colorado (2003), Minnesota (2003), Missouri (2003), New Mexico (2003), and Ohio (2004).[13]

We obtained our county crime data from the University of Michigan's Interuniversity Consortium for Political and Social Research, which maintains the most comprehensive collection of UCR data. Unfortunately, county-level crime data for 1993 is currently unavailable. The National Archive of Criminal Justice Data recently discovered an error in the crime data imputation procedure for 1993 and for this reason, has made 1993 data inaccessible until the error has been corrected. Thus, for all of the following tables with estimates using our updated county data, we are missing values for 1993.

In Table 2, we will replicate and extend the Table 1 NRC estimates correcting for three errors: 1) some data errors that were transmitted to the NRC when they used the Lott county data set; 2) a clear specification error in the arrest rate controls; and 3) the failure to use both robust and clustered standard errors.

---

[12] We also add 0.1 to *all zero* crime values before taking the natural log in our county-level data set, as the NRC did.

[13] Kansas and Nebraska adopted RTC laws in 2006, which is too late to be captured in our analysis, since we assume a state to be an "RTC state" beginning in the first *full* year after a law's passage.

1. The Lott Data Errors Used in the NRC Estimates

In our original efforts at trying to replicate the NRC estimates derived from their Lott data set, we discovered a number of small errors in that data set.[14] First, Philadelphia's year of adoption is coded incorrectly—as 1989 instead of 1995. Second, Idaho's year of adoption is coded incorrectly—as 1991 instead of 1990. Third, the area variable, which is used to compute county density, has missing data for years 1999 and 2000. Fourth, we determined that the NRC data set was missing all county identifiers for 1999 and 2000, which meant that that both these years were dropped for the NRC estimates depicted in Table 1. Our analysis corrects all these errors.

2. Lott and Mustard's Erroneous Arrest Rate Variables

Since the NRC report followed the Lott-Mustard specification, the regressions it presented (which we reproduce in Table 1) used arrest rates as the sole criminal justice control variable in estimating the effect of RTC laws. Although we have already noted Lott's claim that his is "the most comprehensive set of control variables yet used in a study of crime," in fact, the Lott and Mustard model omits controls for police and incarceration, which many studies -- e.g., Kovandzic, Vieraitis, and Boots, (2009) -- have found to be key influences on crime (we will re-introduce those variables in Section VIII).

Lott and Mustard's use of the arrest rate variables is not a good modeling choice in general, and the particular approach that Lott and Mustard employed is especially problematic.[15]

[14] We know all too well how easy it is to make these small but annoying errors in creating these data sets, since regrettably we had a few similar errors in our own data set in the Aneja, Donohue, Zhang (2011) published version, which are all corrected here. None of the main conclusions of the published paper were altered by those errors, which are set forth in footnote 18.

[15] Even apart from the considerable data problems with the county arrest rates, the measure is also not well defined. Ideally, one might like a measure showing the likelihood that one who commits a certain crime will be arrested. The Lott and Mustard arrest rates instead are a ratio of arrests to crimes, which means that when one person kills many,

To see the concern, note that the NRC's model (Table 1b in this paper) is trying to explain the level of seven individual Index I crime categories while using a control that is computed as a crime-specific arrest rate, which is the number of arrests for a given crime divided by the contemporaneous number of crimes. Thus, murder in 1990 is "explained" by the ratio of arrest to murders in 1990. Econometrically, it is inappropriate to use this contemporaneous measure since it leaves the dependent variable on both sides of the regression equation (at a minimum, a better approach would lag this variable one year, as discussed in Ayres and Donohue (2009)). Better still, one could alternatively use the broad categories of violent and property crimes to compute arrest rates, as have many recent papers (such as, Moody and Marvell, 2008). We adopt this latter approach for all of our regressions in this paper, and also lag the arrest rate one year to reduce the endogeneity problem.

3. The Erroneous Standard Errors in the NRC Estimates

Surprisingly, when the NRC presented its estimates (which we reproduce in Table 1), the NRC report did not make the very basic adjustment to their standard errors to correct for heteroskedacticity. Since Hal White's paper discussing this correction has been the single most cited paper in all of economics since 1970,[16] the failure to make this standard adjustment was unexpected. Accordingly, in all of our own estimates, we use robust standard errors.

Even more significant in terms of the results, though, is the issue of whether one must cluster the standard errors. The statistical consequence of the NRC committee's failure to use

---

for example, the arrest rate falls, but when many people kill one person, the arrest rate rises since only one can be arrested in the first instance and many can in the second. The bottom line is that this "arrest rate" is not a probability and is frequently greater than one because of the multiple arrests per crime. For an extended discussion on the abundant problems with this pseudo arrest rate, see Donohue and Wolfers (2009).

[16] Kim, E.H.; Morse, A.; Zingales, L. (2006). "What Has Mattered to Economics since 1970?". Journal of Economic Perspectives 20 (4): 189–202.

robust and clustered standard errors is to massively understate the reported standard errors (and consequently to overstate the level of significance). Unlike the issue of robust standard errors, the Committee report actually addressed the issue of clustering, concluding that this adjustment was not necessary. In Section V, we will show that this was an error. Therefore, we will from this time forward only present results based on the clustering adjustment to our standard errors.

C. Improving on the Table 1 Estimates by Using Better Data and Slightly Improved Lott and Mustard Models

Having just identified three problems with the estimates presented by the NRC, we now seek to fix them. To be clear about our approach, we use annual county-level crime data (and later, state-level data) for the United States from 1977 through either 2000 (to conform to the NRC report) or 2006. We explore the impact of RTC laws on seven Index I crime categories by estimating the reduced-form regression:

$$Y_{it} = \eta RTC_{jt} + \alpha_i + \theta_t + \beta_{jt} + \gamma X_{ijt} + \varepsilon_{it} \qquad (1)$$

where the dependent variable Y_{it} denotes the natural log of the individual violent and property crime rates for county *i* and year *t*. Our explanatory variable of interest—the presence of an RTC law within state *j* in year *t*—is represented by RTC_{jt}. The exact form of this variable shifts according to the three variations of the model we employ (these include the Lott and Mustard dummy and spline models, as well as the Ayres and Donohue hybrid model.)[17]

The variable α_i indicates county-level fixed effects (unobserved county traits) and θ_t indicates year effects. As we will discuss below, there is no consensus on the use of state-specific time trends in this analysis, and the NRC report did not address this issue. Nevertheless,

[17] As noted previously, in the dummy variable approach, the RTC variable is a dichotomous indicator that takes on a value of one in the first full year that a state j has an RTC law. In the spline model, the RTC variable indicates the number of post-passage years. The hybrid specification contains both dummy and trend variables.

we will explore this possibility, with β_{jt} indicating state-specific trends, which are introduced in selected models. Since neither Lott and Mustard (1997) nor the NRC (2004) examines state trends, this term is dropped when we estimate their models. The term X_{ijt} represents a matrix of observable county and state characteristics thought by researchers to influence criminal behavior. The components of this term, however, vary substantially across the literature. For example, while Lott uses only "arrest rates" as a measure of criminal deterrence, we discuss the potential need for other measures of deterrence, such as incarceration levels or police presence, which are measured at the state level.

Table 2 reproduces the regressions depicted in Table 1, while correcting for the three problems mentioned above (the inaccurate Lott data, the poorly constructed Lott arrest ratios, and the incorrect standard errors) and using our reconstruction of the county dataset from 1977 through 2000 (which omits the flawed 1993 county data). Tables 2a and 2b represent our improved estimates of what the NRC reported and we depict in Tables 1a and 1b. Table 2b appends our hybrid model, which estimates the effect of RTC laws with both a dummy and a spline component (thus nesting the individual dummy and spline models).

The bottom line is that the superior Table 2 estimates look nothing like the Table 1 estimates presented in the NRC report. Table 1 shows estimated effects that are almost uniformly statistically significant -- at times suggesting crime increases and at times suggesting crime decreases. Table 2 shows far fewer statistically significant effects, but every one of which suggests RTC laws *increase* crime -- for aggravated assault, robbery, auto theft, burglary, and larceny. There is not even a hint of any crime declines.

Recall that James Q. Wilson thought that the most important regressions to look at were those presented in Table 1b, because they provided the full set of controls from the Lott and

Mustard specification. While for six of the seven crime categories the story that emerged from Table 1b varied sharply on whether one looked at the dummy or the spline model, Wilson was content to find a beneficial RTC effect on murder because the Table 1 estimates for murder both appeared to be negative and significant.

Table 2

Table 2a

Estimated Impact of RTC Laws – with ADZ Changes – No Controls, All Crimes, 1977-2000

Dataset: ADZ Updated 2013 County Data (without 1993 data)

Changes: Updated Dataset, Robust and Clustered Standard Errors

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-0.16	32.23	24.04	24.94*	31.45	31.92	38.55
	(7.92)	(22.87)	(18.78)	(14.35)	(20.11)	(21.30)	(24.90)
Spline Model:	0.69	4.67*	4.34*	3.20	4.72*	5.08*	6.03*
	(0.96)	(2.70)	(2.22)	(2.02)	(2.62)	(2.61)	(3.15)

Table 2b

Estimated Impact of RTC Laws – with ADZ Changes – Lott-Mustard Controls, All Crimes, 1977-2000

Dataset: ADZ Updated 2013 County Data (without 1993 data)

Changes: Updated Dataset, Lagged Violent/Property Arrest Rates, Robust and Clustered Standard Errors

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-1.61	15.64	18.52***	10.31*	18.99**	11.18*	21.77***
	(6.96)	(10.57)	(6.48)	(6.05)	(7.18)	(5.64)	(6.43)
Spline Model:	-0.09	1.36	2.07**	0.74	2.08**	0.99	2.13*
	(0.88)	(1.39)	(0.98)	(0.91)	(0.85)	(0.78)	(1.10)
Hybrid Post-Passage Dummy:	-1.77	14.33	14.21**	10.36*	14.85**	10.11*	18.55***
	(7.52)	(9.11)	(5.80)	(6.03)	(7.30)	(5.61)	(4.83)
Trend Effect:	0.04	0.32	1.04	-0.01	1.00	0.26	0.78
	(0.94)	(1.23)	(0.96)	(0.82)	(0.68)	(0.76)	(0.91)

When we switch to Table 2b, however, we see that there is nothing resembling a statistically significant impact of RTC laws on murder. In fact, we see that assault, auto theft, and larceny now have estimates that are simultaneously statistically significant and positive for both the dummy and spline model. Thus, the results that Professor Wilson found to be consistent evidence of RTC laws reducing murder (see Table 1b) disappear with better data and a superior specification.[18]

In fact, this was essentially the message of the NRC report. Small changes made the estimates bounce around so much that it was difficult to reach any conclusion about the true causal impact of RTC laws. Perhaps it might have been helpful to Wilson if the majority had gone one step further and presented something like the alternative results from Table 2. As we will see in the ensuing sections, there are many additional avenues that could have been explored to probe the robustness of the Table 1b findings that Wilson had accepted so unquestioningly. We will explore these factors in subsequent sections: Section VI will explore whether one should control for individual state trends in crime, section VII will look at additional years of data (adding data beyond 2000 to 2006), section VIII will alter the Lott and Mustard

[18] In the process of reviewing our previous published models and data from ADZ (2011), we discovered some errors in the two data sets that we had constructed (the so-called updated 2009 county data and updated 2009 state data), which are corrected in this paper. For the county data set, we miscoded the state trend variable for Arkansas. Second, we incorrectly coded Oregon's year of adoption as 1989 instead of 1990. Third, Kansas counties have been incorrectly coded as belonging to Kentucky for years 1997-2006. Fourth, our spline and hybrid models had included a counter variable to capture the effect of a post-passage trend, but they inadvertently omitted the overall trend variable off of which this post-passage trend was to be estimated.

In addition to these errors that we discovered, Moody, Lott, Marvell, and Zimmerman (2012) identified two other errors: observations for county 2060 for Alaska was duplicated 73 times for 1996 and Kansas' year of adoption was coded incorrectly as 1996 instead of 2006. All of these errors have been corrected in the tables prepared for this paper. Additionally, the real per capita income measures from our previous datasets had been calculated incorrectly, and these changes have been made for real per capita income, income maintenance, unemployment insurance, and retirement payments.

Moody, Lott et al also claimed that Florida's year of adoption was coded incorrectly as 1989 instead of 1987, and South Dakota's year of adoption was coded incorrectly as 1986 instead of 1985. We disagree on these two points. First, our county data does not provide crime category information for Florida counties for 1988 (this is evident in the NRC data set as well), so we elected to drop the observations for this year for all Florida counties. Thus, it may seem that our first year of adoption is erroneously coded as 1989. However, this simply reflects the fact that we have not included observations for 1988. Note that we maintained consistency with our other trend variables by beginning the post-passage variable counter with a value of "2" in year 1989 to demonstrate 2 years since the passage of RTC legislation. The second issue – regarding South Dakota's year of adoption as 1986 instead of 1985 – is disputed. Our research indicates that the law was passed in 1986 (and thus has been appropriately coded with 1987 as the first full year of implementation).

For our state data set that we will employ below, we note the following changes from ADZ (2011): North Dakota should show RTC adoption in year 1985 with a post-passage trend variable beginning in 1986; South Dakota should show RTC adoption in year 1986 with a post-passage trend variable beginning in 1987. The state dataset has been re-constructed with the most recently available data, the sources of which are provided with this paper at http://works.bepress.com/john_donohue/.

specification (beyond the already mentioned correction for the contemporaneous, crime-specific arrest rates), section IX will go beyond the county data to look at state data, and Section X will consider some additional problems of potential omitted variable bias. But a key aspect of the Table 2 results is that the standard errors were adjusted using the cluster command, and this is one area where the majority stumbled in concluding that this adjustment was not needed. Section V will now address the clustering question.

V. Debate over the Clustering of Standard Errors

A. Is Clustering Necessary?

Aside from the neglecting to use heteroskedastic-robust standard errors, the NRC committee also did not use a cluster adjustment. Research has found that the issue of whether to "cluster" the standard errors has a profound impact on assessments of statistical significance. This issue gained prominence beginning primarily with a 1990 paper by Brent Moulton. Moulton (1990) pointed to the possible need for the clustering of observations when treatments are assigned at a group-level. In such cases, there is an additive source of variation that is the same for all observations in the group, and ignoring this unique variation leads to standard errors that are underestimated. Lott, however, suggests that clustered standard errors are not needed (Lott 2004), claiming that county-level fixed effects implicitly control for state-level effects, and therefore, clustering the standard errors by state is unnecessary.

The NRC committee (2004) sided with Lott on this point, stating that "there is no need for adjustments for state-level clustering." (p. 138). However, we *strongly* believe the committee was mistaken in this decision. One must account for the possibility that county-level disturbances may be correlated within a state during a particular year by clustering the standard errors by state. There is also a second reason for clustering that the NRC report did not address. Specifically, serial correlation in panel data can lead to major underestimation of standard errors. Indeed, Bertrand, Duflo, and Mullainathan (2004) point out that even the Moulton correction alone may be insufficient for panel-data estimators that utilize more than two periods of data due to autocorrelation in both the intervention variable and the outcome variable of interest. Wooldridge (2003, 2006), as well as Angrist and Pischke (2009), suggest that clustering the standard errors by state (along with heteroskedasiticity-robust standard errors) will help address this problem, and at least provide a lower bound on the standard errors.

B. Using Placebo Laws to Test the Impact of Clustering

Our Table 2 estimates (which include clustering) reveal that this adjustment makes a major difference in the results generated by the Lott and Mustard models that the NRC report adopted in its analysis -- completely wiping out any sign of statistically significant crime reductions attributable to RTC laws. But who is correct on the clustering issue—Lott, Mustard, and the NRC panel on the one hand, or Angrist, Pischke, and several other high-end applied econometricians on the other? To address this important question we run a series of placebo tests. In essence, we randomly assign RTC laws to states, and re-estimate our model iteratively (1000 times), recording the number of times that the variable(s) of interest are "statistically significant" at the 5% level. For this experiment, we use our most flexible model: the hybrid model (that incorporates both a dummy and a trend variable) with the controls employed by the NRC.

We run five versions of this test. In our first test, we generate a placebo law in a random year for all 50 states and the District of Columbia. Once the law is applied, it persists for the rest of our data period, which is how laws are coded in the original analysis. We run 1000 trials (where each trial consists of a randomly generated set of RTC passage years) and then proceed to take a simple average of the percentage of significant dummy variable and spline variable estimates. In our second test, we apply a placebo law in a random year to the 32 states that actually implemented right-to-carry laws during the period we are analyzing. The remaining 19 states assume no RTC law.[19] Here again we run 1000 trials in which each iteration consists of randomly generated RTC passage years and proceed to take a simple average of the percentage of significant estimates. Third, we randomly select 32 states to receive a placebo law in a random year (to ensure that any random sample of 32 states does not have the potential to inaccurately bias results, we repeat this entire procedure 5 times – that is, we take 5 samples of 32 random states and for each sample, run the aforementioned process of assigning a random year of RTC adoption 1000 times). Then, we take a simple average of the number of statistically significant dummy variable and spline estimates. Thus, we are, in effect, counting the number of significant dummy and trend estimates generated from 5000 hybrid regressions. Fourth, we apply a placebo law in a random year to the 19 states which did not pass RTC laws within the period, dropping the other 32 states from our dataset, and take the simple average of the statistically significant dummy variable and spline estimates. Finally, we randomly select 12 of the 19 states (to correspond to the previous randomly generated 32 states) to receive an RTC in a randomized year of adoption, and iterate this process 1,000 times over five separate samples. The results of these five tests are presented in Table 3.

Given the random assignment, one would expect to reject the null hypothesis of no effect of these randomized "laws" roughly 5 percent of the time if the standard errors in our regressions

[19] For the purposes of this analysis we do not consider Nebraska or Kansas to have passed an RTC law during this period. These states passed RTC laws in 2006, however, we assume a state to be an "RTC state" beginning in the first *full* year after a law's passage.

are estimated correctly. Instead, the table reveals that the null hypothesis is rejected 21-69 percent of the time for murder and robbery with the dummy variable and even more frequently with the trend variable (35-71 percent). Clearly, this exercise suggests that the standard errors used in the NRC report are far too small.

Table 3b replicates the exercise of Table 3a, but now uses the cluster correction for standard errors (on state). Table 3b suggests that clustering standard errors does not excessively reduce significance, as the NRC panel feared. In fact, the percentages of "significant" estimates produced in all three versions of the test still lie well beyond the 5% threshold. Similar results are found when we replicate Tables 3a and 3b while employing the dummy model instead of the hybrid model (we do not show those results here). All of these tests show that if we do *not* cluster the standard errors, the likelihood of obtaining significant estimates is astonishingly (and unreasonably) high. The conclusion we draw from this exercise is that clustering is clearly needed to adjust the standard errors in these panel-data regressions. Accordingly, we use this clustering adjustment for all remaining regressions in this paper.

Table 3[20]

Table 3a

Percentage of Significant Estimates (5% Level) – Lott-Mustard Controls, 1977-2006 – **No Clustered Standard Errors**
Dataset: ADZ Updated 2013 County Data (without 1993 data)
Hybrid Model

All figures reported in %		Dummy Variable	Trend Variable
1. All 50 States + DC:	Murder	51.6	63.1
	Robbery	55.9	64.5
2. Exact 32 States:	Murder	65.8	71.4
	Robbery	69.3	71.1
3. Random 32 States:	Murder	52.9	66.6
	Robbery	56.2	65.1
4. All 19 States:	Murder	20.9	35.4
	Robbery	38.6	46.9
5. Random 12 States:	Murder	23.9	48.4
	Robbery	39.6	46.3

Table 3b

Percentage of Significant Estimates (at the 5% Level) – Lott-Mustard Controls, 1977-2006 – **With Clustered Standard Errors**
Dataset: ADZ Updated 2013 County Data (without 1993 data)
Hybrid Model

All figures reported in %		Dummy Variable	Trend Variable
1. All 50 States + DC:	Murder	10.6	9.2
	Robbery	8.7	9.1
2. Exact 32 States:	Murder	10.9	11.4
	Robbery	10.5	9.5
3. Random 32 States:	Murder	9.4	11.9
	Robbery	9.0	9.3
4. All 19 States	Murder	12.7	14.3
	Robbery	12.8	14.1
5. Random 12 States:	Murder	15.6	22.6
	Robbery	13.9	14.5

[20] Simulation based on NRC with-controls model, which similar to above estimations, includes year fixed effects, county fixed effects, and weighting by county population. The control variables (adopted from the Lott-Mustard model) include: lagged arrest rate, county population, population density, per capita income measures, and 36 demographic composition measures indicating the percentage of the population belonging to a race-age-gender group. All ten tests use robust standard errors.

VI. Debate over the Inclusion of Linear Trends

An important issue that the NRC did not address was whether there was any need to control for state-specific linear trends. Inclusion of state trends could be important if, for example, a clear pattern in crime rates existed before a state adopted an RTC law that continued into the post-passage period. On the other hand, there is also a potential danger in using state-specific trends if their inclusion inappropriately extrapolates a temporary swing in crime long into the future or otherwise mar the estimate of the dynamic effect of the policy shock (Wolfers 2006). Lott and Mustard (1997) never controlled for state-specific trends in analyzing handgun laws, while Moody and Marvel (2008) always controlled for these trends. Ayres and Donohue (2003a) presented evidence with and without such trends.

Table 4 replicates the NRC's full model (with the appropriate clustering adjustment) from Table 2b with one change: here we add a linear state trend to this county-data model. Strikingly, Table 4 suggests that RTC laws increase aggravated assault by roughly 3-5 percent each year, but no other statistically significant effect is observed. Thus, the addition of state trends eliminates the potentially problematic result of RTC laws increasing property crimes, which actually increases our confidence in these results. Certainly an increase in gun carrying and prevalence induced by a RTC law could well be thought to spur more aggravated assaults. Nonetheless, one must at least consider whether the solitary finding of statistical significance is merely the product of running seven different models, is a spurious effect flowing from a bad model, or reflects some other anomaly (such as changes in the police treatment of domestic violence cases, which could confound the aggravated assault results).[21]

Table 4[22]

Estimated Impact of RTC Laws – Lott-Mustard Controls, 1977-2000 – Clustered Errors and State Trends
Dataset: ADZ Updated 2013 County Data (without 1993 data)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-2.44	-8.41	11.34	0.03	5.26	0.34	6.16
	(5.78)	(11.36)	(7.11)	(7.24)	(8.45)	(7.46)	(7.51)
Spline Model:	-0.19	-3.47	4.41**	1.29	1.64	0.47	1.64
	(1.55)	(4.85)	(1.90)	(2.51)	(2.04)	(2.04)	(2.14)
Hybrid Post-Passage Dummy:	-2.40	-4.72	6.70	-1.59	3.62	-0.22	4.59
	(5.20)	(7.12)	(6.38)	(6.82)	(7.93)	(6.95)	(6.76)
Trend Effect:	-0.03	-3.17	3.99**	1.39	1.41	0.48	1.35
	(1.45)	(4.55)	(1.78)	(2.51)	(1.93)	(1.97)	(2.03)

VII. Extending the Data Through 2006

Thus far we have presented panel-data regression results for the period 1977-2000. Since more data are now available, we can further test the strength of the MGLC premise over time by estimating the NRC Lott and Mustard covariates specification on data extended through 2006. Table 5a presents our estimates (with clustering), which can be compared with Table 2b (which

[21] We tested this theory by creating a new right-hand side dummy variable that identified if a state passed legislation requiring law enforcement officials to submit official reports of all investigated domestic violence cases. Eight states have passed this legislation of which we are aware: Florida (1984), Illinois (1986), Louisiana (1985), New Jersey (1991), North Dakota (1989), Oklahoma (1986), Tennessee (1995), and Washington (1979). We included this dummy variable when running both the NRC specification (through 2000) and our preferred specification (through 2006), and found that this dummy indicator of domestic violence reporting statutes did not undermine the finding that RTC laws *increase* aggravated assaults.

[22] Estimations include year and county fixed effects, and are weighted by county population. Robust standard errors are provided beneath point estimates in parentheses. The control variables (adopted from the Lott-Mustard model) include: lagged arrest rate, county population, population density, per capita income measures, and 36 demographic composition measures indicating the percentage of the population belonging to a race-age-gender group. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

also clusters the standard errors in the main NRC model, but is estimated on the shorter time period). This comparison reveals that the additional six years of data do not substantially change the picture that emerged in Table 2b showing that RTC laws *increase* aggravated assault, auto theft, burglary, and larceny (although the results showing an increase in aggravated assault are stronger with the additional years of data).

Table 5b simply adds state trends to the Table 5a model, which can then be compared to Table 4 (clustering, state trends, and 1977-2000 data). Collectively, these results suggest that the added six years of data do not appreciably change the results from the shorter period. The inclusion of state trends on the longer data set suggests that RTC laws *increase* aggravated assault by roughly 9-10 percent.

Table 5[23]

Table 5a

Estimated Impact of RTC Laws – Lott-Mustard Controls, 1977-2006 – Clustered Standard Errors
Dataset: ADZ Updated 2013 County Data (without 1993 data)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-3.42	13.57	16.33***	6.65	16.82**	10.70**	16.80***
	(6.18)	(13.68)	(5.12)	(4.95)	(7.36)	(4.78)	(5.35)
Spline Model:	-0.20	1.03	1.25*	0.47	1.03	0.44	1.14**
	(0.61)	(1.22)	(0.73)	(0.55)	(0.65)	(0.48)	(0.55)
Hybrid Post-Passage Dummy:	-3.20	11.29	13.53***	5.72	15.48**	11.16**	14.78***
	(6.26)	(11.69)	(3.78)	(5.61)	(6.79)	(5.31)	(5.39)
Trend Effect:	-0.05	0.49	0.60	0.20	0.29	-0.10	0.43
	(0.61)	(0.94)	(0.69)	(0.62)	(0.52)	(0.53)	(0.56)

Table 5b

Estimated Impact of RTC Laws – Lott-Mustard Controls, 1977-2006 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 County Data (without 1993 data)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-1.48	-10.47	9.91**	4.38	8.25	3.75	6.83
	(4.79)	(10.47)	(4.57)	(5.79)	(6.45)	(5.81)	(6.29)
Spline Model:	-0.43	-5.26	1.73	-0.53	-0.88	-1.30	-1.43
	(1.05)	(4.48)	(1.50)	(1.85)	(1.82)	(1.58)	(1.69)
Hybrid Post-Passage Dummy:	-1.29	-8.10	9.19**	4.67	8.74	4.40	7.57
	(4.83)	(9.09)	(4.31)	(5.59)	(6.03)	(5.48)	(5.85)
Trend Effect:	-0.40	-5.09	1.54	-0.62	-1.06	-1.39	-1.59
	(1.06)	(4.39)	(1.47)	(1.84)	(1.76)	(1.54)	(1.64)

[23] Estimations include year and county fixed effects, and are weighted by county population. Robust standard errors are provided beneath point estimates in parentheses. The control variables (adopted from the Lott-Mustard model) include: lagged arrest rate, county population, population density, per capita income measures, and 36 demographic composition measures indicating the percentage of the population belonging to a race-age-gender group. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

VIII. Revising the Lott-Mustard Specification

We have already suggested that the Lott and Mustard specification that the NRC employed is not particularly appealing along a number of dimensions. The most obvious problem – omitted variable bias has already been alluded to: the Lott and Mustard (1997) model had no control for incarceration, which Wilson considered to be one of the most important influences on crime in the last 20 years. In addition to a number of important omitted variables, the Lott-Mustard model adopted by the NRC includes a number of questionable variables, such as the highly dubious ratio of arrests to murders, and the 36 (highly collinear) demographic controls.

To explore whether these specification problems are influencing the regression estimates, we revise the NRC models in a number of ways. First, we completely drop Lott and Mustard's flawed contemporaneous arrest rate variable and add in two preferable measures of state law enforcement/deterrence: the incarceration rate and the rate of police.[24] Second, we add two additional controls to capture economic conditions: the unemployment rate and the poverty rate, which are also state-level variables. Finally, mindful of Horowitz's admonition that the Lott-Mustard model might have *too many* variables (including demographic controls that are arguably irrelevant to the relationship between the guns and crime, and may have a spurious, misleading effect), we decided not to follow the NRC in using the 36 demographic controls employed by Lott-Mustard. Instead, we adhered to the more customary practice in the econometrics of crime and controlled only for the demographic groups considered to be most involved with criminality

[24] We also estimated the model with the arrest rate (lagged by one year to avoid endogeneity concerns), and the results were qualitatively similar to Table 6a except that the dummy variable estimates for Rape (10%), Assault (1%), Robbery (5%), Auto (5%), Burglary (1%), and Larceny (1%) are now all significant. For Table 6b, the dummy variable estimates for murder shifts from negative to positive (insignificant) and assault, robbery, auto theft, burglary, and larceny all switch from negative to positive (though only assault is significant and only at the 10% level).

(as offenders and victims), namely the percentage of black and white males between ages 10 and 40 in each county.[25]

[25] To test the robustness of this specification to changes in the demographic controls, we also estimated the following models: only black males between ages 10 and 40; only black males between ages 10 and 30; and black and white males between ages 10 and 30. The results were again qualitatively similar across our tests.

Table 6[26]

Table 6a

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2006 – Clustered Standard Errors
Dataset: ADZ Updated 2013 County Data (without 1993 data)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	0.40	25.34	23.46	21.24	26.69	29.78	32.17
	(7.10)	(18.71)	(18.46)	(14.18)	(20.78)	(21.86)	(25.30)
Spline Model:	0.36	3.00	3.42*	2.66*	3.18	3.82	4.40
	(0.84)	(1.84)	(1.99)	(1.56)	(2.28)	(2.41)	(2.75)
Hybrid Post-Passage Dummy:	-2.00	13.57	7.98	10.35	14.18	13.84	13.02
	(7.01)	(14.86)	(15.53)	(11.90)	(16.41)	(17.83)	(19.89)
Trend Effect:	0.46	2.28	3.00	2.11	2.42	3.09	3.71
	(0.88)	(1.47)	(1.82)	(1.43)	(2.00)	(2.19)	(2.43)

Table 6b

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2006 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 County Data (without 1993 data)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-3.54	-18.09	-2.00	-0.05	-0.69	-6.51	-6.37
	(4.86)	(11.60)	(9.71)	(7.37)	(9.86)	(10.92)	(12.35)
Spline Model:	0.00	-7.11	1.27	-0.26	-0.49	-1.57	-1.83
	(1.30)	(5.16)	(2.56)	(2.36)	(2.81)	(2.94)	(3.26)
Hybrid Post-Passage Dummy:	-3.58	-15.03	-2.60	0.07	-0.47	-5.86	-5.60
	(5.11)	(10.54)	(10.22)	(7.72)	(10.33)	(11.49)	(12.90)
Trend Effect:	0.07	-6.81	1.32	-0.27	-0.48	-1.45	-1.72
	(1.34)	(5.12)	(2.64)	(2.41)	(2.89)	(3.05)	(3.37)

[26] Estimations include year and county fixed effects and are weighted by county population. Robust standard errors are provided beneath point estimates in parentheses. The control variables for this "preferred" specification include: incarceration and police rates (lagged one year to avoid potential endogeneity issues), unemployment rate, poverty rate, population density, per capita income measures, and six demographic composition measures. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

The results with this new specification are presented in Tables 6a-6b (which correspond to Tables 5a-5b estimated using the Lott and Mustard specification). Note that had the NRC panel used our preferred specification while maintaining its view that neither clustering nor controls for state trends are needed, we would have overwhelming evidence that RTC laws *increase* crime.[27] We don't show these regression results since we are convinced that clustering is needed, although of course when we cluster in Table 6a, the point estimates remain the same (while significance is drastically reduced). Table 6b shows that this model is far less sensitive to whether we control for state trends. Essentially, our preferred specification shows no statistically significant crime effects (but the large standard errors reflect a considerable degree of uncertainty).

IX. State versus County Crime Data

In their initial study, Lott and Mustard (1997) tested the "More Guns, Less Crime" hypothesis by relying primarily on county-level data from the FBI's *Uniform Crime Reports* (UCR).[28] These FBI reports present yearly estimates of crime based on monthly crime data from local and state law enforcement agencies across the country. The NRC report followed Lott and Mustard in this choice and presented regression estimates using only county data. Unfortunately, according to criminal justice researcher Michael Maltz, the FBI's county-level data is highly problematic.

---

[27] Re-estimating Table 6a without clustering (no state trends) shows all dummy variable point estimates (except murder) positive and significant at the 1% level. The murder dummy variable is positive, but not significant. For the spline model, all spline estimates (except murder) are positive and significant at the 1% level, whereas murder is positive and significant at the 5% level.

[28] Lott and Mustard present results based on state-level data, but they strongly endorse their county-level over their state-level analysis: "the very different results between state- and county-level data should make us very cautious in aggregating crime data and would imply that the data should remain as disaggregated as possible" (Lott and Mustard, 1997, p. 39).

The major problem with county data stems from the fact that law enforcement agencies voluntarily submit crime data to the FBI. As a result, the FBI has little control over the accuracy, consistency, timeliness, and completeness of the data it uses to compile the UCR reports. In a study published in the *Journal of Quantitative Criminology*, Maltz and Targonski (2002) carefully analyzed the shortcomings in the UCR data set and concluded that UCR county-level data is unacceptable for evaluating the impact of RTC laws. For example, in Connecticut, Indiana, and Mississippi, over 50% of the county-level data points are missing crime data for more than 30% of their populations (Maltz and Targonski 2002). In another thirteen states, more than 20% of the data points have gaps of similar magnitude. Based on their analysis, Maltz and Targonski (2002) concluded that:

> "County-level crime data cannot be used with any degree of confidence…The crime rates of a great many counties have been underestimated, due to the exclusion of large fractions of their populations from contributing to the crime counts. Moreover, counties in those states with the most coverage gaps have laws permitting the carrying of concealed weapons. How these shortcomings can be compensated for is still an open question…it is clear, however, that in their current condition, county-level UCR crime statistics cannot be used for evaluating the effects of changes in policy" (pp. 316-317).

Because of the concerns raised about county-level crime data, it is prudent to test our models on state-level data. According to Maltz and Targonski (2003), state-level crime data are less problematic than county-level data because the FBI's state-level crime files take into account missing data by imputing all missing agency data. County-level files provided by NACJD, however, impute missing data only if an agency provides at least six months of data; otherwise, the agency is dropped completely (Maltz 2006). As with our estimations using county-level data, we compiled our state-level data from scratch, and will refer to it as "Updated 2013 State-level Data."

Unsurprisingly, the regression results reproduced using state-level data are again different from the NRC committee's estimates using county-level data. This is shown in Table 7a, which presents the results from the NRC's specification (the Lott-Mustard model) on state data, with the cluster adjustment.[29] Table 7b simply adds state trends. When we compare these state-level estimates to the county-level estimates (using the updated 2013 county-level data set), we see that there are marked differences. Considering the preceding discussion on the reliability—or lack thereof—of county data, this result is unsurprising. Importantly, state-level data through 2006 show not a hint of statistically significant evidence that RTC laws reduce murder.[30] Looking across the models with and without state linear trends, there is evidence of increases in aggravated assault and larceny, and decreases in auto theft and rape.

[29] Our placebo test on county data showed that standard errors needed to be adjusted by clustering. In Appendix A, we again find that clustering is needed for state data. Thus, all our state-level estimates include clustering.

[30] We also estimate the model on data through 2000 (the last year in the NRC report), though those results are not shown here. The results similarly do show not any statistically significant evidence that RTC laws reduce murder. Moreover, we also estimate the NRC's no-controls model through 2006 on the state-level data. See Appendix B for these results.

Table 7[31]

Table 7a

Estimated Impact of RTC Laws – Lott-Mustard Controls, 1977-2006 – Clustered Standard Errors
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-2.79	-2.88	2.68	-1.62	2.91	-2.87	3.30**
	(3.68)	(2.11)	(4.45)	(3.60)	(4.59)	(2.09)	(1.49)
Spline Model:	0.44	-0.38	1.05	0.32	-0.44	-0.34	0.43
	(0.49)	(0.44)	(0.73)	(0.58)	(0.45)	(0.44)	(0.28)
Hybrid Post-Passage Dummy:	-5.01	-1.83	-1.07	-3.15	5.06	-2.03	2.16
	(3.81)	(2.36)	(3.57)	(3.85)	(4.58)	(2.07)	(1.63)
Trend Effect:	0.65	-0.31	1.10	0.45	-0.64	-0.25	0.34
	(0.49)	(0.48)	(0.71)	(0.61)	(0.41)	(0.45)	(0.31)

Table 7b

Estimated Impact of RTC Laws – Lott-Mustard Controls, 1977-2006 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-1.67	-3.43	-1.77	-2.41	1.65	-1.91	0.79
	(2.90)	(2.09)	(2.72)	(3.05)	(2.72)	(1.30)	(1.06)
Spline Model:	0.55	0.40	2.48***	0.84	-1.60**	0.54	0.48
	(0.60)	(0.85)	(0.59)	(0.82)	(0.68)	(0.64)	(0.48)
Hybrid Post-Passage Dummy:	-2.11	-3.82*	-3.58	-3.08	2.86	-2.35	0.47
	(3.01)	(2.20)	(2.86)	(3.10)	(2.76)	(1.43)	(1.22)
Trend Effect:	0.63	0.56	2.63***	0.97	-1.72**	0.64	0.46
	(0.63)	(0.88)	(0.61)	(0.83)	(0.69)	(0.67)	(0.51)

Tables 8a and 8b below repeat Tables 7a and 7b, but use our preferred set of explanatory variables instead of the Lott and Mustard (1997) variables. The main question raised by these

[31] Estimations include year and state fixed effects, and are weighted by state population. Robust standard errors are provided beneath point estimates in parentheses. The control variables (adopted from the Lott-Mustard model) include: lagged arrest rate, state population, population density, per capita income measures, and 36 demographic composition measures indicating the percentage of the population belonging to a race-age-gender group. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

estimations is whether state trends are needed in the regression models. If not, there is some evidence that RTC laws increase rape, assault, robbery, auto theft, burglary, and larceny. If state trends are needed, some muddiness returns but RTC laws appear to increase aggravated assault and perhaps auto theft.

Table 8[32]

Table 8a

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2006 – Clustered Standard Errors
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	0.92	10.71**	8.56**	14.18**	16.84**	11.24**	11.66***
	(5.43)	(5.04)	(4.00)	(6.69)	(7.47)	(5.38)	(4.19)
Spline Model:	0.16	0.82	1.37**	1.31	1.20	0.81	1.05*
	(0.78)	(0.69)	(0.68)	(0.96)	(0.85)	(0.69)	(0.58)
Hybrid Post-Passage Dummy:	0.25	9.13**	3.13	10.76*	14.84**	9.87**	9.03**
	(4.29)	(4.23)	(3.37)	(6.07)	(6.20)	(4.36)	(3.57)
Trend Effect:	0.15	0.35	1.21*	0.76	0.45	0.31	0.58
	(0.76)	(0.63)	(0.71)	(0.97)	(0.70)	(0.61)	(0.56)

Table 8b

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2006 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-0.27	-2.85	-1.30	2.29	8.41**	1.02	1.64
	(2.70)	(2.19)	(3.00)	(3.39)	(3.95)	(2.04)	(1.87)
Spline Model:	0.89	0.34	2.93***	0.61	-1.35	-0.01	-0.14
	(0.79)	(0.84)	(0.89)	(1.13)	(0.88)	(0.84)	(0.67)
Hybrid Post-Passage Dummy:	-0.67	-3.04	-2.65	2.04	9.13**	1.04	1.72
	(2.79)	(2.23)	(2.94)	(3.55)	(3.99)	(2.17)	(2.00)
Trend Effect:	0.91	0.43	3.01***	0.56	-1.61*	-0.04	-0.19
	(0.80)	(0.83)	(0.90)	(1.17)	(0.86)	(0.87)	(0.70)

[32] Estimations include year and state fixed effects, and are weighted by state population. Robust standard errors are provided beneath point estimates in parentheses. The control variables for this "preferred" specification include: incarceration and police rates (lagged one year to avoid potential endogeneity issues), unemployment rate, poverty rate, population density, per capita income measures, and six demographic composition measures. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 9[33]

Table 9a

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2010 – Clustered Standard Errors
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	2.03	10.84**	8.57**	13.14*	16.44**	11.48**	10.82***
	(5.87)	(4.79)	(3.89)	(6.83)	(7.80)	(5.16)	(4.03)
Spline Model:	0.45	0.73	1.13*	1.19	1.05	0.76	0.92*
	(0.63)	(0.61)	(0.61)	(0.80)	(0.71)	(0.56)	(0.48)
Hybrid Post-Passage Dummy:	-0.01	9.08**	4.19	9.34	14.09**	9.70**	8.05**
	(4.95)	(4.22)	(3.45)	(6.33)	(6.86)	(4.64)	(3.63)
Trend Effect:	0.45	0.39	0.98	0.85	0.53	0.40	0.62
	(0.59)	(0.58)	(0.64)	(0.80)	(0.61)	(0.53)	(0.48)

Table 9b

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2010 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-1.79	-2.79	-0.32	1.53	7.89*	0.51	0.93
	(3.81)	(2.17)	(3.05)	(2.90)	(4.13)	(2.04)	(1.72)
Spline Model:	0.87	-0.09	2.01**	-0.06	-1.55**	-0.09	-0.56
	(0.71)	(0.67)	(0.77)	(0.82)	(0.77)	(0.61)	(0.51)
Hybrid Post-Passage Dummy:	-2.29	-2.78	-1.41	1.58	8.84**	0.57	1.24
	(3.81)	(2.19)	(2.95)	(3.01)	(4.09)	(2.17)	(1.83)
Trend Effect:	0.93	-0.02	2.05***	-0.10	-1.78**	-0.11	-0.59
	(0.70)	(0.67)	(0.76)	(0.84)	(0.75)	(0.63)	(0.54)

[33] These regressions include year and state fixed effects, and are weighted by state population. Robust standard errors are provided beneath point estimates in parentheses. The control variables for this "preferred" specification include: incarceration and police rates (lagged one year to avoid potential endogeneity issues), unemployment rate, poverty rate, population density, per capita income measures, and six demographic composition measures.
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 9 extends the data through 2010. Extending the analysis to 2010 has a clear impact on Table 8a, which is the comparable table using ADZ variables to the no-state trends Lott and Mustard model that was presented in the NRC report. Here the ostensible evidence that RTC laws increase crime is very strong: all three models in Table 9a have positive coefficients for every crime category (with the exception of Hybrid Post-Passage Dummy on murder, which is -0.01), and 11 of the 28 coefficients are statistically significant.

In contrast, little changes when the four years of data are added for the model that controls for state trends: Table 9b once again shows highly significant evidence (in the spline model and in the trend effect of the hybrid model) that RTC laws increase aggravated assault. Similarly, both dummy effects are positive and significant at a 10% level or greater for auto theft, while both trend effects are negative and significant at a 5%. Most of the remaining coefficients are not statistically significant.

Table 10 restricts the analysis period to 1999-2010 in order to completely avoid the run up and run down from the crack epidemic. When restricting the analysis to this date range, and without controlling for state trends, there appears to be little effect of RTC laws on crime, except for a statistically significant increase in murder using the trend models. However, if state trends do need to be controlled for, the results become more varied. The passage of RTC laws is associated with declines in rape, auto theft, and larceny.

Table 10[34]

Table 10a

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1999-2010 – Clustered Standard Errors
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	5.85	1.71	4.07	-3.54	-1.48	1.19	-3.41
	(5.47)	(3.11)	(3.38)	(4.48)	(3.66)	(3.23)	(2.60)
Spline Model:	1.41**	0.32	1.05	0.12	-0.60	0.56	0.17
	(0.57)	(0.42)	(0.67)	(0.44)	(0.73)	(0.39)	(0.33)
Hybrid Post-Passage Dummy:	5.51	1.63	3.82	-3.57	-1.33	1.05	-3.46
	(5.78)	(3.11)	(3.48)	(4.55)	(3.66)	(3.37)	(2.64)
Trend Effect:	1.38**	0.31	1.03	0.13	-0.60	0.56	0.19
	(0.54)	(0.42)	(0.67)	(0.44)	(0.73)	(0.40)	(0.33)

Table 10b

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1999-2010 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	2.77	2.53	3.74	-0.56	-0.88	0.56	-1.31
	(3.43)	(2.46)	(2.93)	(5.40)	(3.51)	(3.11)	(2.02)
Spline Model:	0.31	-2.80**	-2.09	-1.55	-5.14*	-1.09	-1.53*
	(2.76)	(1.26)	(1.39)	(1.88)	(2.82)	(1.47)	(0.85)
Hybrid Post-Passage Dummy:	2.73	3.07	4.15	-0.28	0.06	0.76	-1.03
	(3.33)	(2.29)	(2.94)	(5.67)	(3.88)	(3.38)	(2.13)
Trend Effect:	0.23	-2.89**	-2.22	-1.55	-5.14*	-1.11	-1.50*
	(2.77)	(1.31)	(1.39)	(1.95)	(2.88)	(1.54)	(0.88)

[34] These regressions include year and state fixed effects, and are weighted by state population. Robust standard errors are provided beneath point estimates in parentheses. The control variables for this "preferred" specification include: incarceration and police rates (lagged one year to avoid potential endogeneity issues), unemployment rate, poverty rate, population density, per capita income measures, and six demographic composition measures. The states that adopted shall issue laws during the time period are Colorado (2004), Kansas (2006), Michigan (2002), Minnesota (2004), Missouri (2002), Nebraska (2006), New Mexico (2004), and Ohio (2005).
* Significant at 10%; ** Significant at 5%; *** Significant at 1%.

X. Additional Concerns in the Evaluation of Legislation Using Observational Data

We now turn to three critical issues that must be considered when using panel data to evaluate the impact of legislation and public policy (and gun laws in particular). First, we discuss the possibility of difficult-to-measure omitted variables and how such variables can shape estimates of policy impact. We are particularly concerned with how the crack epidemic of the 1980s and 1990s may bias results in the direction of finding a beneficial effect. Second, we explore pre-adoption crime trends in an attempt to examine the potentially endogenous adoption of right-to-carry legislation. Finally, given that the intent of right-to-carry legislation is to increase gun-carrying in law-adopting states, we explore whether these laws may have had a particular effect on gun-related assaults (which is the one crime category that has generated somewhat consistent results thus far).

A. Further Thoughts on Omitted Variable Bias

As discussed above, we believe it is likely that the NRC's estimates of the effects of RTC legislation are marred by omitted variable bias. In our attempt to improve (at least to a degree) on the original Lott-Mustard model, we included additional explanatory factors, such as the incarceration and police rates, and removed extraneous variables (such as unnecessary and collinear demographic measures). We recognize, however, that there are additional criminogenic influences for which we cannot fully control. In particular, we suspect that a major shortcoming of all of the models presented is the inability to account for the possible influence of the crack-cocaine epidemic on crime.[35]

[35] Although Lott and Mustard (1997) do attempt to control for the potential influence of crack-cocaine through the use of cocaine price data based on the U.S. Drug Enforcement Agency's STRIDE data, we find their approach wanting for both theoretical and empirical reasons. First, a control for crack should capture the criminogenic influence of the crack trade on crime. We know that prior to 1985, there was no such influence in any state and that after some point in the early to mid-1990s this criminogenic influence declined strongly. Since there is little reason to believe that cocaine prices would be informative on the criminogenic influence of crack in particular geographic

Many scholars now suggest that rapid growth in the market for crack cocaine in the late 1980s and the early 1990s was likely one of the major influences on increasing crime rates (and violent crimes in particular) during this period (Levitt 2004). Moreover, the harmful criminogenic effect of crack was likely more acute in urban areas of states slow to adopt RTC laws. Meanwhile, many rural states adopted such laws during this era. If this was indeed the case, this divergence between states could account for much of the purported "crime-reducing" effects attributed by Lott and Mustard to gun laws (which were then supported by scholars such as James Q. Wilson). The regression analysis would then identify a relationship between rising crime and the failure to adopt RTC legislation, when the actual reason for this trend was the influence of crack (rather than the passage of the RTC law).

We now explore how results from our main models vary when we restrict the analysis to the time periods before and after the peak of the American crack epidemic. According to Fryer *et al*. (2005), the crack problem throughout most of the country peaked at some point in the early 1990s. Coincidentally, the original Lott-Mustard period of analysis (1977-1992) contains years that likely represent the height of crack-induced crime problem. With this in mind, we run our main regressions after breaking up our dataset into two periods: the original Lott-Mustard period of analysis (1977-1992) as well as the post-Lott-Mustard period (1993-2006). We first present the results for the era that includes the crack epidemic (1977-1992) on our preferred model. We run these regressions (with clustered standard errors) on state-level data, with and without state

---

areas, it is hard to see how the cocaine price data could be a useful control. Second, the data that Lott and Mustard use is itself questionable. Horowitz (2001) argues forcefully that STRIDE data is not a reliable source of data for policy analyses of cocaine. The data are mainly records of acquisitions made to support criminal investigations in particular cities, and are not a random sample of an identifiable population. Moreover, since the STRIDE data is at the city-level, we are not sure how this would be used in a county-level analysis. The data was collected for 21 cities, while there are over 3,000 counties in the U.S. In addition, the data is missing for 1988 and 1989, which are crucial years in the rise of the crack epidemic in poor urban areas. Lott and Mustard drop those years of analysis when including cocaine prices as a control.

trends. These results are presented in Tables 11a and 11b. We then estimate the same models on the post-crack period (see Tables 12a and 12b).

Note that, with a simple naive reading, the regression results in Table 11 from the initial Lott and Mustard 16-year time period (1977-1992) do suggest that crime rates are dampened by RTC laws if state trends are not needed and that murder may have declined if state trends are needed. If we look at the following 14 year period from 1993 – 2006 in Table 12, however, there is no longer any evidence of a decline in violent crimes. Instead, RTC laws are associated with higher rates of murder, rape, aggravated assault and robbery. This evidence supports the theory that the initial Lott and Mustard finding was likely the result of the crime-raising impact of crack in non-RTC states.

Table 11[36]

Table 11a

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-1992 – Clustered Standard Errors
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-7.49*	-6.90**	-5.84	-4.54	6.13	-1.81	1.28
	(4.29)	(3.31)	(4.04)	(4.91)	(4.89)	(2.98)	(1.33)
Spline Model:	-2.30	-0.99	0.59	-2.13***	0.38	0.23	0.48*
	(1.54)	(0.71)	(1.63)	(0.64)	(0.90)	(0.66)	(0.26)
Hybrid Post-Passage Dummy:	-2.50	-6.89**	-11.63**	1.49	8.13	-3.78	0.10
	(5.73)	(2.97)	(5.76)	(6.38)	(5.68)	(3.98)	(1.72)
Trend Effect:	-1.95	-0.00	2.25	-2.35**	-0.78	0.77	0.46
	(1.92)	(0.61)	(1.48)	(0.96)	(0.83)	(0.70)	(0.33)

Table 11b

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-1992 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-6.66**	-4.59	-1.51	-3.23	-1.70	-3.12	1.62
	(3.26)	(3.44)	(3.27)	(4.62)	(3.27)	(2.52)	(1.95)
Spline Model:	-6.68***	-0.32	-0.49	-4.62*	-2.84*	-2.93**	-0.08
	(2.27)	(1.21)	(1.48)	(2.38)	(1.62)	(1.18)	(1.23)
Hybrid Post-Passage Dummy:	3.37	-6.30*	-1.30	4.42	3.16	1.19	2.62
	(5.27)	(3.53)	(3.84)	(6.84)	(4.03)	(4.23)	(2.29)
Trend Effect:	-7.54***	1.29	-0.15	-5.75*	-3.65*	-3.23*	-0.75
	(2.74)	(1.21)	(1.74)	(3.13)	(1.90)	(1.79)	(1.40)

[36] Estimations include year and state fixed effects, and are weighted by state population. Robust standard errors are provided beneath point estimates in parentheses. The control variables for this "preferred" specification include: incarceration and police rates (lagged one year to avoid potential endogeneity issues), unemployment rate, poverty rate, population density, per capita income measures, and six demographic composition measures. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 12[37]

Table 12a

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1993-2006 – Clustered Standard Errors
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	2.24	-3.02	1.23	3.11	3.18	3.75	0.29
	(3.80)	(2.76)	(3.81)	(4.64)	(4.58)	(2.56)	(2.59)
Spline Model:	1.23*	0.22	2.08***	1.93**	0.92	0.78	0.20
	(0.72)	(0.72)	(0.76)	(0.86)	(1.08)	(0.76)	(0.54)
Hybrid Post-Passage Dummy:	1.55	-3.16	0.05	2.03	2.67	3.32	0.18
	(4.00)	(2.88)	(4.23)	(4.63)	(4.30)	(2.58)	(2.60)
Trend Effect:	1.22	0.25	2.08***	1.91**	0.90	0.75	0.20
	(0.73)	(0.71)	(0.77)	(0.87)	(1.07)	(0.76)	(0.55)

Table 12b

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1993-2006 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 State Data

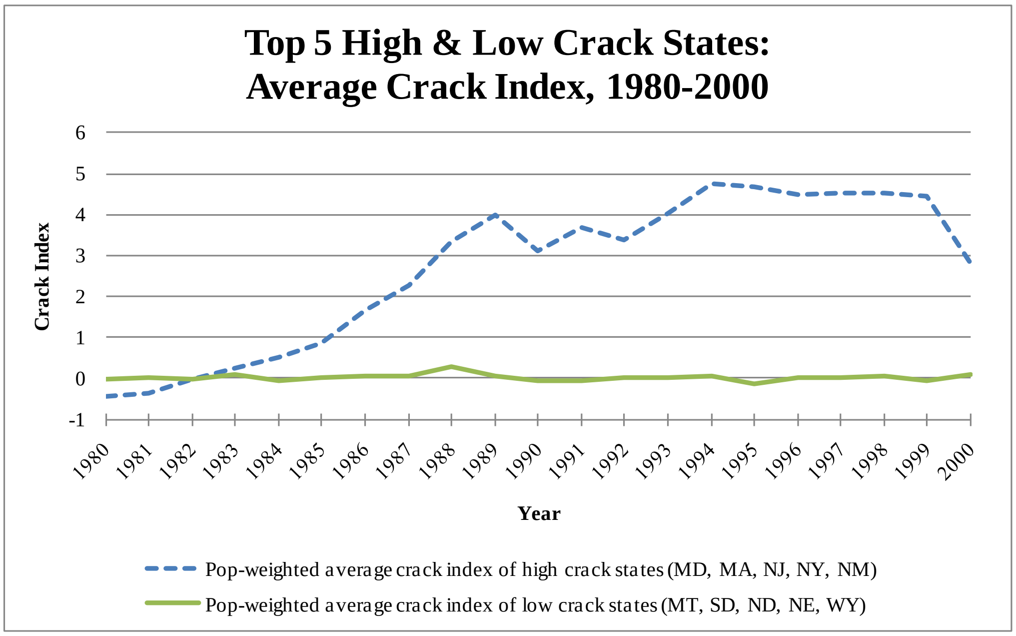
All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	3.35	0.28	2.33	3.79	3.00	0.80	0.38
	(3.34)	(2.70)	(2.69)	(2.99)	(3.40)	(1.96)	(1.79)
Spline Model:	-1.98	1.97	4.00**	0.02	-5.66*	-0.84	-1.57
	(1.74)	(1.22)	(1.54)	(1.94)	(2.97)	(1.73)	(1.38)
Hybrid Post-Passage Dummy:	4.44	-0.60	0.65	4.00	5.74	1.23	1.12
	(3.69)	(2.48)	(2.49)	(3.31)	(3.57)	(2.18)	(2.00)
Trend Effect:	-2.53	2.05*	3.92**	-0.47	-6.37*	-0.99	-1.71
	(1.88)	(1.18)	(1.61)	(2.09)	(3.22)	(1.85)	(1.52)

Figure 8 depicts a measure of crack prevalence for the period 1980-2000 in the five states with the greatest crack problem as well as the five states with the least crack, according to Fryer *et al.* (2005). Figure 9 shows the murder rates over time for these two sets of states. We see that

[37] Estimations include year and state fixed effects, and are weighted by state population. Robust standard errors are provided beneath point estimates in parentheses. The control variables for this "preferred" specification include: incarceration and police rates (lagged one year to avoid potential endogeneity issues), unemployment rate, poverty rate, population density, per capita income measures, and six demographic composition measures. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

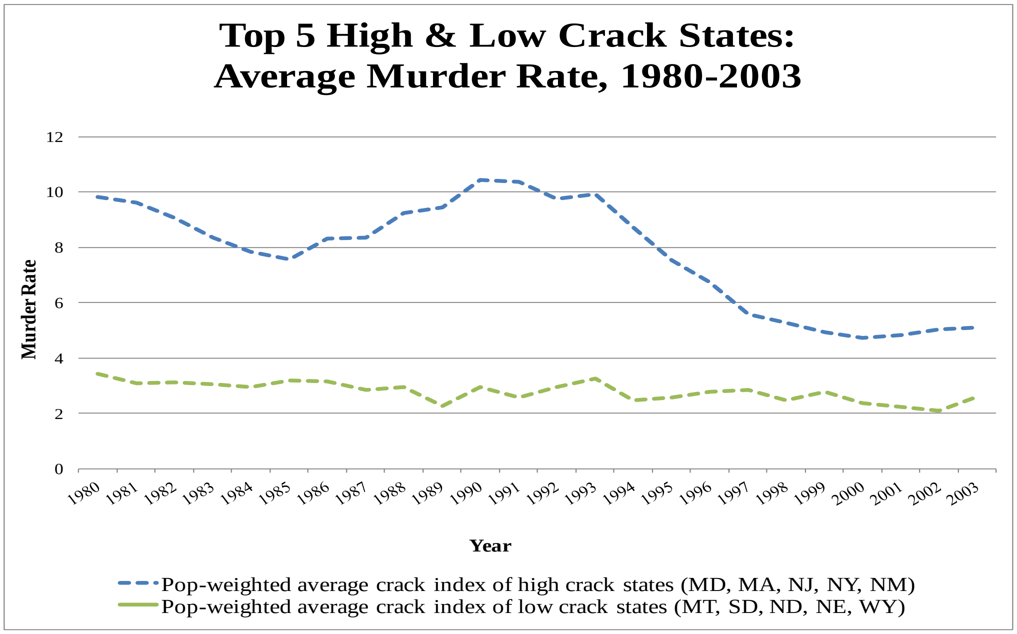
crime rose in the high crack states when the crack index rises in the mid-to-late 1980s, but that the crack index does not turn down in those states at the time crime started to fall. Apparently, the rise of the crack market triggered a great deal of violence but once the market stabilized, the same level of crack consumption could be maintained while the violence ebbed.

Figure 8: Prevalence of Crack in the 5 Most and 5 Least Crack-affected States



Source: Authors' calculations based on the crack index of Fryer et al. (2005)

Figure 9: Murder Rates in the 5 Most and 5 Least Crack-affected States



Source: FBI UCR Data.

Of course, omitting an appropriate control for the criminogenic influence of crack is problematic if the high-crack states tend not to adopt RTC laws and the low-crack states tend to adopt. This is in fact the case: all of the five "high-crack" states are non-RTC states during this period, whereas four of the five "low-crack" states are RTC states (all four adopted an RTC law by 1994).[38] The only exception is Nebraska, a state that did not adopt an RTC law until 2006, which is outside the scope of our current analysis.[39]

Moreover, as Table 13 reveals, the 12 states that adopted RTC laws during the initial Lott-Mustard period (1977-1992) had crack levels substantially below the level of the five high-crack states shown in Figures 8 and 9. Of the RTC adopters shown in Table 13, the largest has

[38] New Mexico, one of the five highest crack states, adopted its RTC law in 2003. Wyoming and Montana adopted RTC laws in 1994 and 1991, respectively. North Dakota and South Dakota adopted their laws prior to the start of our data set (pre-1977), although the dates are contested (Lott and Mustard, 1997; Moody and Marvell 2008).

[39] Nebraska's adoption of its RTC law falls outside of our analysis period, because the effect of the law is not realized until 2007, one year after passage. Out of the ten states with the lowest crack cocaine index, seven adopted an RTC law by 1994. The exceptions are Nebraska, Minnesota (2003), and Iowa (2011).

an average crack index of 1.21, while Figure 9 reveals that the high-crack states had a crack level in the neighborhood of four or five.

Table 13: Population-weighted Statistics of RTC-Adopting States between 1977 and 1990[40]

State	Year of RTC Law Adoption	Murder Rate	Crack Index
Indiana	1980	6.53	0.23
Maine	1985	2.54	0.10
North Dakota	1985	1.28	0.03
South Dakota	1986	2.08	-0.03
Florida	1987	11.88	1.00
Virginia	1988	7.96	0.85
Georgia	1989	12.39	1.21
Pennsylvania	1989	5.72	0.88
West Virginia	1989	5.65	0.43
Idaho	1990	3.58	0.36
Mississippi	1990	11.69	0.32
Oregon	1990	4.87	0.98

Source: Fryer et al. (2004), FBI UCR Data.

In other words, over the initial Lott-Mustard period of analysis (ending in 1992), the criminogenic influence of crack made RTC laws look beneficial since crack was raising crime in non-RTC states. In the later period, crime fell sharply in the high-crack states, making RTC states look bad in comparison. Therefore, the effects estimated over this entire period will necessarily water down the initial Lott-Mustard results. The hope is that estimating the effect over the entire period will wash out the impact of the omitted variable bias generated by the lack of an adequate control for the effect of crack.

[40] The crack index data comes from Fryer et al (2005), which constructs the index (beginning in 1980) based on several indirect proxies for crack use, including cocaine arrests, cocaine-related emergency room visits, cocaine-induced drug deaths, crack mentions in newspapers, and DEA drug busts. The paper does suggest that these values can be negative. The state with the lowest mean value of the crack index over the data period from 1980 to 1990 is South Dakota (-0.03), and the state with the highest mean value is New York (1.58).

As an additional test for potential omitted variable bias in both the NRC and our own preferred model specification, we perform an analysis inspired by Altonji et al (2005). In their influential paper the authors provide a practical method to test the extent to which potential omitted variable bias drives the results of a multivariate analysis. This test assumes that the selected, observable variables are chosen from a broader set of possible controls, and then explores how strong selection on unobserved variables would have to be relative to selection on observed variables to produce an OLS estimate if the true effect (in our case the effect of RTC laws on crime trends) were zero. See Appendix E for details of the test procedure.

Using the Altonji et al (2005) test procedure we analyzed the relative strength of the Table 1b estimate from the NRC Report that RTC laws were associated with an 8.33% reduction in murder rates. Using the NRC model, we calculated a potential bias of -1.03, which implies that the OLS estimate would be solely driven by selection bias if selection on unobservables were only 8 percent as strong as selection on observables. This is strong evidence that the NRC/Lott model suffers fatally from omitted variable bias. In comparison, an analogous test of our preferred specification using state data from 1977 to 2006 (Table 8a) – which showed an estimated *increase* in murder of 0.92% (albeit not statistically significant) – shows that the potential bias in the murder effect was -0.29. In other words, in our case, the implied bias is negative, which means that the positive and statistically insignificant effect of RTC laws on murder that we found is a *lower* bound for the true effect.

B. Endogeneity and Misspecification Concerns

To this point, our analysis has remained within the estimation framework common to the NRC/Lott-Mustard analyses, which implicitly assumes that passage of right-to-carry legislation in a given state is an exogenous factor influencing crime levels. Under this assumption, one can interpret the estimated coefficient as an unbiased measure of RTC laws' collective impact.

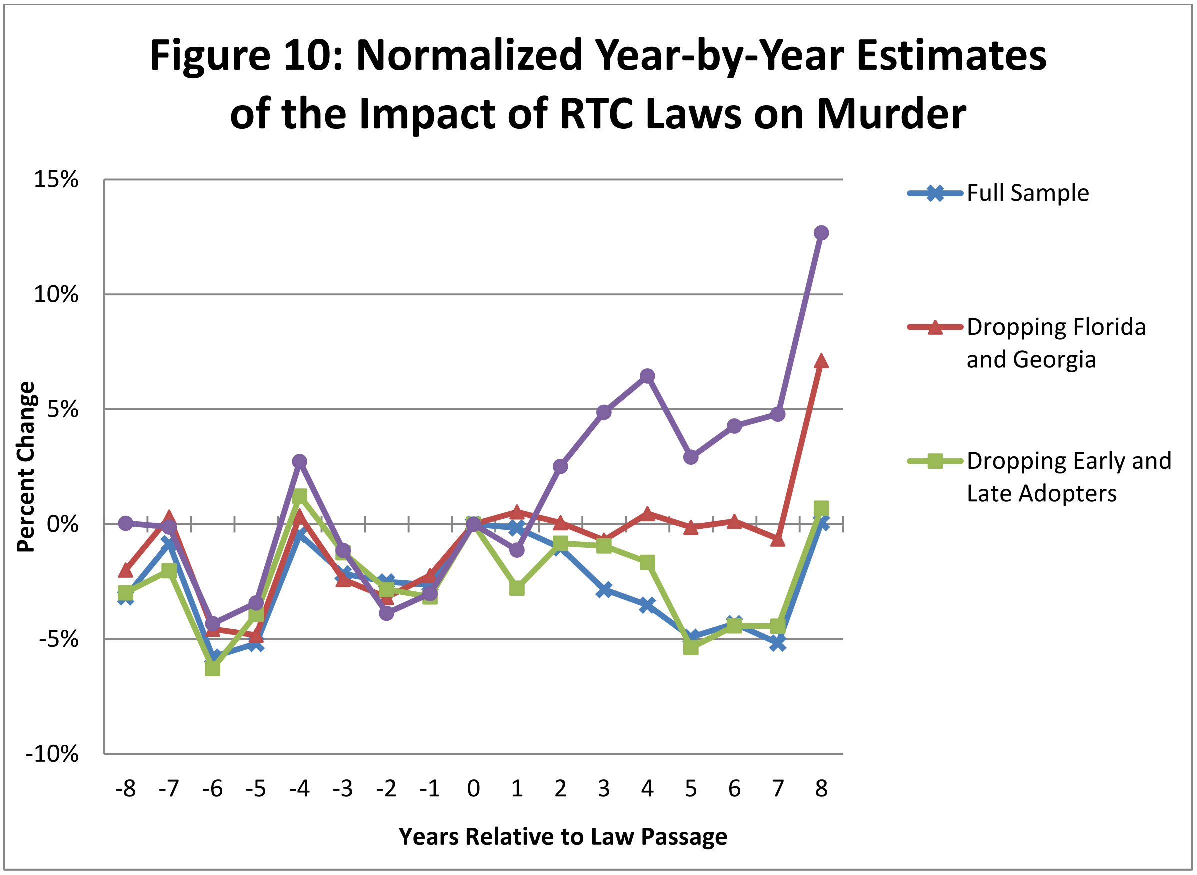
We probe the validity of this strong claim by estimating a more flexible year-by-year specification, adding pre- and post-passage dummy variables to the analysis.[41] Pre-passage dummies can allow us to assess whether crime trends shift in unexpected ways prior to the passage of a state's RTC law. Autor, Donohue, and Schwab (2006) point out that when analyzing the impact of state-level policies using panel data, one would ideally see lead dummies that are near zero. The graphs that we present below, though, suggest the possible presence of systematic differences between RTC law adopters that can complicate or thwart the endeavor of obtaining clean estimates of the impact of right-to-carry laws.

Figures 10 through 13 present the results from this exercise in graphical form. Using our preferred model as the base specification, we introduce dummies for the eight years preceding and the first eight years following adoption. We first estimate this regression for each violent crime category over the full sample of 50 states plus the District of Columbia. However, because of the presence of one state that adopted its RTC law just three years after our dataset begins, and eight states that adopted laws within the five years before our dataset ends, we have nine states that cannot enter into the full set of pre- and post-adoption dummy variables. Because Ayres and Donohue (2003) showed that the year-by-year estimates can jump wildly when states drop in or out of the individual year estimates, we also estimate the year-by-year model after dropping out the earliest (1980) and latest (post-2000) law-adopting states. In this separate series of regressions, our estimates of the full set of lead and lag variables for the 25 states that adopted

[41] In Appendix C, we further analyze the issue of misspecification and model fit by analyzing residuals from the regression analysis.

RTC laws between 1985 and 1996 are based on a trimmed data set that omits the 9 early and late adopters.[42]

Figure 10[43]



Unfortunately, the graphs raise concerns about the presence of endogenous adoption that complicate our thinking about the influence of right-to-carry laws on violent crime. If one looks at the four lines in Figure 10, one sees four different sets of year-by-year estimates of the impact

---

[42] The states that drop out (with dates of RTC law passage in parentheses) include: Indiana (1980), Michigan (2001), Colorado (2003), Minnesota (2003), Missouri (2003), New Mexico (2003), Ohio (2004), Kansas (2006), Nebraska (2006).

[43] Estimations include year and county fixed effects, state trends, and are weighted by county population. The control variables include: incarceration and police rates, unemployment rate, poverty rate, county population, population density, per capita income, and six demographic composition measures.

of RTC laws on murder. The lines have been normalized to show a zero value in the year of adoption of a RTC law. Let's begin with the bottom line (looking at the right hand side of the figure) and the line just above it. The lower line represents the naive year-by-year estimates from the preferred model estimated on the 1977-2006 period, while the line just above it drops out the early and late adopters, so that the estimated year-by-year estimates are based on the "clean" sample of all non-adopting states plus the 25 RTC adopters for which complete data is available from 8 years prior to adoption through 8 years after adoption. One sees that the trimmed estimates are different and less favorable to the "More Guns, Less Crime" hypothesis, as evidenced by the higher values in the post-passage period.

How should we interpret these trimmed sample estimates? One possibility is to conclude that on average the pre-passage estimates are reasonably close to zero and then take the post-passage figures as reasonable estimates of the true effect. If we do this, none of the estimates would be statistically significant, so one could not reject the null hypothesis of no effect. But note that the pre-passage year-to-year dummies show an oscillating pattern that is not altogether different from what we see for the post passage values. Without the odd drop when moving from year four to five and subsequent rise in values through year 8, the zero effect story would seem more compelling, but perhaps the drop merely reflects a continuation of the pre-passage oscillations, which are clearly not the product of the passage of RTC laws (since they precede passage).[44]

Perhaps, though, what is most important than the oscillating pattern is the trend just prior to passage. This might suggest that rising crime in fact increases the likelihood that a state would adopt a RTC law. In particular, since murder is typically the crime most salient in the

[44] The ostensible pronounced drop in murder five years after adoption exists for the full data set, as well, but it is part of a continuing downward trend in murder that simply reaches a trough five years after passage.

media, we suspect it has the greatest effect on the implementation of purported crime control measures such as RTC legislation. Of course, this would suggest an endogeneity problem that would also likely lead to a bias in favor of finding a deterrent effect. The mechanism driving this bias would presumably be that rising crime strengthens the NRA push for the law, and the mean reversion in crime would then falsely be attributed to the law by the naive panel data analysis (incorrectly premised on exogenous RTC law adoption). Post-adoption murder rates again decline—often to within the neighborhood of pre-law levels. We do, however, uncover some interesting findings when estimating (more cleanly) the year-by-year effects on the 25 states for which we have observations across the full set of dummy variables.

Another striking feature we note is the strong influence of Florida and Georgia on our estimates of the impact of RTC laws on murder (Figure 10) and rape (Figure 11). When we remove these two states, the post-adoption trend lines for murder and rape shift upwards substantially. Moreover, when dropping them from the set of RTC states that already excludes the early and late adopters—still leaving us with 23 RTC states to analyze—we see that murder increases in each post-adoption year except one. As previous papers have noted, Florida experienced enormous drops in murder during the 1990s that may have been completely unrelated to the passage of its right-to-carry policy. Donohue (2003) points out that the 1980 Mariel boat lift temporarily added many individuals prone to committing crimes to Florida's population, causing a massive increase in crime in Florida during the 1980s. Thus, it is plausible that the massive 1990s crime reductions in Florida were not driven by the adoption of the state's RTC law but rather a return to traditional population dynamics that were less prone to violent crime (again, a reversion to the mean). This is important to consider given the strong downward pull of Florida on aggregate murder rates.

The line based on dropping Florida and Georgia from the trimmed sample would suggest that for the 23 other states, the impact of RTC laws on murder was highly pernicious -- and increasingly so as the sharp upward trend in the last three years would suggest. Again a number of interpretations are possible: 1) Florida and Georgia are unusual and the best estimate of the impact of RTC laws comes from the trimmed sample that excludes them (and the early and late adopters); 2) there is heterogeneity in the impact of RTC laws, so we should conclude that the laws help in Florida and Georgia, and tend to be harmful in the other 23 states; and 3) omitted variables mar the state-by-state estimates but the aggregate estimates that include Florida and Georgia may be reasonable if the state-by-state biases on average cancel out.

Note that Figure 11, which presents the comparable year-by-year estimates of the impact of RTC laws on rape, shows a similar yet even more extreme pattern of apparent spikes in crime leading to adoption of RTC laws followed by a substantial amount of mean reversion. The somewhat unsettling conclusion from Figures 10 and 11 is that RTC laws might look beneficial if one only had data for four or five years, but this conclusion might be substantially reversed if a few additional years of data were analyzed. Taken as a whole, these two figures show the sensitivity of the estimates to both the time period and sample of states that are analyzed.

Figure 11[45]

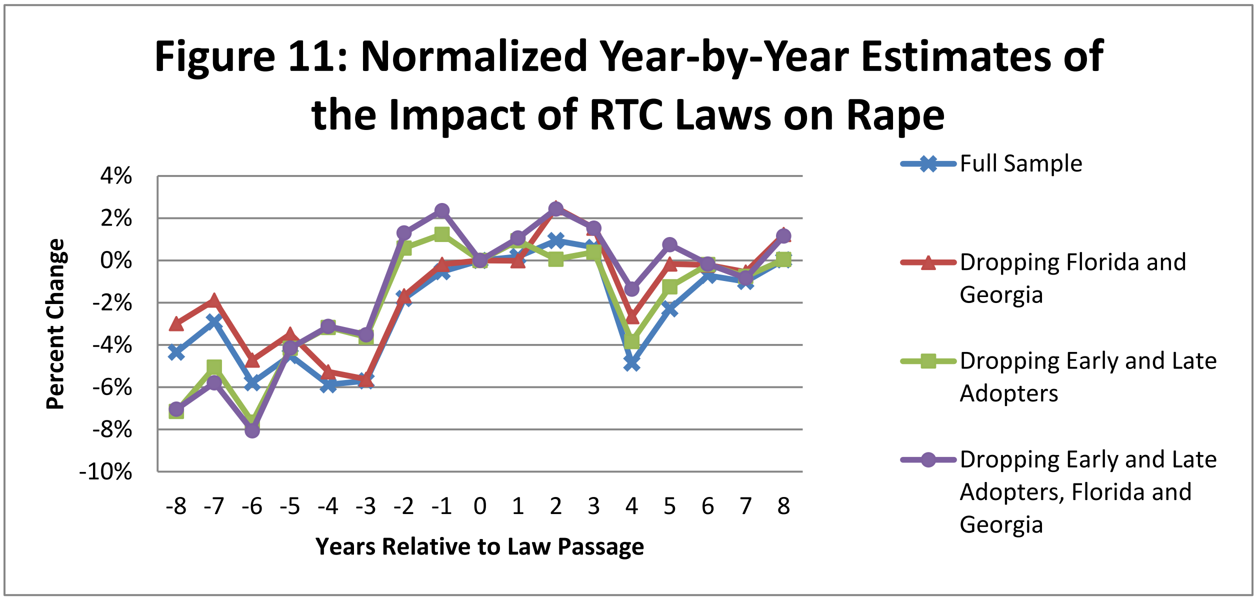


Figure 12[46]

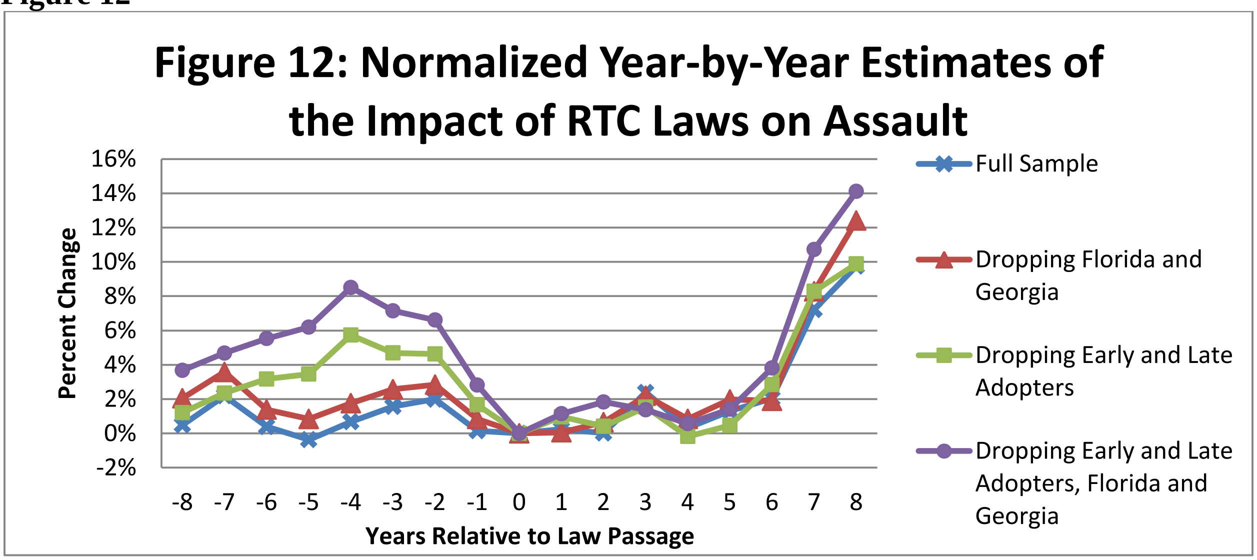


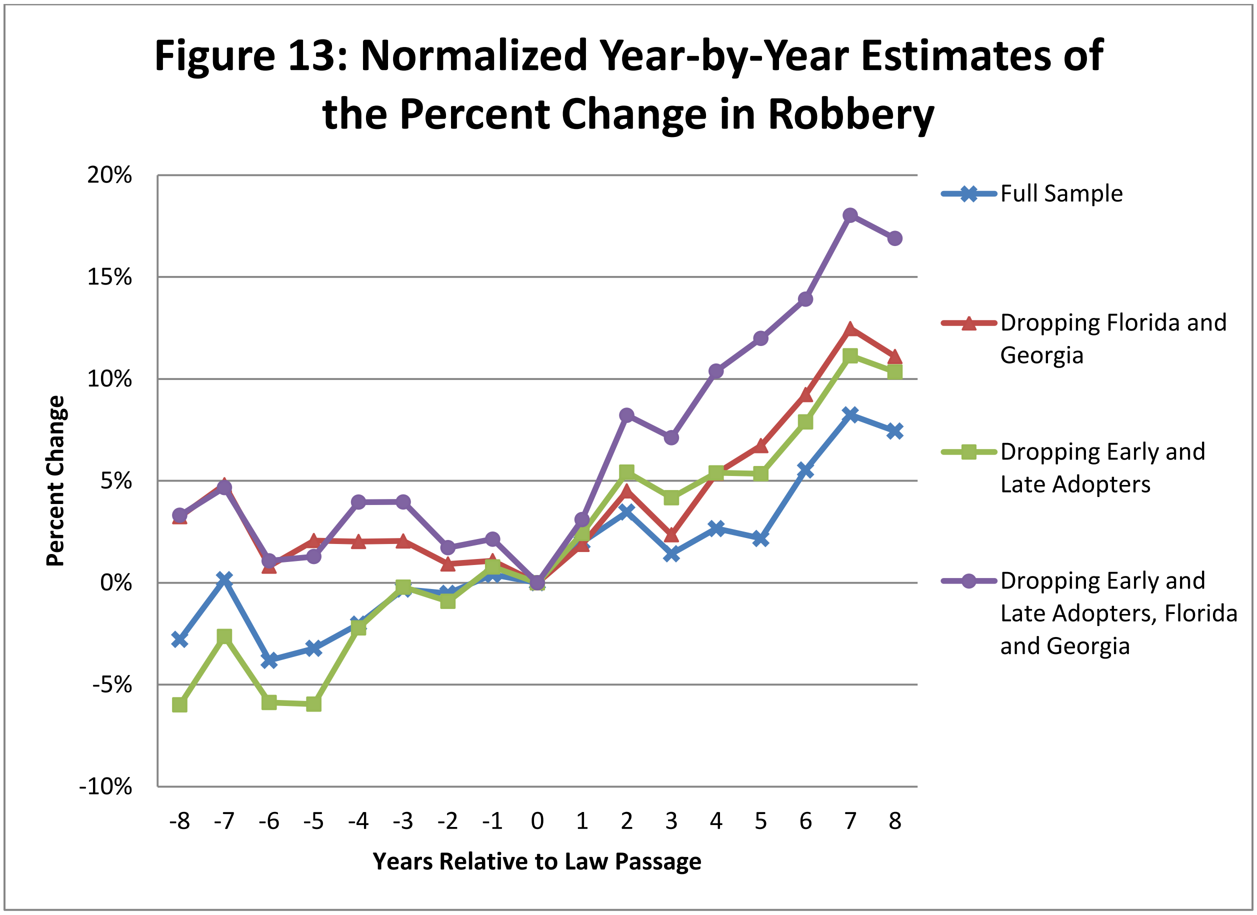
Figure 13[47]

---

[45] Estimations include year and state fixed effects, state trends, and are weighted by state population. The control variables include: incarceration and police rates, unemployment rate, poverty rate, state population, population density, per capita income, and six demographic composition measures.

[46] Estimations include year and state fixed effects, state trends, and are weighted by state population. The control variables include: incarceration and police rates, unemployment rate, poverty rate, state population, population density, per capita income, and six demographic composition measures.

[47] Estimations include year and state fixed effects, state trends, and are weighted by state population. The control variables include: incarceration and police rates, unemployment rate, poverty rate, state population, population density, per capita income, and six demographic composition measures.



Further casting doubt on the possibility that drops in murder and rape could be attributed to the passage of RTC laws, a dramatically different picture emerges from our year-by-year analysis of these laws' impact on assault and robbery rates. The general story here seems to be that assault increases markedly over the time period after law passage, which squares with our results discussed in previous sections. One observes positive coefficient changes that are initially modest, but that increase dramatically and uniformly over the second half of the post-passage period. Moreover, in contrast to the year-by-year murder and rape estimates, assault trends are not demonstrably different when we alter the sample to exclude early and late adopters, as well as Florida and Georgia. The pattern is generally unaffected by sample, giving us some confidence that RTC laws may be having an adverse impact on the rate of assault. Robbery rates similarly increase over time after the passage of RTC laws, although not as dramatically.

Something to consider, however, is how one should interpret the assault trends in light of the murder trends just discussed. Unlike the pattern in murder which at first looked favorable following RTC passage and then turned more pernicious, there is no sign of any improvement but then a sudden upward turn in assaults after year six. If the near uniform increases in assault coefficients means that aggravated assault did actually increase over time with the passage of right-to-carry legislation, this would strongly undercut the "More Guns, Less Crime" thesis. Interestingly, the robbery data (Figure 13) either suggests a pernicious effect similar to that on aggravated assault (particularly for the trimmed estimates dropping only early and late adopters) or a strong upward trend in crime, starting well before passage, that might be taken as a sign of the absence of any impact of RTC laws on robbery.

C. Effects of RTC Laws on Gun-related Assaults

A general concern in evaluating the impact of generic law X is that there is not some other law or policy Y that is generating the observed effect. In this case, the apparent finding that RTC laws increase aggravated assaults raises the question of whether changes in reporting or documenting aggravated assaults might be a possible confounding factor. Specifically, over the last two decades a number of states and municipalities have launched programs designed to combat domestic violence by increasing the arrests of likely perpetrators. These programs could influence the count of aggravated assaults appearing in the FBI crime data we employ. If such programs are more likely to be adopted in either RTC or non-RTC states than the potential for bias must be considered.

One way to address this problem would be to collect data on the various state or municipal initiatives that lead to higher rates of arrest of those committing acts of domestic violence. However, collecting uniform panel data along these lines that also fully captures the nature and intensity of the police initiatives is extremely difficult. An alternative approach is to look at assaults that we think are less likely to be influenced by these domestic violence initiatives (or by other shifts in likelihood of arrest for potentially assaultive conduct), but which are most likely to be influenced by RTC laws (if there is in fact such an influence). Counts of gun assaults would seem to meet these two criteria, because assaults with a gun tend to be serious enough that the level of discretion as to whether to arrest is reduced, and because gun assaults are precisely the types of crimes that we might expect would be influenced if more guns are on the street because of the passage of RTC laws. For this reason, we may get more reliable estimates of the impact of RTC laws by looking at gun-related aggravated assaults than at overall aggravated assaults.

To test this possibility, we estimate our preferred regression using gun-related aggravated assaults as the dependent variable (both with and without state-specific trends) in Table 14 below. Comparing these new results with the assault estimates in Tables 8a and 8b above, our bottom-line story of how RTC laws increase rates of aggravated assault does not change much when limiting our analysis to assaults involving a gun. Without state trends, we see large positive estimates, some of which are significant at the 5% and 10% level. With state trends, we again see some significant evidence that gun-related aggravated assault rates are increased by RTC legislation. These results again suggest that RTC laws may be generating higher levels of assaultive conduct, although more refined tools (or cleaner data) will be needed before confident predictions can be made.

Table 14[48]

Estimated Impact of RTC Laws on Gun-Related Aggravated Assaults – ADZ Preferred Controls, 1977-2006 – Clustered Standard Errors

Dataset: ADZ Updated 2013 State Data

All figures reported in %	Gun-Related Aggravated Assault (No State Trends)	Gun-Related Aggravated Assault (With State Trends)
Dummy Variable Model:	23.34**	1.57
	(11.03)	(6.72)
Spline Model:	2.71	4.45
	(1.74)	(2.67)
Hybrid Post-Passage Dummy:	14.47*	-0.53
	(7.75)	(6.49)
Trend Effect:	1.97	4.46*
	(1.66)	(2.66)

XI. Conclusion

In this paper, we have explored the question of the impact of RTC laws on crime and the NRC panel's 2004 report concluding that the then-current literature was too fractured to reach a conclusion on what that impact is. We agree with the conclusion that the NRC panel reached at

[48] Estimations include year and state fixed effects, and are weighted by state population. Robust standard errors are provided beneath point estimates in parentheses. The control variables for this "preferred" specification include: incarceration and police rates (lagged one year to avoid potential endogeneity issues), unemployment rate, poverty rate, population density, per capita income measures, and six demographic composition measures. * Significant at 10%; ** Significant at 5%; *** Significant at 1%. The gun assault data comes from the FBI master file, available upon request from the agency. The data is provided at the local level; thus for state values we sum the reported gun assaults over all of the reporting agencies by year. However, not all agencies report their estimates during each reporting period, leaving our gun assault figures likely to be undervalued.

that time, as well as with the pointed rebuke the panel gave to James Q. Wilson who argued -- without scientific merit according to the NRC majority -- that RTC laws reduce murder. But while the NRC majority was correct in both its main conclusion and its rejection of Wilson's position, the NRC majority report is also subject to criticism on a number of grounds.

First, as we show in this paper, there is a clear need to employ the cluster correction to the standard errors when estimating panel data models of crime, and the NRC majority erred when it concluded otherwise. As our placebo tests show, the standard errors that the NRC presented in their panel data models were far too low and greatly exaggerated the statistical significance of their results. Indeed, the clustering gaffe was on top of the NRC failure to use the robust correction for heteroskedascticity, which further downward biased the standard errors (although less dramatically then the failure to cluster). Both corrections are needed, and this error alone set the stage for Wilson's dissent. With correct standard errors, none of the estimates that Wilson thought established a benign effect of RTC laws on murder would have been statistically significant. Thus, getting the standard errors right might have kept Wilson from writing his misguided dissent -- to the benefit of Wilson, the NRC majority, and the public.

Second, beyond getting the standard errors correct and therefore undermining the ostensible statistical significance of their presented murder regression, the NRC majority could have said much more than they did to refute Wilson's reliance on two Lott and Mustard regressions to endorse the view that RTC laws reduce murder. Wilson's conclusion essentially rested on the NRC report's presentation of two Lott and Mustard models (the dummy and the spline) based on county data from 1977-2000. The NRC majority did point out that the estimates for six out of 7 crimes were contradictory (some suggesting crime increases and some suggesting crime decreases), so the fact that for the seventh crime -- murder -- both models suggested RTC laws reduced crime might well be a spurious result. But the NRC majority could have given many more reasons to be cautious about relying on the two Lott and Mustard regressions.

Specifically, the NRC response to Wilson could easily have noted that Wilson had previously written that incarceration was perhaps the most important factor explaining the drop in crime in the United States in the 1990s, and he had also written on the importance of police (Wilson, 2008). Yet the Lott and Mustard model that the NRC presented (and that Wilson relied on) did not control for either of these factors.[49] Thus, on this grounds alone, one would have thought Wilson would have been particularly wary not to rely on a regression, which was potentially subject to a charge of omitted variable bias. Neither the NRC majority nor Wilson ever noted this omission.

Moreover, we note in this paper some of the data problems with the Lott data set that the NRC panel used, and then address an array of issues about data and model specification that Wilson ideally should have explored before he uncritically accepted the ostensible finding of a RTC impact on murder. Issues such as the inclusion of state-specific linear trends, the danger of omitted variable bias concerning the crack epidemic, and the choice of county over state-level data -- all of which the NRC neglected to discuss—clearly have enough impact on the panel data estimates to influence one's perception of the "More Guns, Less Crime" theory and thus warrant closer examination.

Perhaps Wilson was so wedded to his position that nothing could have persuaded him not to write his ill-conceived dissent, but the NRC majority could have done more to buttress their

[49] The Lott and Mustard model omitted a control for the incarceration and police rates (which is indicated implicitly--though not explicitly—in the notes to each table of the NRC report, which listed the controls included in each specification).

entirely correct assessment that "the scientific evidence does not support [Wilson's] position" (pg. 275) As a result, Lott, with at best questionable support for his view that RTC laws reduce murder, now claims that Wilson, one of the most eminent criminologists of our time, supports his position (Lott, 2008). If one of the goals of the NRC report was to shield the public and policymakers from claims based on inadequate empirical evidence, the Wilson dissent represents a considerable failure.

A number of important lessons emerge from this story for both producers and consumers of econometric evaluations of law and policy. The first and most obvious is that a single statistical study cannot resolve an important question. Instead, one must wait until a literature has developed. But even then, the conclusion that emerges may be one of uncertainty as the NRC report showed.

A second lesson is how easy it is for mistakes to creep into these empirical studies. The pure data errors that entered into the NRC data set when Lott transmitted an imperfect data set or the error in the 1993 Uniform Crime Reports data (or the errors that entered into our own work in Aneja et al, 2011) were not major enough to have an impact, but at times the errors will be decisive (and the process of peer review is not well-equipped to detect such errors). This episode underscores the value of making publicly available data and replication files that can re-produce published econometric results. This exercise can both help to uncover errors prior to publication and then assist researchers in the process of replication, thereby aiding the process of ensuring accurate econometric estimates that later inform policy debates.

A third lesson is that the "best practices" in econometrics are evolving. Researchers and policymakers should keep an open mind about controversial policy topics in light of new and better empirical evidence or methodologies. Prior to the important work of Bertrand, Duflo, and

Mullainathan (2004) on difference-in-differences estimation, almost no one understood that clustering standard errors on the state-level in order to account for serial correlation in panel data was necessary. The results in many published papers would be wiped out with this adjustment. Even with its impressive array of talent, the NRC report in 2004 got this important issue wrong, even though most applied econometricians today would make this cluster adjustment to avoid greatly increasing the level of Type I error.

While the NRC majority decision of uncertainty was clearly influenced by the sensitivity of the estimates to various modeling choices, the statement by Horowitz was even more categorical in its nihilism, essentially rejecting all applied econometric work on RTC legislation, as indicated by his following independent statement in an appendix to the NRC's (2004) report:

> "It is unlikely that there can be an empirically based resolution of the question of whether Lott has reached the correct conclusions about the effects of right-to-carry laws on crime." (p. 304, NRC Report.)

Of course, if there can be no empirically based resolution of this question, it means that short of doing an experiment in which laws are randomly assigned to states, there will be no way to assess the impact of these laws. But there is nothing particularly special about the RTC issue, as the recent National Research Council report on the deterrence of the death penalty shows (essentially adopting the Horowitz position on the question of whether the death penalty deters murders). The econometrics community needs to think deeply about what these NRC reports and the Horowitz appendix imply more broadly for the study of legislation using panel data econometrics and observational data.

Finally, despite our belief that the NRC's analysis was imperfect in certain ways, we agree with the committee's cautious final judgment on the effects of RTC laws: "with the current evidence it is not possible to determine that there is a causal link between the passage of right-to-carry laws and crime rates." Our results here further underscore the sensitivity of guns-crime estimates to modeling decisions.[50] But not being able to "determine" with the level of certainty of beyond a reasonable doubt does not mean that one cannot offer conclusions at some lower level of certainty such as "more probable than not." Since policymakers need to act, it is useful to offer guidance as to which evidence is likely to be most reliable.

If one had to make judgments based on all of the panel data models of the type used in the NRC report, one would have to conclude that RTC laws likely increase the rate of aggravated assault.[51] In addition, we would consider our preferred regression models run on either the most complete data (state data from 1977-2010) or the data likely to be free of the confounding effect of the crack cocaine epidemic (state data from 1999-2010) as likely to yield the most reliable estimates of the effect of RTC laws on crime. If we estimate both the dummy and spline models using our preferred specification for each of these two time period (overall or after 1999), then we have 4 estimates of the impact of RTC laws for each of seven crime categories. In six of the seven crime categories, at least one of these four estimates suggests that RTC laws increase crime at the .05 level of significance. For the seventh crime category, the evidence that RTC laws increase robbery is only statistically significant at the .10 level. These crime increases are substantial, with the dummy variable model for the complete period (Table 9a) suggesting that RTC laws increased every crime category by at least 9 percent, except murder (in that model, murder rose 2 percent by it is not statistically significant). For the post-1999 regressions, spline estimate (Table 10a) suggests that RTC laws increased the rate of murder by 1.4 percentage

[50] For a quick and clear sense of how sensitive estimates of the impact of right-to-carry laws are, see Appendix D, where we visually demonstrate the range of point estimates we obtain throughout our analysis.
[51] Note that the assaults can be committed either by RTC permit holders or those who have acquired their guns -- either via theft or appropriation of lost guns.

points each year (significant at the .05 level). In none of those 28 regressions was there any statistically significant estimate suggesting that RTC laws decreased crime.

The evidence that RTC laws increase crime is strongest if one accepts the dummy variable model with our preferred specification for the full time period and accepts the Wolfers (2006) critique that one should avoid controlling for state trends (the Table 9a results). But even here questions remain. While one can imagine that RTC laws increase violent crime either directly by putting guns into the hands of those who then commit crimes or indirectly by making guns more available to criminals (when lost or stolen), it is not clear why the property crimes of burglary, auto theft, and larceny would rise as a result of RTC passage.

Three possibilities come to mind. First, the results are correctly capturing the impact of RTC laws and perhaps the indirect effect of increasing the weapons available to criminals (through loss or theft) facilitates all criminal activity (perhaps by emboldening newly armed criminals) or the increase in violent crime diverts police resources so that property crime is stimulated. Second, it is possible that states adopting RTC laws were less successful in fighting crime than non-adopting states, so the RTC law was not itself increasing crime but was simply a proxy for states that on the whole adopted less successful crime-fighting strategies over the last quarter century. Third, it is possible that states chose to adopt RTC laws at a time when crime was on the rise, so their post-passage crime experience reflects an adverse crime shock that is incorrectly causally attributed to RTC laws. If this endogenous timing argument is correct, then it might suggest that post-1999 estimates of Table 10a are preferable, since that has been a period of greater crime stability (as opposed to the dramatic crime swings of the late 1980s and 1990s). The Table 10a estimates show that RTC laws only affected one crime category – with the laws causing a substantial *increase* in murder. Further research will hopefully further refine our conclusions as more data and better methodologies are employed to estimate the impact of RTC laws on crime.

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber. 2005. "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools." *Journal of Political Economy* 113(1): 151-184.

Angrist, Joshua and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

Autor, David, John J. Donohue, and Stewart Schwab. 2006. "The Costs of Wrongful-Discharge Laws." *Review of Economics and Statistics* 88(2): 211-231.

Ayres, Ian and John J. Donohue. 2003a. "Shooting Down the More Guns, Less Crime Hypothesis." *Stanford Law Review*, 55(4): 1193-1312.

Ayres, Ian and John J. Donohue. 2003b. "The Latest Misfires in Support of the More Guns, Less Crime Hypothesis." *Stanford Law Review*, 55(4): 1371-1398.

Ayres, Ian and John J. Donohue. 2009. "More Guns Less Crime Fails Again: The Latest Evidence from 1977-2006." *Econ Journal Watch*, 6(2): 218-238. http://www.aier.org/aier/publications/ejw_com_may09_ayresdonohue.pdf

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

Black, Dan A., and Daniel S. Nagin. 1998. ''Do Right-to-Carry Laws Deter Violent Crime?'' *Journal of Legal Studies*, 27: 209-219.

Collins, Gail. 2009. "Have Gun, Will Travel." *The New York Times*. 31 July 2009.

Donohue, John J. 2003. "The Impact of Concealed-carry Laws." *Evaluating Gun Policy*. J. Ludwig & P. J. Cook (Eds.). Washington, DC: Brookings Institution Press. 287–324.

Donohue, John J. 2004. “Guns, Crime, and the Impact of State Right-to-Carry Laws.” *Fordham Law Review*, 73: 623-652.

Donohue, John J. and Justin Wolfers. 2009. “Estimating the Impact of the Death Penalty on Murder.” *American Law and Economics Review, 11 (2): 249-309*

Elder, Todd and Christopher Jepsen. 2013. “Are Catholic Primary Schools More Effective Than Public Primary Schools?” *Journal of Urban Economics*, forthcoming.

Fryer, Roland, Paul Heaton, Steven Levitt, and Kevin Murphy. 2005. “Measuring the Impact of Crack Cocaine.” NBER Working Paper Series No. W11318. National Bureau of Economic Research, Cambridge, MA.

Horowitz, Joel L. "Should the DEA's STRIDE Data Be Used for Economic Analyses of Markets for Illegal Drugs?" *Journal of the American Statistical Association*, 96(465): 1254-1271.

Kovandzic, T. V., Vieraitis, L. M. and Boots, D. P. (2009), Does the death penalty save lives? *Criminology & Public Policy*, 8: 803–843.

Levitt, Steven D. 2004. “Understanding Why Crime Fell in the 1990’s: Four Factors that Explain the Decline and Six that Do Not,” *Journal of Economic Perspectives*, 17: 163-190.

Lott, John R. 2000. *More Guns, Less Crime*. Chicago: University of Chicago Press.

Lott, John R. 2004. “Right-to-Carry Laws and Violent Crime Revisited: Clustering, Measurement Error, and State-by-State Breakdowns.” http://ssrn.com/abstract=523002 or doi:10.2139/ssrn.523002.

Lott, John R. 2008. "Do Guns Reduce Crime?" *Intelligence Squared Debate Series*. http://intelligencesquaredus.org/wp-content/uploads/Guns-Reduce-Crime-102808.pdf

Lott, John R. and David Mustard. 1997. "Crime, Deterrence and Right-to-Carry Concealed Handguns." *Journal of Legal Studies*, 26(1): 1-68.

Ludwig, J. 1998. "Concealed Gun-carrying Laws and Violent Crime: Evidence from State Panel Data." *International Review of Law and Economics*, 18: 239-254.

Maltz, Michael D. 2006. *Analysis of Missingness in UCR Crime Data*. NCJ 215343 Washington: U.S. Department of Justice.

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

Maltz, Michael D., & J. Targonski. 2003. "Measurement and other errors in county-level UCR data: A reply to Lott and Whitley." *Journal of Quantitative Criminology*, 19: 199-206.

Moody, Carlisle E. and Thomas B. Marvell. 2008. "The Debate on Shall-Issue Laws." *Econ Journal Watch*, 5(3): 269-293. http://www.aier.org/ejw/archive/doc_view/3610-ejw-200809?tmpl=component&format=raw.

Moody, Carlisle E. John R Lott, Jr., Thomas B. Marvell, and Paul R. Zimmerman. 2012. "Trust but Verify: Lessons for the Empirical Evaluation of Law and Policy."

Moulton, Brent. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics*, 72: 334-338.

Nagin, Daniel S. and John V. Pepper, editors, 2012. *Deterrence and the Death Penalty*. Washington: The National Academies Press.

National Research Council. 2004. *Firearms and Violence: A Critical Review*. Washington: The National Academies Press.

Plassman, Florenz and John Whitley. 2003. "Confirming More Guns, Less Crime." *Stanford Law Review*, 2003: 1313-1369.

Wilson, James Q. 2000. "Guns and Bush." *Slate Politics*. http://slate.msn.com/?id=91132. (accessed on November 29 2009).

Wilson, James Q. 2008. "What Do We Get From Prison?" *The Volokh Conspiracy*. http://volokh.com/posts/chain_1213046814.shtml. (accessed on 20 November 2009).

Wooldridge, Jeffrey M. 2003. "Cluster-sample Methods in Applied Econometrics" *American Economic Review*, 93: 133-138.

Wooldridge, Jeffrey M. 2006. "Cluster-Sample Methods in Applied Econometrics: An Extended Analysis." Unpublished manuscript. Michigan State University.

Wolfers, Justin. 2006. "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results." American Economic Review, 96(5): 1802-1820.

Zimring, Franklin and Gordon Hawkins. 1997. "Concealed handguns: The counterfeit deterrent." *The Responsive Community*, 7:46-60.

Appendix A: Using Placebo Laws to Test the Impact of Clustering in the State Data

Table 3 reports the results of our placebo tests using county data. In this appendix, we use state-level data to again conduct our experiment with placebo laws to examine the effects of clustering the standard errors. As seen in Tables 1-4 of Appendix A, we find results similar to those generated with our county data: without clustering, the Type 1 error rates are often an order of magnitude too high or worse for our murder and robbery regressions (see Tables A1 and A3). In fact, even *with* clustered standard errors (Tables A2 and A4), the rejection of the null hypothesis (that RTC laws have no significant impact on crime) occurs at a relatively high rate. This finding suggests that, at the very least, we should include clustered standard errors to avoid unreasonably high numbers of significant estimates.

Appendix A[52]

Table A1

Percentage of Significant Estimates (5% Level) – Lott-Mustard Controls, 1977-2006 – Hybrid Model
Dataset: ADZ Updated 2013 State Data

All figures reported in %		Dummy Variable	Trend Variable
1. All 50 States + DC:	Murder	40.8	61.9
	Robbery	44.9	62.7
2. Exact 32 States:	Murder	42.0	55.3
	Robbery	45.3	57.9
3. Random 32 States:	Murder	41.8	55.3
	Robbery	47.9	61.5
4. All 19 States:	Murder	34.3	38.5
	Robbery	40.3	72.8
5. Random 12 States:	Murder	36.8	46.2
	Robbery	48.9	70.5

Table A2

Percentage of Significant Estimates (5% Level) – Lott-Mustard Controls, 1977-2006 – Hybrid Model and **Clustered Standard Errors**
Dataset: ADZ Updated 2013 State Data

All figures reported in %		Dummy Variable	Trend Variable
1. All 50 States + DC:	Murder	14.3	24.4
	Robbery	12.0	18.6
2. Exact 32 States:	Murder	14.1	16.7
	Robbery	11.6	14.7
3. Random 32 States	Murder	14.5	14.5
	Robbery	14.1	18.1
4. All 19 States	Murder	21.9	32.7
	Robbery	20.9	49.4
5. Random 12 States:	Murder	25.1	35.4
	Robbery	28.0	47.5

[52] Simulation based on NRC with-controls model, includes year fixed effects, state fixed effects, and weighting by state population. The control variables (adopted from the Lott-Mustard model) include: lagged arrest rate, state population, population density, per capita income measures, and 36 demographic composition measures indicating the percentage of the population belonging to a race-age-gender group.

Appendix A (Cont.)

Table A3

Percentage of Significant Estimates (5% Level) – Lott-Mustard Controls, 1977-2006 – Dummy Model
Dataset: ADZ Updated 2013 State Data

All figures reported in %		Dummy Variable
1. All 50 States + DC:	Murder	39.6
	Robbery	45.6
2. Exact 32 States:	Murder	36.1
	Robbery	38.0
3. Random 32 States:	Murder	40.0
	Robbery	47.6
4. All 19 States:	Murder	31.9
	Robbery	40.5
5. Random 12 States:	Murder	34.6
	Robbery	46.6

Table A4

Percentage of Significant Estimates (5% Level) – Lott-Mustard Controls, 1977-2006 – Dummy Model and **Clustered Standard Errors**
Dataset: ADZ Updated 2013 State Data

All figures reported in %		Dummy Variable
1. All 50 States + DC:	Murder	13.8
	Robbery	12.3
2. Exact 32 States:	Murder	8.1
	Robbery	4.9
3. Random 32 States	Murder	13.4
	Robbery	12.3
4. All 19 States	Murder	20.8
	Robbery	22.4
5. Random 12 States:	Murder	24.1
	Robbery	27.0

Appendix B – Panel Data Models over the Full Period with No Covariates

The NRC panel sought to underscore the importance of finding the correct set of covariates by presenting panel data estimates of the impact of RTC without covariates but including county and year fixed effects. For completeness, this Appendix presents these same estimates for the preferred models (with and without state trends) on both county and state data for the period from 1977-2006.

If one compares the results from these four tables with no controls with the analogous tables using the preferred model for the same time period, one sees some interesting patterns. For example, if we compare the county results without state trends from both our preferred specification (Table 6a) and the no-controls specification (Table B1), we see that the results are quite similar in terms of magnitude and direction, although adding in our suggested covariates seems to both dampen the coefficients and reduce their significance. The basic story from our analysis is again strengthened: there seems to be virtually no effect of RTC laws on murder, while if there is *any* RTC effect on other crimes generally, it is a crime-*increasing* effect. The results are slightly less similar when we compare those from the models that include state trends (Tables 6b and B2). While we see that estimates are similar for murder, rape, and burglary, the estimates for assault, robbery, auto theft, and larceny change in either magnitude or direction (or both) when adding controlling factors to the model. In general, though, we only see decreases when adding state trends to either specification, and even then, the results are much too imprecise to make causal inferences.

When we shift to a comparison of the state-level results, we again see similarities between the preferred and no-controls specifications. When looking at the results without state trends, we see that the estimates are fairly similar in terms of direction, although the no-controls estimates are often larger in magnitude and more statistically significant. When doing a similar comparison of the specifications that now add in state trends, we also see similar results for nearly all crimes. The exception is aggravated assault, where we see that our preferred specification's estimate that RTC laws increase aggravated assault when state trends are added. Again, when the comparison is taken as a whole, consistent and robust support is lacking for the view that RTC laws lead to reductions in crime.

Appendix B[53]

Table B1

Estimated Impact of RTC Laws – No Controls, 1977-2006 – Clustered Standard Errors
Dataset: ADZ Updated 2013 County Data (without 1993 data)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	0.03	32.80	28.77	27.32*	34.99	38.15*	40.45
	(8.40)	(22.49)	(18.91)	(14.54)	(21.43)	(21.96)	(25.43)
Spline Model:	0.36	3.36*	3.22*	2.88**	3.42*	3.86*	4.27*
	(0.76)	(1.94)	(1.67)	(1.36)	(2.02)	(2.01)	(2.30)
Hybrid Post-Passage Dummy:	-2.87	20.59	15.84	16.50	23.35	24.31	24.35
	(8.39)	(18.58)	(17.19)	(12.95)	(17.86)	(19.30)	(22.00)
Trend Effect:	0.52	2.19*	2.32*	1.94*	2.09	2.49	2.89
	(0.74)	(1.27)	(1.31)	(1.11)	(1.51)	(1.58)	(1.74)

Table B2

Estimated Impact of RTC Laws – No Controls, 1977-2006 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 County Data (without 1993 data)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-3.15	-15.78	3.42	2.18	3.64	-0.79	0.24
	(5.11)	(10.98)	(9.01)	(7.63)	(9.00)	(10.15)	(11.19)
Spline Model:	0.07	-5.47	2.67	0.36	0.53	-0.31	-0.23
	(1.29)	(4.40)	(2.41)	(2.44)	(2.64)	(2.53)	(2.68)
Hybrid Post-Passage Dummy:	-3.24	-12.39	1.70	1.98	3.35	-0.59	0.40
	(5.25)	(10.16)	(9.73)	(8.15)	(9.67)	(11.05)	(12.17)
Trend Effect:	0.14	-5.18	2.63	0.31	0.45	-0.30	-0.24
	(1.31)	(4.39)	(2.50)	(2.52)	(2.73)	(2.68)	(2.84)

[53] Estimations include year and county fixed effects, and are weighted by county population. Robust standard errors are provided beneath point estimates in parentheses. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Appendix B (Cont.)

Table B3

Estimated Impact of RTC Laws – No Controls, 1977-2006 – Clustered Standard Errors
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-1.77	9.46	11.41**	19.72**	23.88**	19.35***	17.11***
	(7.61)	(8.40)	(4.71)	(7.93)	(11.01)	(7.19)	(5.44)
Spline Model:	0.15	1.01	1.50**	2.00**	1.92	1.71**	1.55**
	(0.89)	(0.92)	(0.64)	(0.97)	(1.17)	(0.81)	(0.66)
Hybrid Post-Passage Dummy:	-3.70	5.66	4.69	12.68*	19.29**	14.42**	12.41***
	(6.90)	(5.79)	(3.78)	(7.27)	(8.29)	(6.01)	(4.61)
Trend Effect:	0.35	0.70	1.24*	1.30	0.84	0.91	0.86
	(0.89)	(0.74)	(0.63)	(0.96)	(0.90)	(0.71)	(0.60)

Table B4

Estimated Impact of RTC Laws – No Controls, 1977-2006 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 State Data

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-0.67	-4.56**	0.30	3.05	8.55	2.14	1.54
	(3.71)	(2.08)	(3.27)	(4.82)	(5.67)	(3.50)	(2.56)
Spline Model:	1.02	-0.30	2.78***	0.62	-0.15	-0.25	-0.43
	(0.93)	(0.93)	(0.91)	(1.37)	(1.43)	(0.79)	(0.60)
Hybrid Post-Passage Dummy:	-1.29	-4.44**	-1.35	2.72	8.75	2.32	1.82
	(3.63)	(2.12)	(3.26)	(5.20)	(5.66)	(3.76)	(2.64)
Trend Effect:	1.05	-0.20	2.81***	0.56	-0.35	-0.31	-0.47
	(0.93)	(0.93)	(0.91)	(1.41)	(1.42)	(0.83)	(0.63)

Appendix C – Trimming the Sample to Address Questions of Model Fit

Given our concerns about how well the guns-crime econometric models fit all 50 US states (plus D.C.), we decided to examine the residuals from various regressions models. For example, one potentially important issue is whether one should include linear state trends in our models. To further explore this issue, we examined the variance of the residuals for the aggravated assault regression estimates using our preferred models on state data for the period through 2006—both with and without state trends.[54] In particular, we found that the residual variance was high for smaller states, even when we do not weight our regressions by population.[55]

We explored how these "high residual-variance" states (defined from the aggravated assault regressions on our preferred model through 2006) might be influencing the results. We estimated our preferred model (both with and without state trends) after removing the 10 percent of states with the highest residual variance. This step is also repeated after removing the highest 20 percent of states in terms of residual variance. Our full-sample results for our preferred specification (which includes clustered standard errors, and is run over the 1977-2006 time period) are shown in Table 8a and 8b (without and with state trends, respectively). The results from our two trimmed set of states are presented below. Tables C1 and C2 should be compared to Table 8a (no state trends) and Tables C3 and C4 should be compared to Table 8b (adding in state trends).

[54] Since our most robust results across the specifications in this paper were for aggravated assault, we focused specifically on the residuals obtained using assault rate as the dependent variable.

[55] We removed the population weight for this exercise because it is likely that when regressions are weighted by population, the regression model will naturally make high-population states fit the data better. As a result, we expect that residuals for smaller states will be higher. We find, however, that the results are qualitatively similar even when we obtain the residuals from regressions that include the population-weighting scheme.

Removing high residual-variance states (based on the aggravated assault regressions) has little impact on the story told in Table 8a (no state trends): there was no hint that RTC laws reduce crime in Table 8a and this message comes through again in Tables C1 and C2. All three of these tables show at least some evidence that RTC laws *increase* aggravated assault. Removing the high residual-variance states from the models with state trends does nothing to shake the Table 8b finding that RTC laws *increase* aggravated assault. The somewhat mixed results for auto theft seen in Table 8b also remain in Table C3 and C4.

Of the states dropped from Tables C1, the following four states adopted RTC laws during the 1977-2006 period (with date of adoption in parentheses): Montana (1991), Maine (1985), West Virginia (1989), and Nebraska (2006).[56] Of the *additional* states dropped from Table C2, the following three states adopted RTC laws during the 1977-2006 period (with date of adoption in parentheses): Tennessee (1994), North Dakota (1985), and Kentucky (1996). Results from Table C3 come from dropping similar RTC states to Table C1, although Vermont is dropped rather than Maine.[57] Finally, in addition to the five RTC states that were dropped in Table C3, Table C4 dropped the following five RTC states: Kentucky (1996), Nevada (1995), North Dakota (1985), Indiana (1980), and Delaware (No RTC law).

[56] Nebraska passed their RTC law in 2006, which is not in our analysis period because the full first year after passage is in 2007.

[57] The dropped states are slightly different between Tables C1 and C3, as well as between Tables C2 and C4, because the state ranks based on residual variances differed when the models were run with and without state trends.

Appendix C[58]

Table C1

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2006 – Clustered Standard Errors
Dataset: ADZ Updated 2013 State Data
Dropping States with Highest Residual Variance (Top 10%: MT, WV, NE, NH, ME)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	1.31	10.74**	10.08**	14.88**	18.64**	11.58**	12.13***
	(5.51)	(4.98)	(3.82)	(6.69)	(7.50)	(5.39)	(4.23)
Spline Model:	0.22	0.77	1.50**	1.35	1.40	0.85	1.12*
	(0.81)	(0.70)	(0.66)	(0.99)	(0.85)	(0.68)	(0.59)
Hybrid Post-Passage Dummy:	0.41	9.48**	4.48	11.56*	16.11**	10.13**	9.27**
	(4.45)	(4.32)	(3.46)	(6.27)	(6.30)	(4.54)	(3.67)
Trend Effect:	0.20	0.29	1.27*	0.75	0.57	0.33	0.65
	(0.79)	(0.65)	(0.70)	(1.01)	(0.70)	(0.61)	(0.57)

Table C2

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2006 – Clustered Standard Errors
Dataset: ADZ Updated 2013 State Data
Dropping States with Highest Residual Variance (Top 20%: MT, WV, NE, NH, ME, TN, ND, VT, HI, KY)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	0.02	10.68**	10.08***	14.05**	17.97**	10.70*	11.41**
	(5.68)	(5.17)	(3.28)	(6.85)	(7.77)	(5.38)	(4.37)
Spline Model:	0.15	0.72	1.44**	1.31	1.42	0.79	1.10*
	(0.83)	(0.71)	(0.67)	(1.02)	(0.86)	(0.69)	(0.61)
Hybrid Post-Passage Dummy:	-0.86	9.66**	4.62	10.55	15.01**	9.30*	8.34**
	(4.50)	(4.57)	(2.81)	(6.46)	(6.65)	(4.61)	(3.72)
Trend Effect:	0.19	0.22	1.20	0.77	0.65	0.31	0.67
	(0.81)	(0.67)	(0.72)	(1.05)	(0.72)	(0.64)	(0.60)

[58] Estimations include year and state fixed effects, and are weighted by state population. Robust standard errors are provided beneath point estimates in parentheses. The control variables for this "preferred" specification include: incarceration and police rates (lagged one year to avoid potential endogeneity issues), unemployment rate, poverty rate, population density, per capita income measures, and six demographic composition measures. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Appendix C (Cont.)

Table C3

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2006 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 State Data
Dropping States with Highest Residual Variance (Top 10%: MT, NH, WV, VT, NE)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	-0.10	-2.74	-0.54	2.65	9.41**	1.02	1.78
	(2.75)	(2.23)	(2.94)	(3.41)	(3.78)	(2.06)	(1.89)
Spline Model:	0.97	0.45	2.43**	0.56	-1.47	-0.06	-0.13
	(0.81)	(0.84)	(0.91)	(1.18)	(0.89)	(0.88)	(0.70)
Hybrid Post-Passage Dummy:	-0.53	-2.97	-1.63	2.43	10.18**	1.06	1.86
	(2.84)	(2.26)	(2.89)	(3.57)	(3.81)	(2.20)	(2.03)
Trend Effect:	0.98	0.53	2.47**	0.49	-1.76*	-0.09	-0.19
	(0.82)	(0.84)	(0.93)	(1.22)	(0.87)	(0.90)	(0.74)

Table C4

Estimated Impact of RTC Laws – ADZ Preferred Controls, 1977-2006 – Clustered Standard Errors and State Trends
Dataset: ADZ Updated 2013 State Data
Dropping States with Highest Residual Variance (Top 20%: MT, NH, WV, VT, NE, KY, ND, NV, DE, IN)

All figures reported in %	Murder	Rape	Aggravated Assault	Robbery	Auto Theft	Burglary	Larceny
Dummy Variable Model:	0.14	-1.82	0.60	3.25	9.86**	0.85	1.86
	(2.82)	(2.24)	(2.85)	(3.32)	(3.87)	(2.18)	(2.01)
Spline Model:	1.03	0.54	2.27**	0.80	-1.67*	-0.14	-0.07
	(0.88)	(0.89)	(0.93)	(1.28)	(0.94)	(0.97)	(0.75)
Hybrid Post-Passage Dummy:	-0.22	-2.02	-0.19	2.99	10.52**	0.91	1.89
	(2.88)	(2.24)	(2.73)	(3.46)	(3.92)	(2.29)	(2.13)
Trend Effect:	1.03	0.59	2.28**	0.74	-1.90**	-0.16	-0.11
	(0.89)	(0.89)	(0.95)	(1.32)	(0.93)	(0.99)	(0.79)

Appendix D – Summarizing Estimated Effects of RTC Laws Using Different Models, State v. County Data, and Different Time Periods

This appendix provides graphical depictions of 16 different estimates of the impact of RTC laws for the dummy and spline models for specific crimes using different data sets (state and county), time periods (through 2000, 2006, or 2010), and models (Lott and Mustard versus our preferred model and with and without state trends). For example, Figure D1 shows estimates of the impact on murder using the dummy model, designed to capture the average effect of RTC laws during the post-passage period. The first bar in each of the first six groupings corresponds to county-level estimates; the second bar corresponds to state-level estimates, for a total of 12 estimates per figure. Additionally, the last four estimates only contain one bar corresponding to state models run through 2010. The value of the figures is that they permit quick visual observation of the size and statistical significance of an array of estimates. Note, for example, that only one of the estimates of RTC laws on murder in either Figure D1 or Figure D2 is significant at even the .10 threshold. This sharp contrast to the conclusion drawn by James Q. Wilson on the NRC panel is in part driven by the fact that all of the estimates in this appendix come from regressions in which we adjusted the standard errors by clustering.

In contrast to the wholly insignificant estimates for murder, the estimates of the impact of RTC laws on aggravated assault in Figure D6 are generally significant as indicated by the shading of the columns, where again no shading indicates insignificance, and the shading darkens as significance increases (from a light grey indicating significance at the .10 level, slightly darker indicating significance at the .05 level, and black indicating significance at the .01 level). Note that the overall impression from Figure D6 is that RTC laws *increase* aggravated assault. Even in Figure D6, though, one can see that some of the estimates differ between county and state level data and tend to be strongest in state data controlling for state trends.

Figure D5, which provides estimates of the effect of RTC laws on aggravated assault using the dummy model (rather than the spline model of Figure D6), reveals that the conclusion that RTC laws *increase* aggravated assault is model dependent: if the dummy model is superior, and if we confine our attention to the complete 1977-2006 data set, the conclusion that RTC laws increase aggravated assault only holds for certain model specifications and not others.

Figure D1. Various Murder Estimates (Dummy Model)

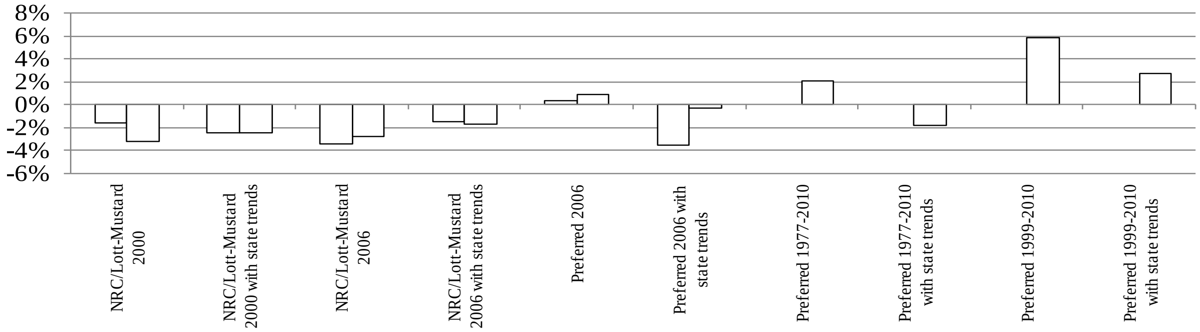


Figure D2. Various Murder Estimates (Spline Model)

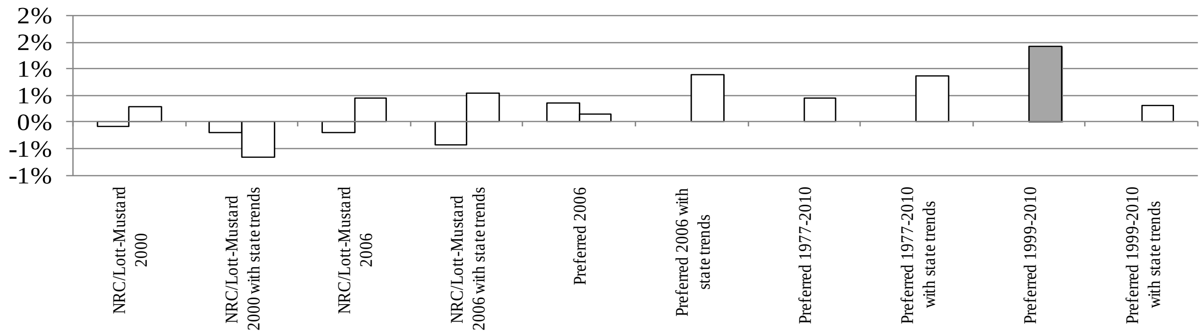


Figure D3. Various Rape Estimates (Dummy Model)

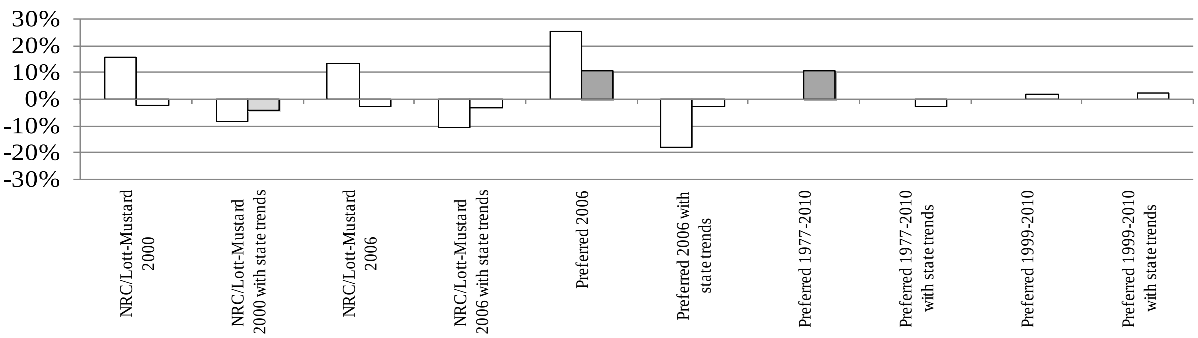


Figure D4. Various Rape Estimates (Spline Model)

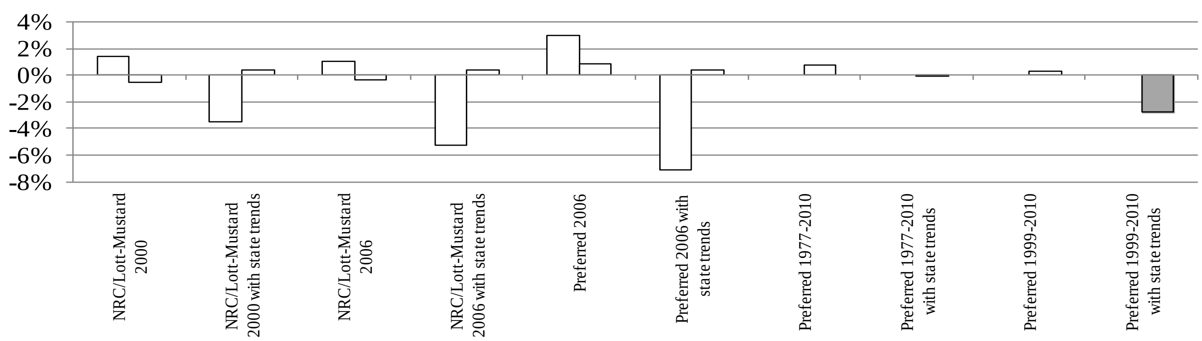


Figure D5. Various Assault Estimates (Dummy Model)

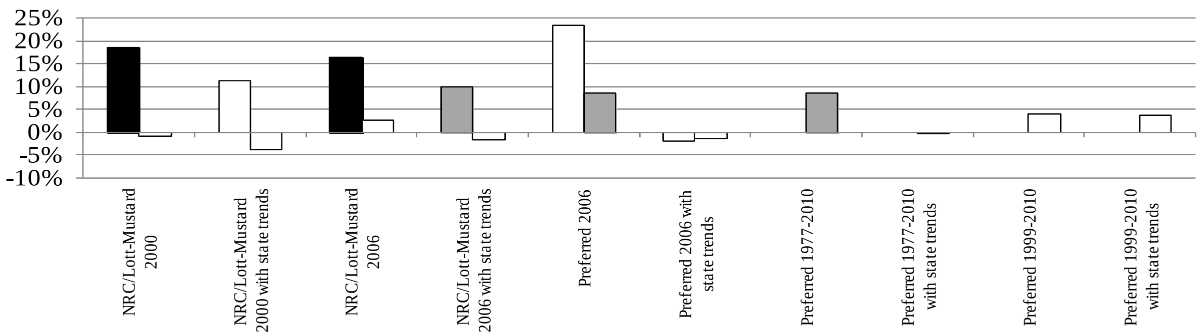


Figure D6. Various Assault Estimates (Spline Model)

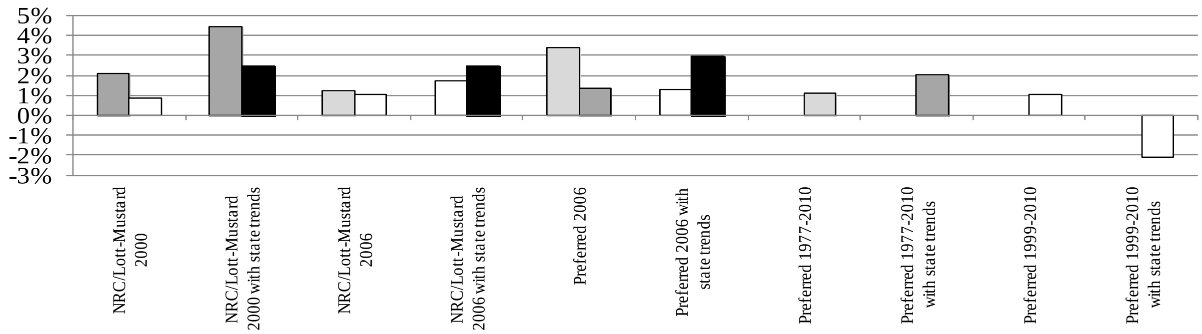


Figure D7. Various Robbery Estimates (Dummy Model)

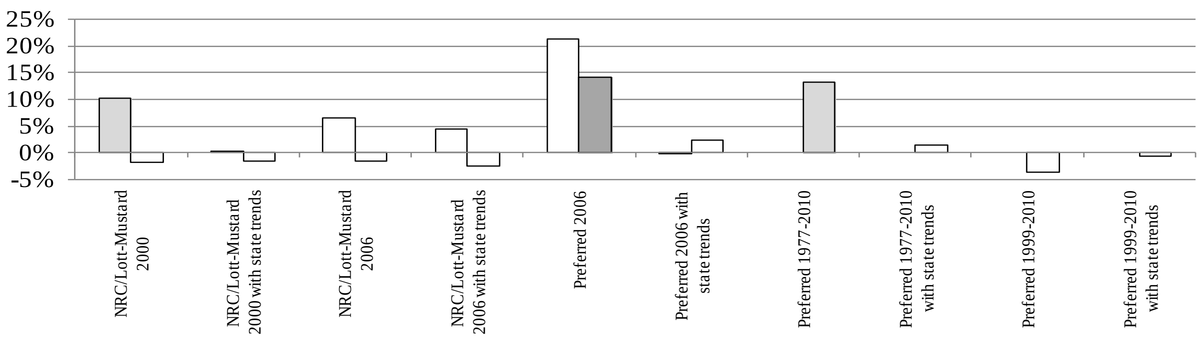


Figure D8. Various Robbery Estimates (Spline Model)

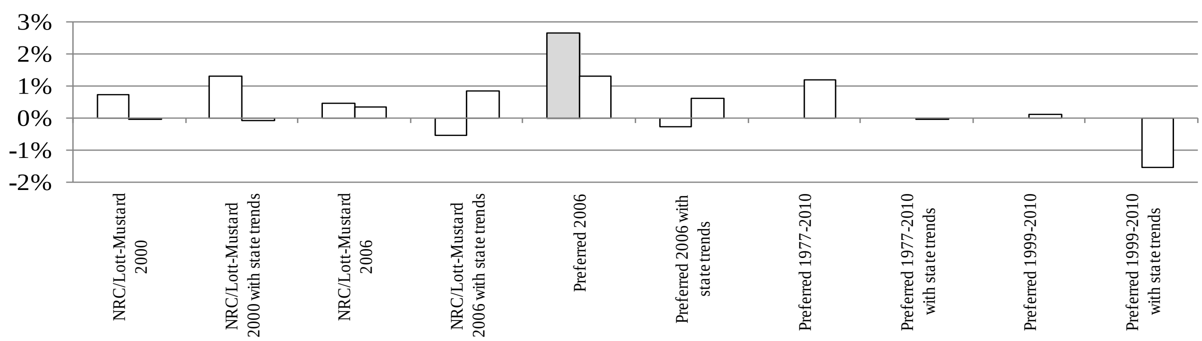


Figure D9. Various Auto Theft Estimates (Dummy Model)

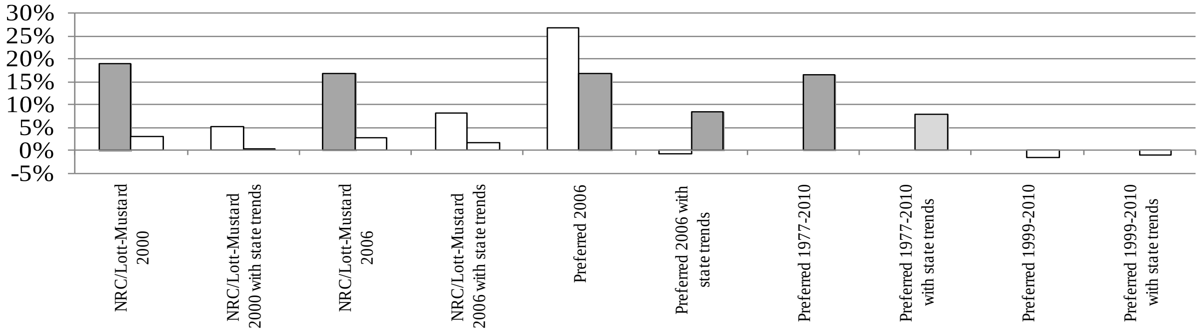


Figure D10. Various Auto Theft Estimates (Spline Model)

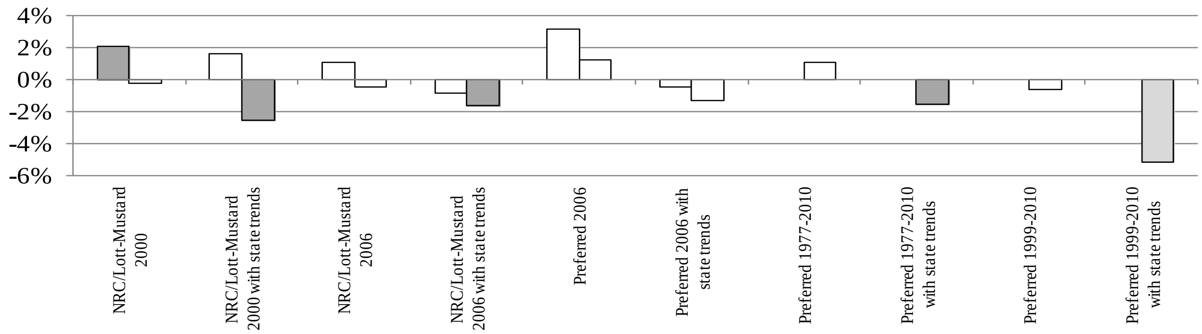


Figure D11. Various Burglary Estimates (Dummy Model)

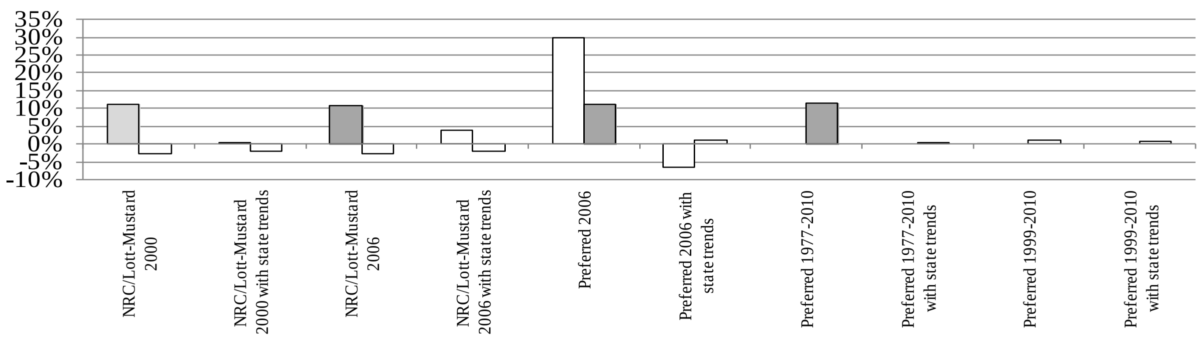


Figure D12. Various Burglary Estimates (Spline Model)

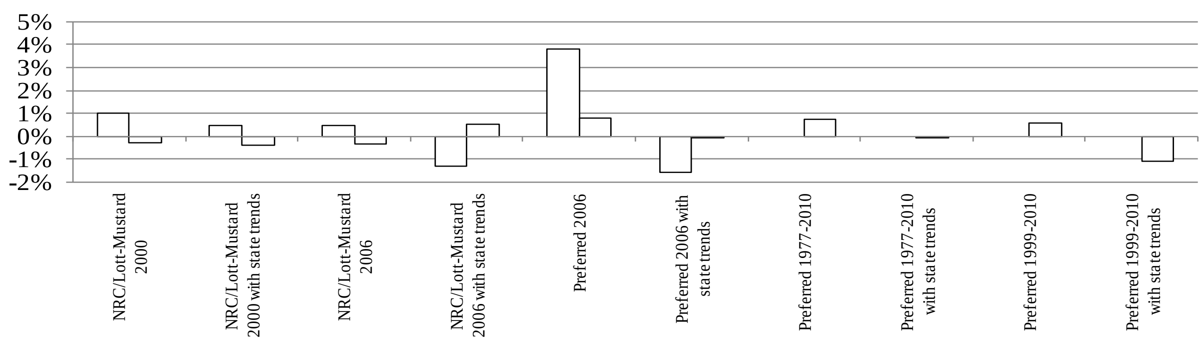


Figure D13. Various Larceny Estimates (Dummy Model)

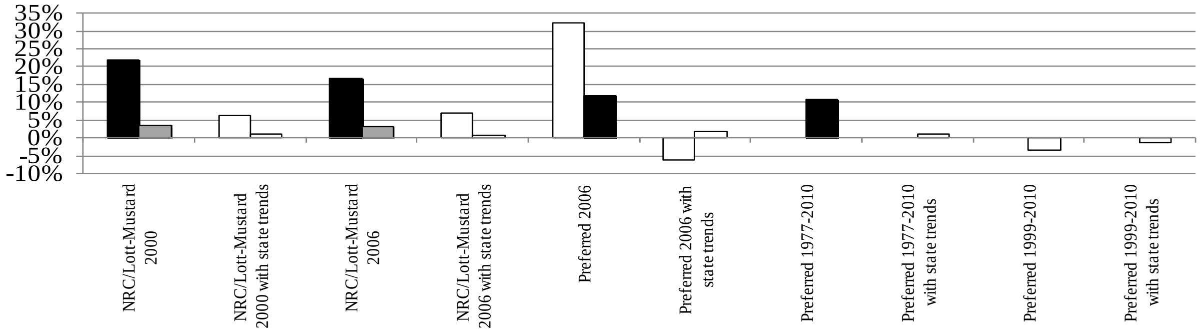
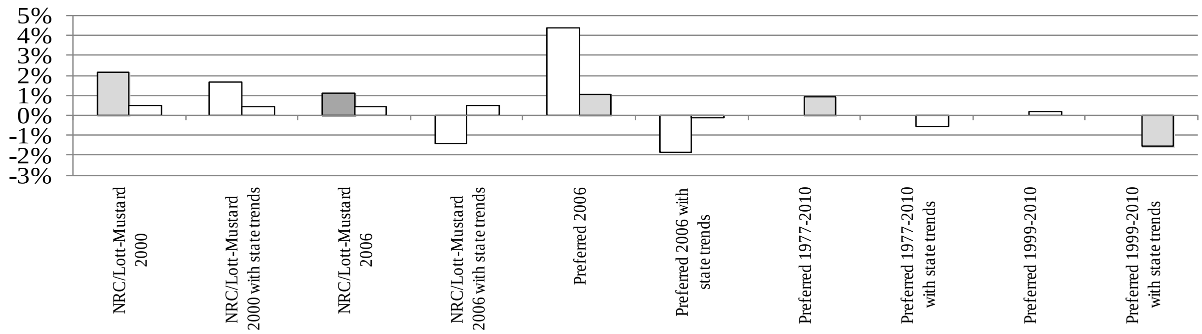


Figure D14. Various Larceny Estimates (Spline Model)



Appendix E – Methodological Description of Using Selection on the Observables to Assess Selection Bias

The analysis in Altonji *et al* (2005) is meant to test the presumption that there is omitted variable bias in a regression and quantify whether selection bias drives the OLS estimate. An underlying assumption of this model is that the observable controls are selected independently from the larger set of possible controls. In Elder and Jepsen (2013), which provides a useful description of the methodological features of the test, footnote 6 implies that the potential bias can be calculated with the given expression $\frac{cov(CS_i,\varepsilon_i)}{var(\widetilde{CS_i})} = \frac{cov(\widetilde{CS_i},\varepsilon_i)}{var(\widetilde{CS_i})}$, where CS is the potentially endogenous variable that corresponds to our RTC dummy variable.

This former expression, and that found in equation (3) of the same paper, can be combined to yield an expression for the potential bias: $\frac{cov(CS_i,X_i'\gamma)\cdot \text{var}(\varepsilon_i)}{var(\widetilde{CS_i})\cdot var(X_i'\gamma)}$. Here $\widetilde{CS_i}$ is given by the formula $CS_i = X_i\beta + \widetilde{CS_i}$ (that is $\widetilde{CS_i}$ is simply the residual from the regression of CS_i on $X_i\beta$). Putting this formula in terms of our RTC dummy variable gives the expression $\frac{cov(Shall_i,X_i'\gamma)\cdot \text{var}(\varepsilon_i)}{var(\widetilde{Shall_i})\cdot var(X_i'\gamma)}$. Because the beta coefficient of the bivariate regression of the RTC dummy on the fitted values of the regression of Y_i (murder rate) on our full set of controls (less the RTC dummy variable) amounts to $\frac{cov(Shall_i,X_i'\gamma)}{var(X_i'\gamma)}$, the only remaining variables needed are $\text{var}(\varepsilon_i)$ and $var(\widetilde{Shall_i})$. With this information one can calculate the "potential bias", which then is compared to the beta coefficients presented in the report.

The ratio of this implied bias to the estimate of the beta coefficient represents how strong selection on unobserved variables would have to be relative to selection on observed variables to attribute the entire estimated effect to selection bias. For the ADZ preferred specification (Table 8a), we find a beta coefficient of 0.0092, with a potential bias of -0.2894. This implied ratio is negative, implying that selection on observables and unobservables would have to be of opposite signs to be consistent with a true effect of zero. This finding implies that our slightly positive coefficient is a lower bound of the true effect of RTC laws on murder. Additionally, for the NRC regression (Table 1b), the report found a statistically significant beta coefficient on murder of -0.0833. However, when compared to the potential bias of -1.0304, it can be inferred that the OLS estimate would be solely driven by selection bias if selection on unobservables were only 8 percent as strong as selection on observables, which is strong evidence of omitted variable bias.