Model Uncertainty and the Effect of Shall-Issue Right-to-Carry Laws on Crime

by

Steven N. Durlauf, Salvador Navarro and David A. Rivers

Working Paper # 2014-4 **October 2014**

CIBC Working Paper Series

Department of Economics
Social Science Centre
Western University
London, Ontario, N6A 5C2
Canada

Model Uncertainty and the Effect of Shall-Issue Right-to-Carry Laws on Crime

Steven N. Durlauf†

Salvador Navarro‡

David A. Rivers‡

†University of Wisconsin at Madison

‡University of Western Ontario

October 15, 2014

Abstract

In this paper, we explore the role of model uncertainty in explaining the different findings in the literature regarding the effect of shall-issue right-to-carry concealed weapons laws on crime. In particular, we systematically examine how different modeling assumptions affect the results. We find little support for some widely used assumptions in the literature (e.g., population weights), but find that allowing for the effect of the law to be heterogeneous across both counties and over time is important for explaining the observed patterns of crime. In terms of model uncertainty, we find that there is substantial variation in the estimated effects for each model across all dimensions of the model space. This suggests that one should be cautious in using the results from any particular model to inform policy decisions.

Durlauf thanks the Institute for New Economic Thinking and Vilas Trust, and Navarro and Rivers thank the Social Sciences and Humanities Research Council, for financial support. We thank Maria Antonella Mancino, Joel Han, Nicholas Tenev, and Kegon Tan for outstanding research assistance. This paper is dedicated to the memory of Eduardo Ley.

1. Introduction

Over the last two decades, there has been a dramatic expansion in the ability of US citizens to carry concealed weapons. The US Supreme Court's 2008 District of Columbia vs. Heller decision established a personal right to carry arms. This decision made inevitable that all US states establish "may-issue" laws: laws which defined criteria under which a permit to issue concealed weapons may be obtained. A form of these laws, "shall-issue," which require permits to be issued unless the applicant falls into certain narrow categories, have been adopted by 41 states. 38 of the shall-issue states adopted the law prior to the Supreme Court decision. Shall-issue laws essentially legalize concealed carry for the vast bulk of a state's population.

A key part of the policy debates preceding the adoption of shall-issue laws has been the claim that legalization reduces certain crime rates. The theoretical argument is that potential criminals considering crimes such as a street robbery will be deterred based on beliefs about the likelihood that a potential victim is carrying a weapon. Studies of the effects of shall-issue laws have also examined other violent crimes, such as murder and rape. This has been done both because the deterrence logic of concealed carry is generally felt not to hold for these types of crimes, and also because of arguments that concealed carry may increase violent crimes by creating opportunities, e.g., during a disagreement between people who are intoxicated.

Empirical evidence for these claims was initiated in work by Lott and Mustard (1997); other prominent examples of work arguing in favor of a deterrent effect include Lott (2000). This body of work in turn generated a set of studies which concluded that claims of a deterrence effect were wrong and an artifice of special modeling assumptions; e.g., Black and Nagin (1998), Ludwig (1998), and Ayres and Donohue (2003). While the various participants in this literature have subsequently argued in favor of their assumptions and criticized those of others, these further arguments have not produced any resolution of the conflicting empirical conclusions, nor is there good reason to think that they can. The different assumptions are not grounded in principled behavioral arguments, i.e., substantive assumptions about economic theory; nor do they involve statistical distinctions for which there is a logical basis for preferring one approach to another.

This paper attempts to understand the conflicting claims in the concealed carry literature. We do this by systematically exploring how different modeling assumptions are linked to different findings on shall-issue's effect on crime. We do not, ex ante, privilege any particular model, as is standard practice in this literature. We are concerned with the relative evidentiary support for a given model, as well as the crime effect found through model averaging, but our primary objective is to explore the evidentiary support across a set of models and to understand which assumptions do and do not matter. Our analysis identifies a set of categories which span much of the existing concealed carry literature. These assumptions range from specification of the nature of the policy effect to choices of control variables to heteroskedasticity corrections to formulations of potential parameter heterogeneity to choices of instrumental variables. A major goal is to understand how these assumptions determine the disparate findings in the literature.

The closest antecedents to this paper are Cohen-Cole, Durlauf, Fagan, and Nagin (2009), Durlauf, Fu, and Navarro (2013), Durlauf, Navarro, and Rivers (2010), and Bartley and Cohen (1998). The first three papers attempted to adjudicate different claims about the deterrent effect of capital punishment and find that affirmative claims of deterrence were fragile in the sense that they required certain sets of modeling assumptions-assumptions which do not produce strong evidentiary support in comparison to other models.[1] In particular, evidence of deterrence required homogeneity in the state-specific deterrence parameter as well as the use of a linear probability model. Bartley and Cohen (1998) used Leamer's (1983) extreme bounds analysis to adjudicate differences between Lott and Mustard (1997) and Black and Nagin (1998) with respect to various choices of control variables. Our analysis considers a much more systematic set of papers and assumptions and focuses on the nature of the heterogeneity results rather than a robustness analysis per se. Further, we employ model averaging rather than extreme bounds analysis as the basis for

[1] See Strnad (2007) for a general argument for the importance of accounting for model uncertainty in empirical legal studies.

aggregating information across models.[2] Our work attempts to follow the spirit of Leamer (1978), who initiated systematic efforts to account for model uncertainty in evaluating empirical claims, although we do not focus on extreme bounds analysis per se.

The modern literature on model uncertainty has focused on the development of model averaging methods to facilitate statistical inferences which are not model-specific.[3] Eduardo Ley played a foundational role in this work, which initially focused on the identification of the determinants of heterogeneity in growth rates across countries. Ley's work, in combination with Sala-i-Martin, Doppelhofer, and Miller (2004), provided a constructive way to address model uncertainty which has seen wide application, and hence the dedication. We will employ model averaging methods to summarize our exploration of the model space. However, our main goal is to understand how assumptions determine conclusions on the effects of concealed carry laws.

As the concealed carry literature is based on observational data, the research may be subjected to the standard critiques that are made by advocates of randomized controlled trials and natural experiments as the gold standard for empirical social science. A National Academy of Sciences report, *Firearms and Violence* (Wellford, Pepper, and Petrie (2004)), expressed deep pessimism about the possibility that observational data would be able to resolve the effect of shall issue laws

> "…with the current evidence it is not possible to determine that there is a causal link between the right to carry laws and crime rates…It is also the committee's view that additional analysis along the lines of the current literature is unlikely to yield results that will persuasively demonstrate a causal link between right-to-carry laws and crime rates (unless substantial numbers of states were to adopt or repeal right-to-carry laws), because of the sensitivity of the results to model specification." (pg. 150-151)

[2]Brock, Durlauf, and West (2003) discuss why model averaging has better decision-theoretic foundations.
[3]Draper (1995) and Raftery, Madigan, and Hoeting (1997) are seminal contributions from statistics; see Doppelhofer (2008) for a concise overview of applications in economics.

This claim led one of the members of the panel, the eminent criminologist James Q. Wilson, to dissent from the report, which is extremely rare. Wilson's decision to dissent is primarily based on his objections to the report's claims that the concealed carry literature was uninformative. Specifically, he criticizes the report for making this conclusion based on the conflicting results on the effects of legalization generated by models with and without controls. Wilson argues that models without controls, even if they contradict other models, are intrinsically uninteresting.

> "Suppose Professor Jones wrote a paper saying that increasing the number of police in a city reduced the crime rate and Professor Smith wrote a rival paper saying that cities with few police officers have low crime rates. Suppose that neither Smith nor Jones used any control variables, such as income, unemployment, population density, or the frequency with which offenders are sent to prison in reaching their conclusions. *If* such papers were published, they would be rejected by the committee out of hand for the obvious reason they failed to produce a complete account of the factors that affect the crime rate." (pg. 270)

Wilson's dissent, in turn, invoked the rejoinder

> "…Everyone (including Wilson and the rest of the committee) agrees that control variables matter, but there is disagreement on the correct set. Thus the fact that there is no way to statistically test for the correct specification and that researchers using reasonable specifications find different answers are highly relevant. Given the existing data and methods, the rest of the committee sees little hope in resolving this fundamental statistical problem." (pg. 273-274)

As discussed in Durlauf, Navarro, and Rivers (2008), Wilson has the stronger side of the argument. With reference to model uncertainty, model averaging and related methods can constructively address the issues raised by the majority in the report. And of course one can just as easily criticize assumptions implicit in randomized trials. Our position, which follows the spirit of Heckman (2005), is that there is no single privileged type of empirical strategy in economics, which does not mean that methodological arguments are vacuous, but rather that judgments need to be exercised with respect to assumptions.

Our empirical findings may be summarized as follows. Some assumptions in the literature, in particular the use of population weights to control for heteroskedasticity in crime rates and the use of linear time trends, have so little evidentiary support that they are reasonably excluded from the analysis. For the remaining assumptions, the effects of the law on crime range across positive and negative values. With a partial exception involving the specification of the way that shall-issue law effects are measured, there is no simple one-to-one mapping between any one assumption and uniformly positive or negative set of estimates conditional on the assumption. Together, the full range of assumption disagreements matter for the disagreements in conclusions.

We also employ model averaging to provide a summary evaluation of the model-specific estimates that we study. Interestingly, the posterior probabilities in the model space are almost entirely concentrated on a single model for both property and violent crimes, although the models turn out to differ. The model-averaged point estimates imply a weak negative (deterrent) effect on crime in the short run and a positive and increasing effect on crime in the long run.

Together, we conclude that any strong claims about the effects of concealed carry legalization are fragile, at least based on the data and model space we examine. By the standard of Leamer's extreme bounds analysis, in which changes of sign across a model space are equated to fragility, there is no assumption that can avoid fragility per se, although one assumption gives most effect estimates the same sign. Model averaging reveals that posterior probabilities pile onto a small set of models all of which include this assumption. However, for these models, the year by year effects of the law change sign. Hence it would take very special priors and a very special loss function to use our findings as evidence in favor of concealed carry. We believe our findings indicate why the same may be said about the papers whose findings we have attempted to analyze.

2. Model uncertainty and policy effects: basic ideas

Our objective in this analysis is to explicitly incorporate model uncertainty into descriptions of the evidentiary support for a given empirical claim. The standard approach to policy evaluation, abstractly, is simply the construction of the posterior probability for outcome o conditional on policy p, data D, and a model m,

$$\Pr\left(o \mid D,p,m\right) \tag{1}$$

This distribution can be used to compute first and second moments $E\left(o \mid D,p,m\right)$ and $\operatorname{var}\left(o \mid D,p,m\right)$, as well as the posterior probability of the data, which provides a measure of goodness of fit:

$$\Pr\left(D \mid p,m\right) \tag{2}$$

Models may be understood as collections of assumptions which together provide a statistical specification (see Brock and Durlauf (2014) for elaboration). The new model uncertainty literature can be understood as relaxing the conditioning in equations (1) and (2) on a single model and providing new ways to report evidence on policy effects. Instead of a single model, one employs a set of candidate models, M; elements m of this set are differentiated by the assumptions from which they are constituted. From this perspective, the identity of the "true" model is one of the unobservables for which an analyst must account in drawing inferences. Further, the true model is itself of no intrinsic interest. The objective of a policy analysis is solely to evaluate the effect of the law on criminal behavior.

When assumptions are justified by economic or econometric theory, then it is of course appropriate to condition empirical conclusions on them. If such *a priori* information exists, then a researcher is justified in employing a dogmatic prior on the model space, i.e., assigning $\Pr\left(m\right)=1$ to some model. Similarly if some assumptions warrant a prior probability 1, then those models which contradict these assumptions should receive prior probability 0. However, for a context such as concealed carry and crime, the differences in modeling assumptions which have been used in competing analyses do not, as we argue below, admit resolution on an *a priori* basis. While arguments can be made in favor of or against particular assumptions, the models in the literature are all plausible candidates for the crime process.

This general critique of the use of a single model applies to the concealed carry literature at several levels. First, it means that there is no principled basis for functional form assumptions or for the choice of control variables that are found in a given model. Becker's (1968) deterrence model involves the marginal costs and benefits of a potential crime, but is mute on the forms of the cost and benefit functions nor does the theory say anything about their determinants. Choices of functional forms and control variables involve judgments. Second, uncertainty exists with respect to the validity of instruments. Unsurprisingly, a number of crime determinants are endogenous and so instruments are used to account for correlation between them and the unobserved heterogeneity in crime rates across localities. However, the absence of a principled basis for constructing the crime model means that the literature has relied on ad hoc informal arguments for the validity of particular instruments.

A third problem is that there is no reason to think that the parameters of a crime regression should be invariant to other characteristics (either measured or unmeasured) of the polities, i.e., it is not clear whether data taken across localities obeys the needed exchangeability assumptions which underlie analyses based on common coefficients. It is easy to identify reasons for parameter heterogeneity. For example, Southern states where people are taught at early ages how to handle weapons, and where guns are very much a part of the culture; and Northeastern states where guns are not common among regular citizens, can be expected to exhibit different coefficients in a crime regression that includes variables such as the conditional probability of arrest given crime commission, gun prevalence, or the presence of a concealed weapons carry law. The Lucas (1976) critique, further, can be argued to imply that the parameters in a regression of the type employed in the concealed carry literature will depend on the polity's criminal sanction regime.

All of these issues suggest that model uncertainty is likely to play a key role in crime regressions. In this respect, the literature on shall-issue carry laws is more of a demonstration of sensitivity of policy effect calculations to model assumptions than a body of work from which conclusions may be drawn. In fact, if the deterrent effects are estimated because of their use in policy evaluation, decision theory requires that model uncertainty be part of the appropriate loss function calculation.

Model averaging, when the loss function does not depend on the “true” model, provides a sufficient characterization of policy effects for policy comparisons if the policymaker is solving a standard statistical decision problem with uncertainty; see Brock, Durlauf, and West (2003) for a formal argument and elaboration. However, there are good reasons why the averaging the probabilities in (2) across models may not be sufficient for an analysis whose objective is to communicate evidence on policy effects. One reason concerns the specification of the priors on the model space. We regard this as a qualitatively different problem from the specification of priors for model-specific parameters. Why? While we have argued that there are not good reasons to possess dogmatic priors based on economic or statistical theory, at least for the context we study, this claim does not speak to the issue of how to construct priors for the model space.

Consider the case in which there is model uncertainty depending on whether one assumes rational or adaptive expectations in an environment. Different economists would have very different priors based on the totality of social science knowledge each possesses, or because each evaluates a common knowledge set differently in terms of its implications for which model is correct. While the concealed carry literature does not contain such stark disagreements on economic principles, authors have made principled arguments in favor of various assumptions, which means that priors over the model space would range widely.

The judgments which underlie an element of a model space reflect different views on assumptions, so while it be reasonable to ask a policymaker to relax a dogmatic model space prior, it is not reasonable to expect him to only be interested in results in which assumptions or their negation receive equal *a priori* weight. Second, the literature on model robustness (Hansen and Sargent (2007)) has given reasons to believe that preferences can exhibit ambiguity aversion with respect to model uncertainty. In such cases, the worst case scenario for a policy will depend on the identity of the true model, which in turn cannot be determined independently from the policymaker's preferences.

These considerations suggest that it is valuable to report information about the distribution of policy effects associated with (1) and (2). As initially suggested in Brock, Durlauf, and West (2007), policy dispersion plots, which visually represent the set of values of a policy effect across a model space, are useful in helping a policymaker to understand policy-effect/model dependence. Durlauf, Fu, and Navarro (2013) extend this idea by creating dispersion plots in which they organize models by classes of assumptions, allowing a reader to identify which assumptions do and do not matter for heterogeneity in model-specific policy effects. Hence, while model averaging is the best way to provide scalar statistics (e.g., moments) for a policy effect, we believe policy dispersion plots are useful as well.

Frequentist vs. Bayesian approaches

While our discussion of model uncertainty has employed Bayesian language, we will report frequentist estimates. When we engage in model averaging, this produces a "Bayes-Frequentist hybrid," a combination that first appeared in economics in Sala-i-Martin, Doppelhofer, and Miller (2004), who dubbed it Bayesian Averaging of Classical Estimates (BACE). We employ a hybrid in order to maintain comparability with the existing literature. While there are assumptions under which point estimates and standard errors can be equated to or will approximate posterior means and variances (see discussion in Raftery (1995) for example), our preferred interpretation is simply that we are providing information on the sensitivity of policy effects to model specification, with what we call model posteriors providing information on relative goodness of fit, adjusted for model complexity.

3. Model space

As we have emphasized, the lack of scholarly consensus over the effect of concealed carry laws on crime has centered on differences in the modeling choices employed by various researchers. Since there is essentially an infinite number of modeling choices available to a researcher, we employ the notion of exchangeability as a guiding principle. The basic idea is to consider whether the model selected by the researcher eliminates enough differences between (sets) of observations, e.g., Alameda county in California and Jackson county in Mississippi, that one can assume that the residuals are no longer a function of the identity of the particular observation. This is not to say that violating exchangeability invalidates a model. Instead one should consider whether the sources of non-exchangeability invalidate the purpose for which the model was built. Hence, when choosing a model space, we consider aspects like control variable choice, specification of the effect of the law, heterogeneity in the effect of the law both in the cross-section and over time, etc., that may lead to violations in exchangeability that are likely to affect our estimates of the effect of the law on crime.

The model space we consider in this paper takes into account some of the key sources of debate in the literature including modeling how the effect of the law varies over time, controlling for demographics, modeling time trends in crime rates, and dealing with the potential endogeneity of the laws. In addition, we also include some dimensions to the model space that we feel are particularly relevant to the issue, but that have received little or no attention in the literature. This includes the introduction of additional relevant control variables, allowing the parameters of interest to be heterogeneous, dealing with the potential endogeneity of arrest rates, and the use of population weights. Below we discuss each dimension of the model space in detail. In addition to identifying assumptions which differentiate existing papers, we also note some cases in which the literature seems to have ignored some natural variations of those specifications which have appeared so far.

i. formulation of shall issue laws as a crime determinant

Perhaps surprisingly, there has been disagreement in the literature as to the formulation of the impact of shall-issue laws in the statistical crime model. These disagreements depend on how one characterizes the dynamic effects of legalization. The first part of the literature identified concealed carry effects using a dummy variable to indicate whether the law is in place or not in a given location-time pair. This specification has a substantive implication, as it entails that the introduction of the law generates a permanent shift up or down in the crime rate. However, suppose that crime rates were increasing in a location prior to the implementation of the law. Suppose further that the effect of the law was to reverse this trend, generating a subsequent decreasing rate of crime. If one just measured crime before and after the introduction of the law, average crime rates would be similar, leading to the incorrect conclusion that the law had little or no effect on crime.

An alternative to the dummy variable model that seeks to account for this possibility, is to allow for different time trends in crime before and after the law. This approach has been used in John Lott's work (Lott and Mustard (1997), Lott (2000)), and has been referred to in the literature as the "spline" model (Ayres and Donohue (2003)). If we let $Y_{i,t}$ denote the log of the crime rate in county i at time t, and t_0 denote the time at which the law is implemented, a (very basic) spline model can be written as

$$Y_{i,t} = (t - t_0)\left[1(t_0 > t)\pi_b + 1(t_0 < t)\pi_a\right] + \epsilon_{i,t}. \tag{3}$$

Here π_b and π_a measure the trends before and after the law is implemented.

The spline model of equation (3) allows for pre- and post-law trends. However, it does not allow for the law to have any immediate impact on crime. As a response, a "hybrid" specification proposed by Ayres and Donohue (2003) incorporates both the trends of the spline model as well as the discrete shift allowed by the dummy model.

These two models are an attempt to account for differing temporal effects of the law in order to improve on the pure dummy model. However, they both have the feature that, as time elapses, the trend will keep pushing crime either up or down indefinitely. As we illustrate next, this feature can lead to paradoxical results.

Take the example in Figure 1 which shows a case in which crime rates are decreasing, then the law is implemented, the trend flattens and after a few years the trend flattens even more. The real effect of the law is that it worsens crime (since it was decreasing at a higher rate before the law was implemented) in the short run, and makes it even worse in the long run.

How will a hybrid trend specification behave relative to this hypothetical pattern? The pre-law trend will match exactly the real trend, but the post-law trend will overestimate the effect in the short run and underestimate it in the long run. More importantly, the dummy will show an immediate shift to a lower crime rate, i.e., it will show that the law decreased crime immediately, even though the only effect of the law is to worsen crime both in the short run and in the long run. The point we want to illustrate is that one should be very careful when interpreting estimates from these models, a task that gets even harder once the effect of the law is allowed to depend on other variables as well (as we do in the section on parameter heterogeneity below).

ii. control variables

Another important source of disagreement surrounds the selection of control variables to include in the empirical model.[4] These disagreements involve different types of variables.

It is common to include control variables for socio-economic status, age, and gender, in aggregated empirical analyses of crime. While their inclusion is, abstractly, justified by choice-theoretic models of criminal behavior such as Becker (1968), there are many potential measures of these underlying characteristics which are available for use, and theory does not provide guidance on how to proceed. For example, Lott includes a set of dummy variables interacting gender, race (white, black, and other), and age (divided into six age bins). In contrast, Ayres and Donohue (2003) argue that the number of such variables is excessive due to these variables being highly collinear and suggest including only the subset pertaining to black and white males, since these groups contribute to the vast majority of crime. They find that limiting the set of

[4]For a discussion of these issues related to measuring the effect of the death penalty on murder see Cohen-Cole, Durlauf, Fagan, and Nagin (2009), and Durlauf, Fu, and Navarro (2013).

demographic controls causes significant changes in the estimated effect of gun laws on crime, and in some instances leads to a change in the sign of the effect

In addition to differences in demographics across localities, there may be other factors related to location that are relevant for explaining crime rates. For example, attitudes towards gun safety, violence, or even gun usage in general may differ across various parts of the country. County-level fixed effects are one way to control for these differences, and are commonly included in empirical models. However, fixed effects do not capture any time-varying differences across counties. In order to address this, we consider two additional variables that vary across both county and time and are potentially related to both crime rates and gun control laws: gun prevalence and the level of urbanization.

If one believes that the legalization of concealed weapons has an effect on crime rates, then the same logic justifying this makes it plausible that the prevalence of guns in the population directly affects crime. Since the presence of shall-issue laws and the degree of gun prevalence might be correlated, one may want to include a measure of gun prevalence in the regression equation.

While gun prevalence is not a standard control variable in the literature, there are a few papers that include it. However, as is the case with demographic controls, there is no consensus on the effect of inclusion. Duggan (2001) estimates a positive correlation between gun prevalence and crime rates, but finds that controlling for gun prevalence does not affect the relationship between gun laws and crime. Moody and Marvel (2005) look only at the correlation between guns prevalence and crime, but find no systematic relationship. They attribute the difference between their results and Duggan’s to the choice of proxy variable for gun prevalence and differences in the number of control variables for age included in the regressions.

In addition to controlling for gun prevalence, we also expand the model space by introducing a measure of the level of urbanization in a county as a control variable. As is the case with gun prevalence, the degree to which a county is urban or rural may be related to both crime rates and the probability that a concealed-carry law is in place.

iii. time trends

Aggregate crime data often contain low frequency movements which are unexplained relative to the substantive theoretical commitments involved in a crime regression. As a result, there is controversy over how to model this type of unobserved heterogeneity, which involves the specification of time trends.[5] One option is to include time-specific dummy variables that control for changes in crimes rates that affect all localities equally. This allows for full flexibility in how crime rates evolve over time, but does not allow these effects to vary by location. An alternative option is to include region-time trends, which allow for differences across region, at the expense of less flexibility in the time dimension.

iv. parameter heterogeneity

A fourth source of model uncertainty is parameter heterogeneity. While parameter heterogeneity has not been systematically explored in the concealed carry literature, it has been shown to matter in other crime contexts, e.g., Shepherd (2004). One exception is Black and Nagin (1998) who argue that affirmative evidence that concealed carry reduces crime is sensitive to whether or not data from Florida are included in the analysis. Black and Nagin argue that this sensitivity reflects the effects of the Mariel boatlift on crime rates; concealed carry legalization in Florida occurs during the post-boatlift crime increase, whose reduction Black and Nagin attribute to the absorption of this population into Florida society and the identification of criminals who came in the boatlift, not concealed carry per se.

Black and Nagin’s argument should be understood in the context of whether crime observations from different US counties can be thought of as draws from a common model with fixed parameters.[6] Essentially, the issue is whether, after controlling for a large set of variables, one should assume that the effect of concealed

[5]See Durlauf, Navarro, and Rivers (2008) for further discussion.

[6]Brock and Durlauf (2001) provide an analysis of parameter heterogeneity as an example of an exchangeability violation and discuss how such violations can, if not accounted for, create spurious policy conclusions.

carry on crime is independent of the identity of the observation. For example, should one assume that the effect of the law is the same for county-time pairs where gun prevalence is very large as for county-time pairs in which there are few guns amongst the population?

The vast majority of the literature implicitly assumes that the effects of a policy, in this case shall-issue laws, are the same for all states and counties, in all time periods.[7,8] However, the response of crime rates to these laws may depend on a number of factors. We consider three factors that we believe could create heterogeneity in the effects of gun laws.

First, we allow for the laws to have differing effects depending on the level of urbanization in a county. Response to the laws may be very different in rural versus urban areas. For example, in rural areas, since people are more spread out, there may be fewer confrontations involving weapons. It may also be the case that in rural areas people know each other better, and therefore are more informed about who has a weapon and who does not, potentially altering the benefit of having a concealed weapon.

We also include a specification that allows for the effects of the laws to vary by region. Some regions may have differing attitudes regarding the use and safety of firearms, which could lead to different effects of the law.

Finally, we allow for the effects of the laws to depend on the prevalence of guns in a county. Since shall-issue laws allow individuals to carry concealed weapons, these laws could have a very different effect in areas with very few guns to begin with, compared to areas with many guns. For example, allowing people to carry concealed weapons in a place where gun prevalence is very low, might have a small effect on crime as very few people have guns. From the perspective of a criminal, the probability

[7]Examples of papers which allow for parameter heterogeneity in a crime context are Black and Nagin (1998), Shepherd (2005), Durlauf, Navarro, and Rivers (2010), and Durlauf, Fu, and Navarro (2013).

[8]The spline and hybrid models discussed above allow for the law to affect the linear trend in crime rates, but do not allow the effect to depend on calendar time, nor do they allow for non-linear trends.

of ending up in a confrontation with an armed victim is minimal, even if they are allowed to carry concealed weapons in principle.

v. instrumental variables

The concealed carry literature has attempted to address the endogeneity of the timing of legalization. Intuitively, relatively high (or low) crime rates may have determined which laws are passed, and not the other way around. We follow the convention in the literature and use the Republican vote share in the most recent presidential election to instrument for the laws.[9]

An additional concern that has received less attention in the literature is that the arrest rates for various crimes may also be endogenously determined. For example, if a county experiences an increase in the rate of a given crime, it may find it more difficult to find and arrest the perpetrators of those crimes with a given set of police resources. Alternatively, a county may respond to an increase in crime by allocating additional resources towards catching criminals. In either case, arrest rates would be endogenous.

In order to control for the endogeneity of arrest rates, we employ an instrumental variable strategy that is common in other literatures, particularly industrial organization, but to our knowledge has not been applied in crime contexts. We use the so-called Hausman-Nevo instruments (see Hausman (1995) and Nevo (2001)). These instruments exploit the panel structure of the data, and are constructed as a function (often the average) of the endogenous variable in other locations. In our context, we implement this in the following way. For each county-time pair, we compute the average arrest rate in all regions of the US in that period, excluding the region in which that county is located.

The identifying assumption behind this instrumental variable strategy is that arrest rates for a given type of crime are correlated across different counties (for instance, it may be easier to catch perpetrators of certain types of crimes), but the arrest

[9]The Republican vote share is used extensively in the literature as an instrument for shall-issue laws, e.g., Lott and Mustard (1997), Ayres and Donohue (2003), among many others.

rates in other regions of the US do not endogenously respond to variation in the local crime rate.

vi. population weights

Our final form of model uncertainty centers on population weights. The use of population weights has become standard practice in empirical crime studies. There are various reasons for why one may want to include weights, but the usual argument is based on concern regarding heteroskedasticity of the residuals. The intuition for wanting to weight aggregate observations (typically at the state or county level) is the following. Consider a simple linear probability model for the probability an individual *j* in county *i* in period *t* commits a crime $\rho_{j,i,t}$:

$$\rho_{j,i,t} = X'_{j,i,t}\beta + \eta_{j,i,t}, \tag{4}$$

where $X_{j,i,t}$ is a vector of observable characteristics, and $\eta_{j,i,t}$ is a mean-zero error term with variance σ^2. If we average this equation over individuals within each county and period, we obtain the following model for the county-time-specific crime rate:

$$P_{i,t} = \frac{\sum_{j=1}^{N_{i,t}} \rho_{j,i,t}}{N_{i,t}} = \overline{X'}_{i,t}\gamma + \varepsilon_{i,t}, \tag{5}$$

where $\overline{X}'_{i,t} = \frac{\sum_{j=1}^{N_{i,t}} X'_{j,i,t}}{N_{i,t}}$ and $N_{i,t}$ is the number of individuals in county c in period *t*.

Since the residual in this new equation is given by $\varepsilon_{i,t} = \frac{\sum_{j=1}^{N_{i,t}} \eta_{j,i,t}}{N_{i,t}}$, its variance will be given by $\frac{\sigma^2}{N_{i,t}}$. The aggregate residual, the one obtained by running county-level panel data regressions, is therefore heteroskedastic in proportion to the county population. Weighting by population makes the residual homoskedastic and generates the correct standard errors in a regression.

However, this argument ignores the possibility that location-time-specific unobservables are present, which we denote as $\psi_{i,t}$. With individual-level data, we could directly control for this form of heterogeneity using distinct location-specific, time-specific, and location-time-specific fixed effects. However, since we only observe county-level aggregate data, unobserved determinants of the crime rate that vary by location and over time, such as unmeasured demographic and socio-economic factors, as well as any other crime-related policy changes will enter the error term via $\psi_{i,t}$. Adding them to equation (5) gives us the following equation for the county-level crime rate:

$$P_{i,t} = \overline{X}'_{i,t}\gamma + \nu_{i,t}, \tag{6}$$

where $\nu_{i,t} = \psi_{i,t} + \varepsilon_{i,t}$. Since the average population in a county is over 100,000, and the variance of ε is equal to $\frac{\sigma^2}{N_{i,t}}$, the variance of ε will tend to be quite small. Therefore, unless the variation in county-level unobservables is extremely small in relation to variation in individual unobservable determinants of crime, the variance of error term in equation (6) will be dominated by the variance of $\psi_{i,t}$. Consequently, the use of population weights will overweight observations from more populous counties, leading to invalid confidence intervals, and potentially misleading point estimates.

Our model space consists of all possible combinations of the model elements described above. The size of the model space is then the product of the dimensions of each element: formulation of the law's effect on crime (3), demographic controls (2), additional covariates (3), time effects (2), parameter heterogeneity—none, gun prevalence (GP), urban, region, GP and urban, GP and region (6), IV/non IV (2), and population weights (2).

When we include interactions between the effect of shall-issue laws with gun prevalence, urbanization, and region we also separately include these variables as controls. As a result some models already contain the additional covariates (gun prevalence and urban dummies).[10] This leaves us with 624 distinct models, for each property and violent crime. In order to organize our discussion of the model space, we first define a baseline model and then fill out the model space by considering deviations from this baseline in each possible dimension. See Figure 2 for a characterization of the model space.

All models employ the log of the crime rate per 100,000 people as the dependent variable. Covariates which always appear are a measure of the shall-issue law, the arrest rate, socio-economic controls, age and race controls, time effects, and county fixed effects. All of the models under consideration are variations of this baseline specification:

$$
\begin{gathered}
\ln\left(\frac{\#Crimes_{i,t}}{Population_{i,t}/100{,}000}\right)= \\
\beta_1 LawDummy_{i,t}+\beta_2 ArrestRate_{i,t}+ \\
\gamma_1 PovertyRate_{i,t}+\gamma_2 UnemploymentRate_{i,t}+ \\
\gamma_3 Population_{i,t}+\gamma_4 RealPerCapitaPersonalIncome_{i,t}+ \\
\gamma_5 RealPerCapitaUnemploymentInsurance_{i,t}+ \\
\gamma_6 RealPerCapitaIncomeMaintenance_{i,t}+ \\
\gamma_7 RealPerCapitaRetirementPayments_{i,t}+ \\
\gamma_8 Age/RaceControls_{i,t}+\sum_t TimeEffects+ \\
\sum_i CountyEffects_i+\varepsilon_{i,t}
\end{gathered} \tag{7}
$$

[10]Region dummies are already implicit in every model given the presence of county-level fixed effects.

In the baseline model, both the shall-issue laws and the arrest rate are treated as endogenous. As we discussed above, we use the share of people who voted Republican in the most recent presidential election and the average arrest rate in the other regions of the county to instrument for each variable respectively.

4. Implementation

In this section we discuss issues related to estimating all the models contained in the model space described above.

While straightforward in principle, the calculations of some objects we use may be complicated. The posterior probability of the data given the model has the form

$$\Pr\left(D|m\right)=\int \Pr\left(D|m,\theta_m\right)\Pr\left(\theta_m|m\right)d\left(\theta_m\right) \qquad (8)$$

Integrals of this form, however, can be very hard to calculate numerically. To see why, consider a Monte Carlo estimator for (8) with diffuse priors on θ_m. In this case, most of the values for θ_m that one would sample to calculate the integral will have very small likelihood (especially if the posterior is concentrated relative to the prior) and the variance of the estimated integral will be large. As a consequence one would need to take an enormous number of draws from the prior in order to obtain precise estimates of the integral in equation (8), otherwise the estimates will be numerically unstable.

Closed form solutions and good approximations for some specific models with specific priors exist which avoid the explicit calculation of the right hand side of (8) (eg. Eicher, Lenkoskia and Raftery (2009)). These results, however, do not generalize to panel data models, mixed effects, and heteroskedasticity, and so cannot be applied to the statistical models which are used to evaluate concealed carry laws. In order to avoid these numerical instabilities, we employ the harmonic mean estimator of Newton and Raftery (1994),

$$\Pr^{-1}\left(D\middle|m\right)=\int\frac{1}{\Pr\left(D\middle|m,\theta_m\right)}d\theta_m \qquad (9)$$

Since this estimator is based on samples from the posterior distribution, instead of from the prior, it avoids the problem of having most of the draws coming from low (or zero) likelihood regions of the support of θ_m.

Estimation is performed via a MCMC Gibbs sampler. The most general version of the class of models we estimate is an instrumental variables model with county fixed effects and population weights. Since all of the models we consider are special cases of this, we only discuss how to implement the general version that includes all of these features.

Let $i=1,...,I$ index counties and $t=1,...,T$ index time. The general version of the model can be written as

$$Y_{i,t}=X'_{i,t}\beta+W'_{i,t}\alpha+\psi_i+\epsilon_{i,t} \qquad (10)$$

where the $j=1,...,J$ elements of $X_{i,t}$ are potentially correlated with $\epsilon_{i,t}$ (i.e., endogenous) while the $h=1,...,H$ elements of $W_{i,t}$ are assumed exogenous. When we estimate models where the law and the arrest rate are assumed exogenous, we simply make $X_{i,t}$ a subset of $W_{i,t}$. We also assume access to a set of $k=1,...,K$ instruments $Z_{i,t}$ such that

$$X_{i,t,j}=Z'_{i,t}\gamma_j+W'_{i,t}\lambda_j+\zeta_{i,j}+\upsilon_{i,t,j} \qquad (11)$$

for all j, with both $Z_{i,t}$ and $W_{i,t}$ assumed to be independent of $\upsilon_{i,t,j}$.

Let $Q_{i,t}$ denote the population of county i at time t, and let $q_{i,t} = \sqrt{Q_{i,t}}$.[11] To allow for heteroskedastic errors, we assume that $\epsilon_{i,t} \sim N\left(0,\frac{\tau_\varepsilon}{Q_{i,t}}\right), \upsilon_{i,t,j} \sim N\left(0,\frac{\tau_{v,j}}{Q_{i,t}}\right)$, jointly independent from ψ,ζ. This allows us to define new random variables $\varepsilon_{i,t} = q_{i,t}\epsilon_{i,t} \sim N(0,\tau_\varepsilon)$ and $v_{i,t} = q_{i,t}\upsilon_{i,t} \sim N(0,\tau_v)$ such that

$$\begin{aligned} Y_{i,t} &= X'_{i,t}\beta + W'_{i,t}\alpha + \frac{\varepsilon_{i,t}}{q_{i,t}} \\ X_{i,t,j} &= Z'_{i,t}\gamma_j + W'_{i,t}\lambda_j + \frac{v_{i,t}}{q_{i,t}} \end{aligned} \tag{12}$$

For the moment we do not specify anything about the dependence of ψ,ζ on either X,W and/or Z.

To construct the Gibbs sampler, we employ an Empirical Bayes approach in which we first estimate the parameters γ,λ and fix them throughout the rest of the analysis. The full details of the algorithm we employ can be found in Durlauf, Navarro and Rivers (2014); here we sketch the basic ideas.

The first step in estimation of the model involves the county fixed effects. Traditional Bayesian approaches deal with fixed effects using a “mixed” effect strategy in which the distribution of the effect is parameterized, allowing it to depend on moments of the distribution of X,W,Z. This approach introduces many additional parameters to estimate, which increases the computational burden, and may lead to imprecision in estimates of the parameters of interest. Furthermore, there is no basis for imposing any particular dependence structure between X,W,Z and ψ,ζ, or making particular distributional assumptions on ψ,ζ. As we discuss below, adopting traditional methods to remove the fixed effects (i.e., within or first differencing), present both implementational and empirical challenges that make them unappealing in our context.

[11]For models that do not include population weights, $Q_{i,t}$ may be set equal to 1 without loss of generality.

Instead, we introduce a new procedure, which we call "random first differencing", that eliminates the fixed effects by differencing the data with respect to a fixed, but randomly chosen period. To see how this procedure works, let r_i denote a random period in which county i is observed in the data and, for any random variable V, let $V_{i,t}^{\Delta} = V_{i,t} - V_{i,r_i}$. By taking differences with respect to the randomly chosen period r_i, we have

$$\begin{aligned} Y_{i,t}^{\Delta} &= X_{i,t}^{'\Delta}\beta + W_{i,t}^{'\Delta}\alpha - \frac{\varepsilon_{i,r_i}}{q_{i,r_i}} + \frac{\varepsilon_{i,t}}{q_{i,t}} \\ X_{i,t,j}^{\Delta} &= Z_{i,t}^{'\Delta}\gamma_j + W_{i,t}^{'\Delta}\lambda_j - \frac{v_{i,r_i}}{q_{i,r_i}} + \frac{v_{i,t}}{q_{i,t}} \end{aligned} \tag{13}$$

The main advantage of this "random first differencing" procedure is that, if we complete the data and condition on ε_{i,r_i} (and the observed q_i), then $\frac{\varepsilon_{i,t}}{q_{i,t}} - \frac{\varepsilon_{i,r_i}}{q_{i,r_i}}$ will be independent over time, and the same is true for v_j.

By comparison, using the traditional within differencing approach leads to a complicated joint distribution of the errors over time. This substantially increases the computational burden of the algorithm compared to the conditionally *i.i.d.* data generated by our procedure.

Random first differencing also solves the problem associated with just first-differencing the data when you have a dummy variable that switches only once per county (as we do for the case of the law). In this case, the only identifying power in the data is given by that one period switch in the dummy, causing the researcher to rely on a very small portion of the variation in the data. Our procedure also overcomes the difficulties associated with differencing with respect to a fixed period, in which case the results can be sensitive to the period that is chosen.[12]

[12]By doing random first differencing, we are effectively integrating against the distribution of the data, which leads to something that resembles the within differences fixed effect type estimator.

In order to deal with the potential heteroskedasticity associated with population weights, we simply pre-multiply the random first differenced data by $q_{i,t}$. Letting $V_{i,t}^{\Delta,q} = q_{i,t}\left(V_{i,t} - V_{i,r_i}\right)$ for any random variable V, the model is given by

$$\begin{aligned} Y_{i,t}^{\Delta,q} &= X_{i,t}^{'\Delta,q}\beta + W_{i,t}^{'\Delta,q}\alpha - \frac{q_{i,t}}{q_{i,r_i}}\varepsilon_{i,r_i} + \varepsilon_{i,t} \\ X_{i,t,j}^{\Delta,q} &= Z_{i,t}^{'\Delta,q}\gamma_j + W_{i,t}^{'\Delta,q}\lambda_j - \frac{q_{i,t}}{q_{i,r_i}}v_{i,r_i} + v_{i,t} \end{aligned} . \tag{14}$$

The remaining step for implementation is the specification of priors for the parameters. For the regression coefficients β,α we specify independent uniform priors $U\left(-a,a\right)$ for a "large" (we employ $a = 500$). In order to specify priors for the distributions of the residuals, notice that one can rewrite the crime equation as

$$\begin{gathered} Y_{i,t}^{\Delta,q} = \\ Z_{i,t}^{'\Delta,q}\sum_{j=1}^{J}\gamma_j\beta_j + W_{i,t}^{'\Delta,q}\left(\sum_{j=1}^{J}\lambda_j\beta_j + \alpha\right) - \frac{q_{i,t}}{q_{i,r_i}}\left(\sum_{j=1}^{J}v_{i,r_i}\beta_j + \varepsilon_{i,r_i}\right) + \sum_{j=1}^{J}v_{i,t}\beta_j + \varepsilon_{i,t} \end{gathered} . \tag{15}$$

Let $v_{i,t,J+1} = \sum_{j=1}^{J} v_{i,t}\beta_j + \varepsilon_{i,t}$ and define $v_{i,t} = \left(v_{i,t,1},\ldots,v_{i,t,J},v_{i,t,J+1}\right)'$ as the $J+1\times 1$ vector of SUR residuals for county i at time t. We assume $v_{i,t} \sim N\left(0,\Pi^{-1}\right)$, where Π is the precision matrix of the SUR equations.[13] Finally, for Π we assume a proper Wishart prior, $\Pi \sim W\left(b,V\right)$, set $b = J+1$ (i.e., the least informative proper Wishart prior) and set V to be an identity matrix divided by the number of observations.

The key to deriving our Gibbs sampling algorithm is that, conditional on the random variable pair $\varepsilon_{i,r_i}, v_{i,r_i}$, equation (14) is a standard instrumental variables model.

[13]Notice that the normality assumptions are in no way crucial since they can be replaced by more flexible distributions like mixtures of normals.

The only difference is that we then have to derive the posterior distribution for $\varepsilon_{i,r_i}, v_{i,r_i}$. Details of the Gibbs sampling algorithm are given in Durlauf, Navarro and Rivers (2014).

Finally, we discuss how we calculate the marginal likelihood of the data. Calculating the marginal likelihood using the harmonic mean estimator in equation (9), requires us to evaluate the conditional likelihoods for each model. Deriving the conditional likelihood for non-IV models is straightforward. Let $\epsilon_i^{\Delta} = \left(\epsilon_{i,1}^{\Delta},...,\epsilon_{i,T_i}^{\Delta}\right)'$. In this case,

$$\Pr\left(D \mid m,\theta_m\right) = \prod_i f_{\epsilon_i^{\Delta}}\left(Y_i^{\Delta} - W_i^{'\Delta}\alpha\right), \tag{16}$$

where $f_{\epsilon_i^{\Delta}}$ is the T_i dimensional multivariate normal density defined by ϵ_i^{Δ}.

Since our focus is on comparing models based on how well they explain crime rates, for IV models we compute the conditional likelihood based only on the partial likelihood of the outcome equation (10), taking first random differences in order to eliminate the county fixed effects. Let $\omega_{i,j}^{\Delta} = \left(v_{i,1,j}^{\Delta},...,v_{i,T_i,j}^{\Delta}\right)'$, the T_i dimensional vector of residuals from the j^{th} first stage regression, and let $\omega_i^{\Delta} = \left(\omega_{i,1}^{\Delta},...,\omega_{i,J}^{\Delta}\right)'$. The partial likelihood is then given by:

$$\Pr\left(Y_{i,t}^{\Delta} \mid X_{i,t}^{\Delta},W_{i,t}^{\Delta},Z_{i,t}^{\Delta},m,\theta_m\right) = \prod_i f_{\epsilon_i^{\Delta}|\omega_i^{\Delta}}\left(Y_i^{\Delta} - X_i^{'\Delta}\beta - W_i^{'\Delta}\alpha\right). \tag{17}$$

5. Data

The main source of data for this paper is a county-level panel dataset covering 1979-2000, and various versions have been used extensively in previous studies on the effect of shall-issue laws on crime.[14,15] The dataset consists of information on crime rates and arrest rates for various crime categories that comes from the FBI's Uniform Crime Reports. This is supplemented with data on the presence of shall-issue laws in each state. The dataset also contains information from the U.S. Census Bureau on a wide set of socio-economic and demographic controls.

In order to expand the model space, we add some additional variables to the dataset. First, we introduce a measure of gun prevalence used previously by Moody and Marvell (2005). Moody and Marvell use the data collected by Duggan (2001) from the General Social Survey (GSS) of the National Opinion Research Center (NORC), which includes questions related to gun ownership. According to Moody and Marvell, this is the only direct measure of gun ownership at the state level.

One drawback of the survey is that it only covers 3000 people, and it is not asked in every state in every year. In order to deal with this, Moody and Marvell show that the percentage of suicides that were committed using guns is a good proxy for gun ownership as measured by the survey. Therefore, we collected data on the percentage of gun suicides from the Centers for Disease Control and Prevention (http://wonder.cdc.gov/).[16] We then scaled this percentage by the correlation between gun ownership and gun suicides found by Moody and Marvell, and use this as our measure of gun prevalence in our empirical analysis.

We also collected data from the U.S. Census Bureau that measures the percentage of people living in urban areas. We define a county as urban if more than 70% of the population live in urban areas and define it as rural if less than 30% live in urban areas. Our results are robust to alternative choices of cutoffs and to the use of a continuous measure of urbanization.

[14] Papers using this dataset include Lott and Mustard (1997), Bartley and Cohen (1998), Black and Nagin (1998), Dezhbakhsh and Rubin (1998), Lott (2000), Plassmann and Tideman (2001), Ayres and Donohue (2003, 2009a, 2009b), Moody and Marvell (2008, 2009), Aneja, Donohue, and Zhang (2011), among many others.

[15] The data set we employ was downloaded from John Lott's website http://www.johnlott.org in January 2013.

[16] These data are available at both the county and state levels. However, due to confidentiality concerns, this data is not reported if the total number of suicides, or number of suicides by guns, is too low. As a result, data are missing for a large number of county-level observations. Therefore, we use the state-level data in our analysis.

The original dataset includes a measure of the percentage of the population that voted Republican in the most recent Presidential election. However, this variable only goes up to 1998, so we updated this series to 2000 using data collected from the Atlas of US Presidential Elections (http://www.uselectionatlas.org).

6. Defining effects

Since the dummy, spline, and hybrid specifications differ in how the law enters the model, and because we have some specifications in which the effects of the law vary across observations, there is no single parameter that can be used to compare the results of the different models. In order to compare the results across models, we compare the crime rates for each observation when we allow the law to be in effect versus when it is not. We call this the marginal effect of the law. We focus on average marginal effects (AME) of the law, where we average the effect over all states in 1998. This means that we are calculating the change in the crime rate which would have occurred if all states had legalized shall issue in 1998. We compute the effect of the law at implementation, as well as one and two years afterwards. We also calculate the average across all three periods. The primary reason for this is that the spline and hybrid models allow the effect of the law to vary over time.[17] In order to capture this, we measure the effect of the law over this three-year window, allowing the values of controls to evolve over the same three-year window.

Other types of policy effects can be calculated. One alternative, which we call AME-Law measures the average effect of the law for *only* those states which had actually implemented shall issue by 1998, using a three-year window for crime rates

[17]The average marginal effects also vary over time when we allow the effect of the law to depend on certain characteristics of the county (e.g., gun prevalence and the level of urbanization). While the levels of gun prevalence and urbanization do change over time, these changes are typically small.

and controls.[18] As a complement to this, we also compute AME-No Law, which measures the effect for states that *did not* implement the law during our sample period. These two measures allow for the possibility that states which implemented the law and those who did not were aware of the heterogeneity in its effects. Our main measure of the effect of the law, AME, is a weighted average of AME-Law and AME-No Law.

In total this gives us twelve different average marginal effect estimates for each model.[19] We compute each of these marginal effect measures for all models in the model space. In order to focus our discussion, we primarily use the version that measures the effect for all states (AME), and we use the average of this measure across all three years. We discuss the cases in which there are important differences between the various measures and their corresponding relevance for gun control policy.

7. Preliminary analysis and identification of a policy-relevant model space

A standard issue in exploring model spaces concerns the high dimension of the space. Madigan and Raftery (1994) propose a way of decreasing a set of candidate models by eliminating highly improbable ones via a procedure they call "Occam's Window". The basic idea of the procedure is to set a threshold for a model's posterior probability and to then eliminate models with lower probabilities than this threshold. For the context of understanding the role of assumptions in outcomes, we follow the spirit of the "Occam's Window" procedure by checking whether certain assumptions are associated with qualitatively lower likelihoods than their (contradictory) counterparts, and so resolve such cases based on goodness of fit. We start with this evaluation in

[18]We also computed the average marginal effect for only states that implemented the law, evaluated when the law was implemented in each state, instead of in 1998. The results were very similar.

[19]We compute each of the three AME measures for each of the three years following the introduction of the law, as well as an average over these three years, for a total of four estimates per AME measure.

order to allow subsequent analyses to focus on assumptions which possess non-negligible posterior weights. Specifically, we identify assumptions such that the model with the highest likelihood from those maintaining the assumption still has a lower likelihood that the model with the lowest likelihood among models which do not maintain it. There are two assumptions whose associated models are dominated by the posterior probabilities found in the rest of the model space.

The first assumption that is associated with dominated models is the use of population weights to control for heteroskedasticity. This finding suggests that the common practice of adding population weights to regressions using county-level or state-level data is not supported by the data. As discussed in Section 3, when one considers the presence of location-time-specific unobservables in the model, the standard argument for using population weights is weakened, and may in fact introduce heteroskedasticity, instead of correcting for it. Further, models with weights tend to overstate the effect of the laws on the crime rate compared to the models without weights. In other words, the models with weights are more likely to find that the laws lead to an increase in crime. As a result, the standard use of population weights in the literature may be leading to an upward bias in the estimate of the effect of the laws on crime in light of the lack of support for the assumption.

The second assumption associated with dominated likelihoods is the use of region-trends to explain the time-series patterns of crime rates (compared to models that use time dummies). This finding indicates that the time-series patterns in crime rates do not follow a linear trend and are better approximated by flexible time dummies, even at the expense of not allowing the time-series pattern to be region-specific.

The model space described in Figure 2 generates 624 models. Half of this set consists of models with population weights, and half of the remaining set consists of models which include region trends. After eliminating these dominated models, 156 models remain for analysis.[20] This means that we now examine model uncertainty with

[20]The product of the dimensions of each remaining element is 216: formulation of the law's effect on crime (3), demographic controls (2), additional covariates (3), parameter heterogeneity (6), and IV/non IV (2). However, as discussed above, when we include interactions between the effect of shall-issue laws with gun prevalence, urbanization, and region we also separately include these variables as controls. As a result some

respect to control variables for demographics and location-specific heterogeneity, parameter heterogeneity, and instrumental variables.[21]

Figure 3 presents dispersion plots which graphically represent the heterogeneity in average marginal effects (AME) that appears as one moves across the model space. Over a third of the models produce a positive effect of the law on crime, i.e., shall issue laws increase crime, whereas the remaining models suggest a negative effect. Magnitudes also vary considerably. For violent crime, the estimates range from the law implying a decrease in crime of 15% to an increase of 18%, and for property crime ranging from a decrease of 15% to an increase of 28%. Based on these results, it is easy to see how the concealed carry literature contains such disparate claims. Depending on the model chosen by the researcher, the results can vary from a large decrease to a large increase in the crime rate. We now examine the sources of this heterogeneity in estimated policy effects.

8. Assumptions and policy effects

In Table 1 we report the average marginal effects (in particular, the three-year average of AME) averaged across each specification within each dimension of the model space. For example, with regard to the formulation of the law dimension, there are three specifications: “dummy”, “spline”, and “hybrid”. For each of these three specifications, we take all models with that specification (e.g., spline) and compute the simple unweighted average of the estimated AME across all of those models. Table 1 illustrates how moving within each dimension of the model space affects the estimated effect of the law, on average. For example, in the bottom right corner of Table 1, we can see that for violent crime, models which correct for endogeneity (IV) on average suggest an increase in crime of 1.4%, whereas models that do not (Non IV) suggest on average a reduction of crime by 3.9%.

---

models already contain the additional covariates (gun prevalence and urban dummies). This leaves us with 156 remaining models.

[21] In Table A in the Appendix, we present the marginal likelihood values for the non-dominated models for both violent and property crime.

As a complement to Table 1, in Figures 4A-4F we present the average marginal effects (three-year average of AME) for all models. These figures are the same as Figures 3A and 3B, except that the models are grouped according to each model dimension. For example, Figures 4A and 4B group the models by the formulation of the law (i.e., dummy, spline, hybrid). These figures illustrate the degree of heterogeneity both across and within each formulation of the law.

i. formulation of shall issue laws as a crime determinant

For both violent and property crime, there is a clear difference in the estimated average marginal effects between the dummy, spline, and hybrid models. As can be seen in Table 1, on average, the dummy specifications generate positive effects of the law, the hybrid generates negative effects, and the spline is in between. Figures 4A and 4B illustrate that even within these specifications there is a substantial amount of heterogeneity. Figure 4A shows that, for violent crime, the dummy model generates the largest effects in absolute value, with some versions producing large negative effects of the law, and others large positive effects. For property crime, Figure 4B illustrates that the average difference between the three specifications in Table 1 is generated primarily by the dummy model having positive effects of the law in almost all specifications, whereas the spline and dummy model produced almost entirely negative effects, but the magnitudes are larger in absolute value for the hybrid models.[22]

In every case, the hybrid or spline specification is preferred to the dummy model (has a higher marginal likelihood value, see table A in the Appendix), and in most cases the hybrid model has a higher marginal likelihood than the spline model. This is evidence that the effect of the law not only affects crime upon implementation (via the dummy portion of the hybrid model), but also changes the trend in the crime rate (via the spline portion).

Together these results suggest that the effect of legalization of concealed carry is heterogeneous over time, and that the use of a single dummy variable is insufficient to

[22]Interestingly, the paper which introduced the hybrid model, Ayres and Donohue (2003), argued against a reduction in crime due to shall issue laws.

capture the effect of the law. The hybrid and spline are reduced form ways of allowing for this but, as illustrated in Figure 1 and its associated discussion, more work needs to be done to investigate the source and implications of this heterogeneity.[23]

ii. control variables

We next examine the effects of alternative control variable choices. The data seem to prefer the specifications with the full set of age and race dummies (36 dummies) to the black/white subset (12 dummies), with the likelihood for the models with a full set of dummies being higher in each case. There is no systematic pattern, however, in the estimated average marginal effects, and the average difference is very small for both types of crime (less than 1%). So while the full set of demographic variables provides a better fit to the data (see Table A in the Appendix), both versions provide similar answers to the question of the effect of shall-issue laws on crime.

For the variables gun prevalence and urbanization, it appears that each is useful for helping to explain crime patterns, as they lead to higher marginal likelihood values but, as was the case with demographic variables, they generate only small changes in the average marginal effects. As we discuss below, the more important impact of these additional variables is through allowing the effect of the law to depend on gun prevalence and urbanization by interacting them with the law variables.

iii. parameter heterogeneity

As illustrated in Figures 4C and 4D, when one allows for the effects of the law to be heterogeneous, there is still a lot of heterogeneity within each specification. On average, the models that allow for interactions with urbanization generate the highest AME's, and those with region interactions generate the lowest AME's.

[23]Black and Nagin (1998) interact the law with a set of five lead and lag dummies, allowing the effect of the law to vary over the 11-year period centered around the enactment of the law, however they fail to find any statistically significant effects of the law on crime.

As shown in Table A-1, for violent crime there is a clear pattern in the marginal likelihoods, and the models that include interactions of the law and region are preferred to the other models, with the models that also include interactions of the law and gun prevalence being the most preferred specifications. These two sets of specifications on average produce lower effects of the law compared to the other models (small negative effects versus effects close to zero). However, as Figure 4C illustrates, the numbers in Table 1 mask a large degree of heterogeneity within each group.

With property crime, the pattern is less clear. The models with interactions are universally preferred to the models without any parameter heterogeneity. In some cases the models that include interactions with urban dummies and gun prevalence are preferred and in a smaller set of cases, the models with region dummy and gun prevalence interactions are preferred. As was the case with violent crime, the specifications with region interactions and region plus gun prevalence interactions generate lower estimated effects of the law (-0.5% and -0.8%) compared to the other sets of models which produce small positive effects of the law (around 2%). There is no systematic pattern in the AME's for the models with urban interactions, but on average they produce the largest positive effects of the law (3.4%)

In general we find that the effect of shall-issue laws on both property and violent crime is increasing and concave in gun prevalence. In other words, the more guns there are in the population, the more likely the law leads to an increase in overall crime. One potential explanation for this result is that as the number of guns in the population increases, any potential deterrent effect of individuals being able to carry concealed weapons is mitigated by the increase in the number of guns actively being carried in public (as opposed to just being owned).

When we allow for the effects of the law to vary with the degree of urbanization, we also find evidence of heterogeneous effects. For property crime, the effect of the law is monotonically decreasing in urbanization. The more urban (rural) a county is, the more likely the law will decrease (increase) crime. With violent crime there seems to be a similar pattern, although it is less pronounced, particularly with the spline and hybrid specifications.

The results with regional heterogeneity are harder to interpret. This is partially due to differences in the crime trends across regions, which we discussed earlier. For violent crime, the dummy model specification suggests that the effect of the law for the South is to increase crime (or decrease crime less) relative to the other regions. However, once we allow for the effect of the law to vary over time via the spline or hybrid models, the result flips, with the law associated with a decrease in crime (or a smaller increase in crime) in the South relative to the other regions of the country. Moreover, when we also allow the law to interact with gun prevalence, the change is even stronger, due to the fact that gun prevalence is higher on average in the South.

For property crime there is also evidence of heterogeneous effects of the law across region, but there is no consistent pattern across models. These results suggest that there is important heterogeneity in the effect of shall-issue laws across regions. While interacting the effect of the law helps to isolate this heterogeneity and better fit the observed patterns of crime, region differences alone do not inform us as to the ultimate source of this underlying heterogeneity, something we discuss more in the conclusion.

iv. instrumental variables

When we compare the models which treat the laws and the arrest rate as endogenous (IV models) to those in which they are assumed to be exogenous (Non IV models), we find that the IV models are preferred to their Non IV counterparts. However, the Non IV models are not dominated as a group as was the case with population weights and region trends. This implies that although there is strong support for treating the laws and arrest rate as endogenous, there are other dimensions of the model space that are equally, if not more, important. On average the IV models generate higher average marginal effects (5.5 and 0.9 percentage points higher for violent and property crime, respectively). Figures 4E and 4F demonstrate the reasons for these differences. For violent crime, it is driven by the fact that almost all Non IV models generate negative effects of the law, but the IV models generate a roughly even mix of negative and positive effects. With property crime, on the other hand, the difference is driven by the magnitude of the effects, as opposed to the sign.

v. differences across the average marginal effect measures

So far we have focused our discussion on the three-year average of AME, one of twelve measures we compute for each model. In many cases the average marginal effects vary little, if at all, across the three measures and across the different years in which the effect is measured. However, there are two situations for which this is not the case, and they are worth highlighting. The first case is when the models contain trends (spline/hybrid). The trend component of these specifications produces an effect of the law that builds over time, and unless the trend effect is equal to zero, the magnitude of the effect depends on how long the law has been in place. In some cases the trend is quite large, which makes it difficult to evaluate the effect of the law compared to other models. In addition, as the discussion surrounding Figure 1 shows, the linear trend specification has the unappealing feature that the magnitude of the effect grows without bound. We believe that while the results from these models suggest that there is important time-heterogeneity in the effects of the laws, the correct specification of these time-effects warrants further research. Furthermore, there may be short-run versus long-run trade-offs in terms of the effects of the laws that need to be considered.

The second situation in which there are important differences across the average marginal effects measures are when the models include interactions of the law with region. This result is presented graphically in Figure 5. When the average marginal effects are computed for states that implemented the law (AME-Law), compared to states that did not implement the law (AME-No Law), we see sizeable differences. (Recall that our main AME measure is just a weighted average of these two measures.) This is reflective of the fact that those regions that were more likely to adopt shall-issue laws experienced different effects of the law compared to regions that were less likely to adopt. This is an important result to note, particularly to policy-makers trying to decide whether to enact such laws. The effects of the law vary across location, and the experiences of those that previously enacted the law are not necessarily going to be replicated in other locations.

This discussion is quite wide ranging. We summarize our findings on the model space as follows.

1. In terms of fit (i.e., marginal likelihoods), models that a) allow for heterogeneity and b) correct for endogeneity, substantially outperform their counterparts.

2. The effect of the law seems to be very heterogeneous across many dimensions. Interacting the law with observables like gun prevalence, region or urban dummies shows that a) having more guns seems to reduce the potential deterrent effect of the law, b) more urbanization increases the potential deterrent effect of the law and c) there seem to be important regional differences in the effect of the law on crime.

3. Models that only allow for the effect of the law to shift crime (i.e., dummy models) find large and positive effects of the law, models that allow for trends in the effect of the law (spline models) find negative but mixed effects of the law, and models that allow for both a trend and a shift (hybrid models) tend to find negative but small effects (i.e., evidence of weak deterrence).

4. There is little support for the substitution hypothesis of Lott (Lott and Mustard (1997), Lott (2000)) in which criminals shift their activity from violent crimes towards less confrontational crimes like property crimes.

5. Finally, all results should be interpreted with caution, as there is enormous heterogeneity in the estimated effects across models, in terms of the signs, magnitudes, and their economic and statistical significance.

9. Posterior Probabilities and Model-Averaged Results

What conclusions would one draw from a model averaging exercise? Employing a uniform prior distribution over the model space and the marginal likelihoods presented in the Appendix, only two models receive posterior probabilities above 1% for violent crime. We label these as model A and model B. The specifications for each are listed below, based on the characterization of the model space in Figure 2.

- Model A: Hybrid, all demographic controls, no additional covariates, time dummies, gun prevalence and region interactions, IV, no population weights.

- Model B: Hybrid, all demographic controls, add urban dummies, time dummies, gun prevalence and region interactions, IV, no population weights.

The difference between the two models is that model B also includes urban dummies as control variables. The posterior probability weights for model A and model B are 98.4% and 1.6%, respectively.

As can be seen in Table 2, both models produce similar results in terms of the AME's. For both models, all three average marginal effect measures suggest a similar pattern. First, as a consequence of the shift effect of the dummy for the law, violent crime falls by about 10% in the first year after the implementation of the law. However, because of the strong positive trend in the effect of the law (which comes from the spline component to the hybrid model), this initial deterrent effect is eliminated by the third year.[24] In fact, because of the linear trend, the effect of the law will become positive and increasingly large over time if extended beyond three years. So while there is some evidence of an immediate decrease in violent crime, the long-run effect seems to be to increase crime, quickly wiping out any initial decrease.

With property crime, the posterior model probabilities load entirely on one model, which we denote as model C.

- Model C: Hybrid, all demographic controls, no additional covariates, time dummies, gun prevalence and urban interactions, IV, no population weights.

[24]Note that when the model includes interactions of variables with the law, the AME measures will not necessarily follow a simple linear trend over time. In fact, depending on how the interacted variables change over time, they need not even be monotone.

For property crime there is initially an economically and statistically significant drop in crime of 8%. However, as was the case with violent crime, there is a steep positive trend in the effect of the law on crime. As a consequence the deterrent effects of the law get quickly wiped out and would become increasingly positive if extended over the long run.

Beyond the estimated effects we just presented, the model averaging exercise illustrates a pattern that warrants further research. The fact that the data prefers models with both dummies and trends on the effect of the law, shows that the law has effects that are not only heterogeneous in the cross section (by region, prevalence of guns, etc) but with strong time heterogeneity as well. However, the practice of introducing this time heterogeneity via simple trends is problematic. For our estimates, the effect on crime of introducing guns moves from negative to positive in three to four years and would simply continue to grow over time. So it seems important to consider richer specifications of the policy effect than have appeared in the literature.

10. Conclusions

Relative to the strong claims made by particular papers in the literature, we find little evidence that shall-issue right-to-carry laws generate either an increase or decrease in crime on average. Furthermore, we do not find any evidence of these laws leading to substitution away from violent crime and into property crime.

However, we do find some interesting results that we believe are revealing as to how shall-issue laws affect crime. For both types of crime, the degree of gun prevalence in a county is an important determinant of how the law affects crime. Interestingly, for violent crime, heterogeneity across regions is also important, whereas the level of urbanization is more important for property crime. While there are regional differences in urbanization, it appears that there are other differences across region that drive violent crime. With property crime, the level of urbanization seems to be an important difference that transcends regional boundaries.

Urban areas differ from rural areas in how closely people live to one another. Property crimes, by definition, do not involve confrontations with victims. Therefore, it seems natural that concealed-carry laws could have different effects on crime in areas in which potential witnesses to property crimes (who may be carrying concealed weapons) are more dispersed compared to areas in which they are tightly clustered. Alternatively, it may be easier to carry a concealed weapon regularly in a rural area in which there may be fewer places in which weapons are not allowed, such as mass transit and state or federal facilities, resulting in differential usage of the law in rural versus urban locations.

For violent crimes, with the exception of murder, there is always at least one witness to the crime, the victim. Therefore, the fact that there might be other armed witnesses might be less relevant for these crimes. Of course, our finding that urbanization matters more for property crimes, and region matters more for violent crimes, does not mean that urbanization is not relevant. Rather, it reveals that there are other factors that vary across regions, which are more important. The use of region interactions helps to identify the level at which these differences exist, but does not provide any direct evidence as to the underlying mechanism. Attitudes towards violence, gun safety, and gun usage in general are examples of potential differences.

We believe that, in particular, our results related to parameter heterogeneity highlight an area of research that is currently underdeveloped in the literature. The fact that interacting the effect of the law with time trends and region helps to better explain crime patterns in the data, suggests the presence of important sources of heterogeneity that are not captured by current models of crime.

We can speculate as to the underlying mechanisms at work here, and we can try to identify additional observable variables to include in the model. However, progress will, we believe, primarily depend on the introduction of a more structural approach to the analysis of shall issue laws. The assumptions underlying both the theories that predict the law will lead to an increase in crime and those that predict a decrease all involve individuals making different decisions in the presence of concealed-carry laws. Therefore, incorporating this behavior into the model could be useful in several dimensions. It would impose constraints on the set of empirical models consistent with the theory. In addition, it would generate testable restrictions that the researcher could evaluate with the available data.

An example of where a more structured approach to modeling crime would help arises in our finding that a time-varying effect of the law is important to explain the patterns in the data. Without a model that explains why these temporal patterns arise, it is not clear how to interpret these changes or how useful it is to even estimate them as opposed to just recovering some “average” effect of the policy. Is this heterogeneity a consequence of restrictions on the “technology” of crime commission, and hence the policy fundamentally alters how criminals commit crimes? Are these patterns the consequence of changes in expectations by the agents, and what we see is simply a temporary effect while agents “figure” out how to identify gun-carrying individuals, for example? Do they arise because of interactions with other (potentially unmeasured) laws that also change over time, etc.? The answers to these questions have different implications for both policy evaluation (e.g., we may not be able to estimate the effect independently of other policies) and implementation (e.g., packages of policies, either concurrently or in particular sequences, may be more effective).[25]

[25] See for example Murphy (2003), Heckman and Navarro (2007), and Cooley, Navarro and Takahashi (2014) for a discussion of issues surrounding time-varying treatment effects.

References

Aneja, A., J. Donohue, and A. Zhang. 2011. "The Impact of Right-to-Carry Laws and the NRC Report: Lessons for the Empirical Evaluation of Law and Policy." *American Law and Economics Review* 13: 565-632.

Ayres, I. and J. Donohue. 2003. "Shooting Down the 'More Guns, Less Crime' Hypothesis." *Stanford Law Review* 55: 1193-1312.

Ayres, I. and J. Donohue. 2009a. "Yet Another Refutation of the More Guns, Less Crime Hypothesis—With Some Help from Moody and Marvell." *Econ Journal Watch* 6: 35-59.

Ayres, I. and J. Donohue. 2009b. "More Guns, Less Crime Fails Again: The Latest Evidence from 1977-2006." *Econ Journal Watch* 6: 218-238.

Bartley, W. and M. Cohen 1998. "The Effect of Concealed Weapon Laws: An Extreme Bound Analysis." *Economic Inquiry* 36: 258-265.

Becker, G. 1968. "Crime and Punishment: An Economic Analysis." *Journal of Political Economy* 78: 169-217.

Black, D. and D. Nagin, 1998. "Do Right-To-Carry Laws Deter Violent Crime?" *Journal of Legal Studies* 27: 209-219.

Brock, W. and S. Durlauf. 2001. "Growth Empirics and Reality." *World Bank Economic Review* 15: 229-272.

Brock, W., S. Durlauf, and K. West. 2003. "Policy Evaluation in Uncertain Economic Environments (with discussion)." *Brookings Papers on Economic Activity* 1: 235-322.

Brock, W., S. Durlauf, and K. West. 2007. "Model Uncertainty and Policy Evaluation: Some Theory and Empirics." *Journal of Econometrics* 136: 629-664.

Brock, W. and S. Durlauf. 2014. "A Sturdy Policy Evaluation." *Journal of Legal Studies,* forthcoming.

Cohen-Cole, E., S. Durlauf, J. Fagan, and D. Nagin. 2009. "Model Uncertainty and the Deterrent Effect of Capital Punishment." *American Law and Economics Review* 11: 335-369.

Cooley, J., S. Navarro, and Y. Takahashi. 2014. "Identification and Estimation of Time-varying Treatment Effects: How the Timing of Grade Retention Affects Outcomes." Mimeo, University of Western Ontario.

Dezhbakhsh, H. and P. Rubin. 1998. “Lives Saved or Lives Lost? The Effects of Concealed-Handgun Laws on Crime.” *The American Economic Review, Papers and Proceedings* 88: 468-474.

Donohue, J. 2004. “Guns, Crime, and the Impact of State Right-to-Carry Laws.” *Fordham Law Review* 73: 623-640.

Doppelhofer, G. 2008. “Model Averaging”. *The New Palgrave Dictionary in Economics*, revised edition. S. Durlauf and L. Blume, eds. London: MacMillan.

Draper, D. 1995. “Assessment and Propagation of Model Uncertainty.” *Journal of the Royal Statistical Society*, *series B*. 57: 45-97.

Duggan, M. 2001. “More Guns, More Crime.” *Journal of Political Economy* 109: 1086-1114.

Durlauf, S., C. Fu, and S. Navarro. 2013. “Capital Punishment and Deterrence: Understanding Disparate Results.” *Journal of Quantitative Criminology* 29: 103-121.

Durlauf, S., S. Navarro, and D. Rivers. 2008. “On the Interpretation of Aggregate Crime Regressions.” *Crime Trends*, A. Goldberger and R. Rosenfeld, eds. Washington DC: National Academies Press.

Durlauf, S., S. Navarro, and D. Rivers. 2010. “Understanding Aggregate Crime Regressions.” *Journal of Econometrics* 158: 306-317.

Durlauf, S., S. Navarro, and D. Rivers. 2014. “A Bayesian MCMC Panel Data Procedure under Endogeneity and Heteroskedasticity.” Mimeo, University of Western Ontario.

Eicher, T., A. Lenkoski, and A. Raftery. 2011. “Bayesian Model Averaging and Endogeneity Under Model Uncertainty: An Application to Development Determinants.” *Journal of Applied Econometrics* 26: 30-55.

Fernandez, C., E. Ley and M. Steel. 2001. “Model Uncertainty in Cross-Country Growth Regressions.” *Journal of Applied Econometrics* 16: 563-576.

Hansen, L. and T. Sargent. 2007. *Robustness*. Princeton: Princeton University Press.

Hausman, J. 1996. “Valuation of New Goods under Perfect and Imperfect Competition”. *The Economics of New Goods*, T. Bresnahan and R. Gordon, eds. Chicago: University of Chicago Press.

Heckman, J. 2005. “The Scientific Model of Causality.” *Sociological Methodology* 35: 1-97.

Heckman, J. and S. Navarro. 2007. “Dynamic Discrete Choice and Dynamic Treatment Effects”, *Journal of Econometrics* 136: 341-396.

Hoeting, J., D. Madigan, A. Raftery, A. and C. Volinsky. 1999. “Bayesian Model Averaging: A Tutorial.” *Statistical Science* 14: 382-401.

Leamer, E. 1978. *Specification Searches*. New York: John Wiley.

Leamer, E. 1983. “Let’s Take the Con Out of Econometrics.” *American Economic Review* 73: 31-43.

Lott, J. 2000. *More Guns, Less Crime: Understanding Crime and Gun Control Laws*, 2nd ed. Chicago: University of Chicago Press.

Lott, J. and D. Mustard. 1997. “Crime, Deterrence and Right-to-Carry Concealed Handguns.” *Journal of Legal Studies* 26: 1-68.

Lucas, R. 1976. “Econometric Policy Evaluation: A Critique.” *Carnegie-Rochester Conference Series on Public Policy* 1: 19-46.

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

Madigan, D. and A. Raftery. 1994. “Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam’s Window.” *Journal of the American Statistical Association* 89: 1535-1546.

Moody, C. and T. Marvell. 2005. “Guns and Crime.” *Southern Economic Journal* 71: 720-736.

Moody, C. and T. Marvell. 2008. “The Debate on Shall-Issue Laws.” *Econ Journal Watch* 5: 269-293.

Moody, C. and T. Marvell. 2009. “The Debate on Shall-Issue Laws, Continued.” *Econ Journal Watch* 6: 203-217.

Murphy S. 2003. “Optimal Dynamic Treatment Regimes.” *Journal of the Royal Statistical Society, Series B (with discussion)* 65: 331-366.

Nevo, A. 2001. “Measuring Market Power in the Ready-to-Eat Cereal Industry.” *Econometrica* 69: 307-342.

Newton, M. and A. Raftery. 1994. “Approximate Bayesian Inference with the Weighted Likelihood Bootstrap.” *Journal of the Royal Statistical Society. series B* 56: 3-48.

Plassmann, F. and T. N. Tideman. 2001. “Does the Right to Carry Concealed Handguns Deter Countable Crimes? Only a Count Analysis Can Say.” *Journal of Law and Economics* 44, S2: 771-798.

Raftery, A. 1995. “Bayesian Model Selection in Social Research (with discussion).” *Sociological Methodology* 25: 111-196.

Raftery, A., D, Madigan, and J. Hoeting. 1997. “Bayesian Model Averaging for Linear Regression Models.” *Journal of the American Statistical Association* 92: 179-191.

Sala-i-Martin, X., G. Doppelhofer, and R. Miller. 2004. “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach.” *American Economic Review* 94: 813-835

Shepherd, J. 2005. “Deterrence Versus Brutalization: Capital Punishment’s Differing Impacts Across States.” *Michigan Law Review* 104: 203-255.

Strnad, J., 2007. “Should Legal Empiricists Go Bayesian?” *American Law and Economics Review* 9: 195-303 .

Wellford, C., J. Pepper, and C. Petrie, eds. 2004. *Firearms and Violence*. Washington DC: National Academy of Sciences Press.

Figure 1: Hybrid Model Example

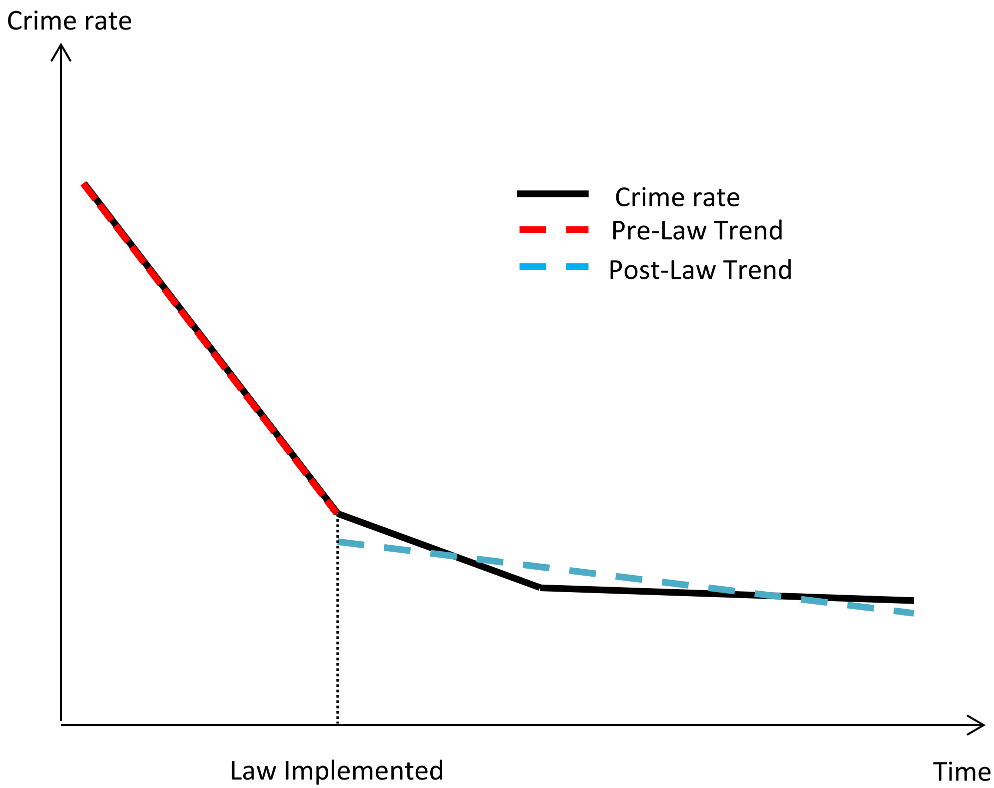
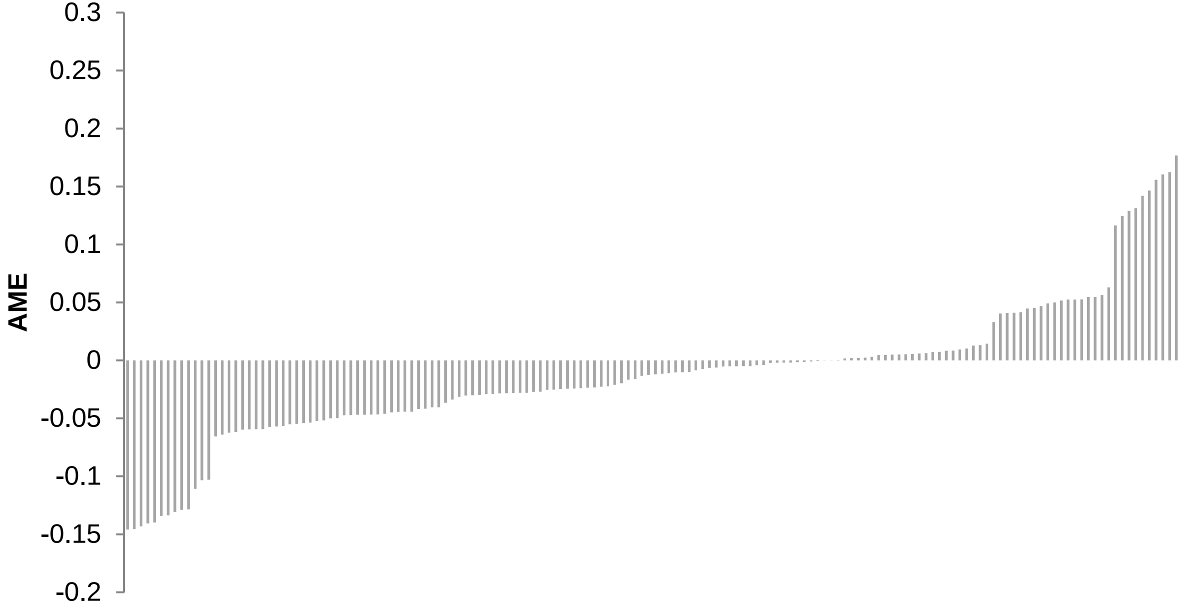


Figure 2: Model Space

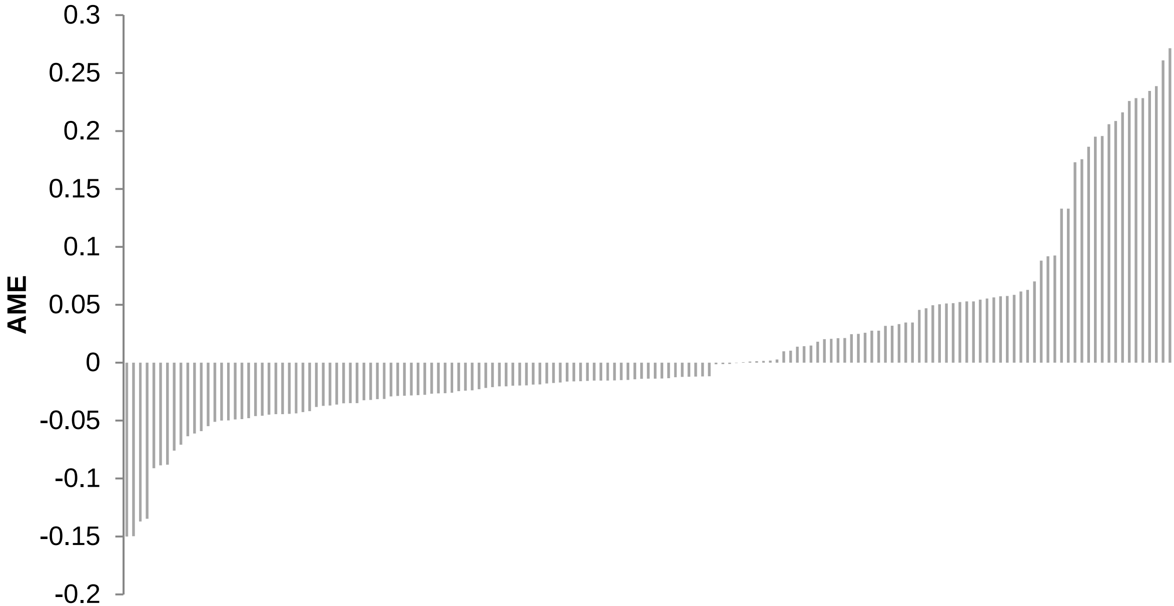
Crime Category	Formulation of the Laws	Demographics	Additional Covariates	Time Effects	Parameter Heterogeneity	Instrumental Variables	Homoskedastic/ Heteroskedastic
1. Property 2. Violent	1. Dummy 2. Spline 3. Hybrid	1. All 2. Collapsed Demographics	1. None 2. Urban 3. Gun Prevalence (GP) and GP Squared	1. Time Dummies 2. Region Trends	1. None 2. GP and GP squared 3. Urban 4. Region 5. GP, GP Squared and Region 6. GP, GP Squared and Urban	1. Non IV 2. IV	1. No weights 2. Population weights

Figure 3A: Average Marginal Effects - Violent Crime



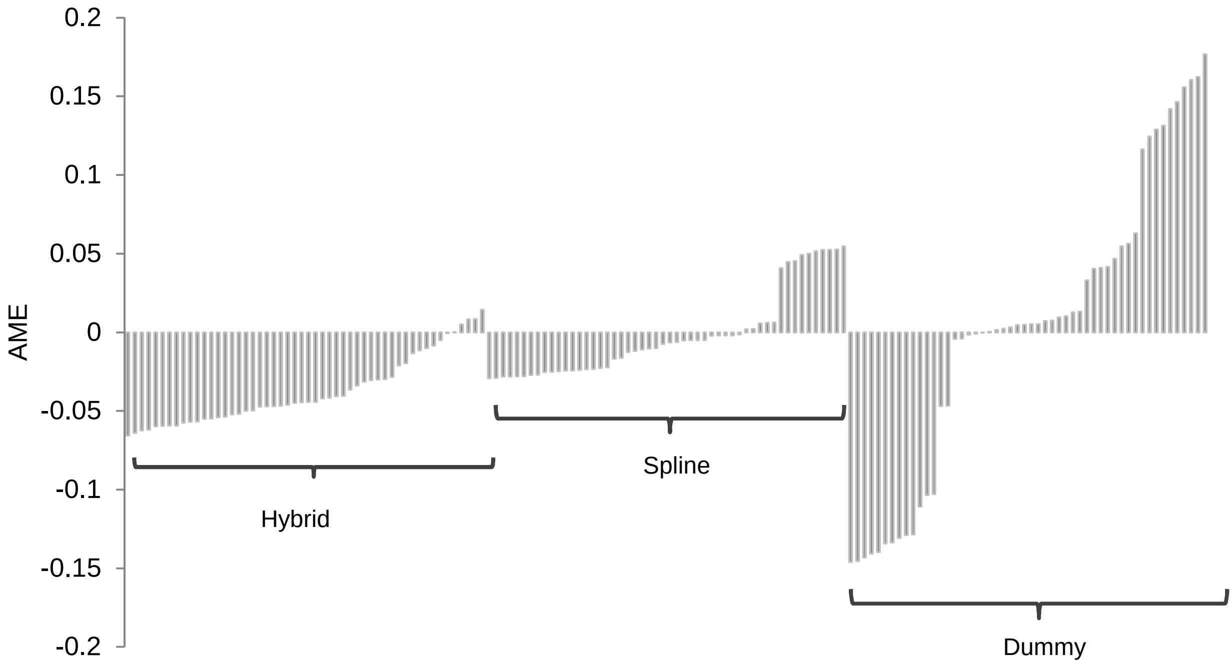
Note: Each bar represents a specific model's three-year average effect of introducing a shall-issue law, computed for all states using their characteristics in 1998-2000 (AME).

Figure 3B: Average Marginal Effects - Property Crime



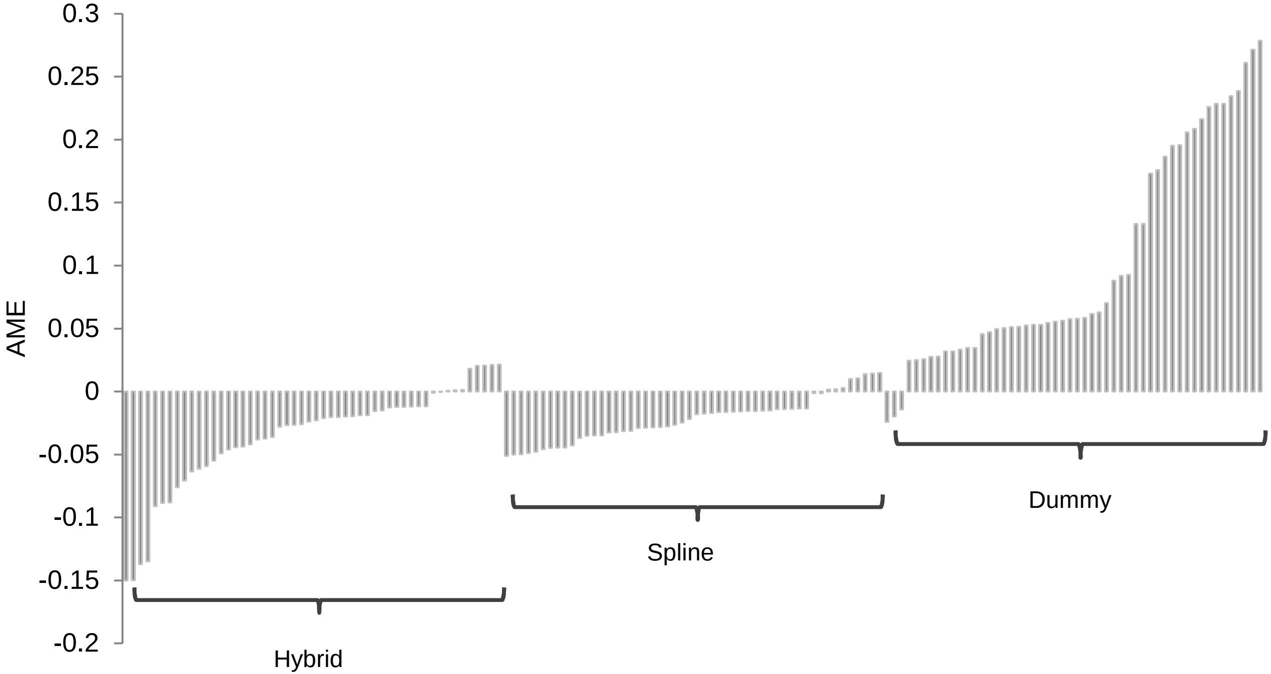
Note: Each bar represents a specific model's three-year average effect of introducing a shall-issue law, computed for all states using their characteristics in 1998-2000 (AME).

Figure 4A: Average Marginal Effects - By Law Specification - Violent Crime



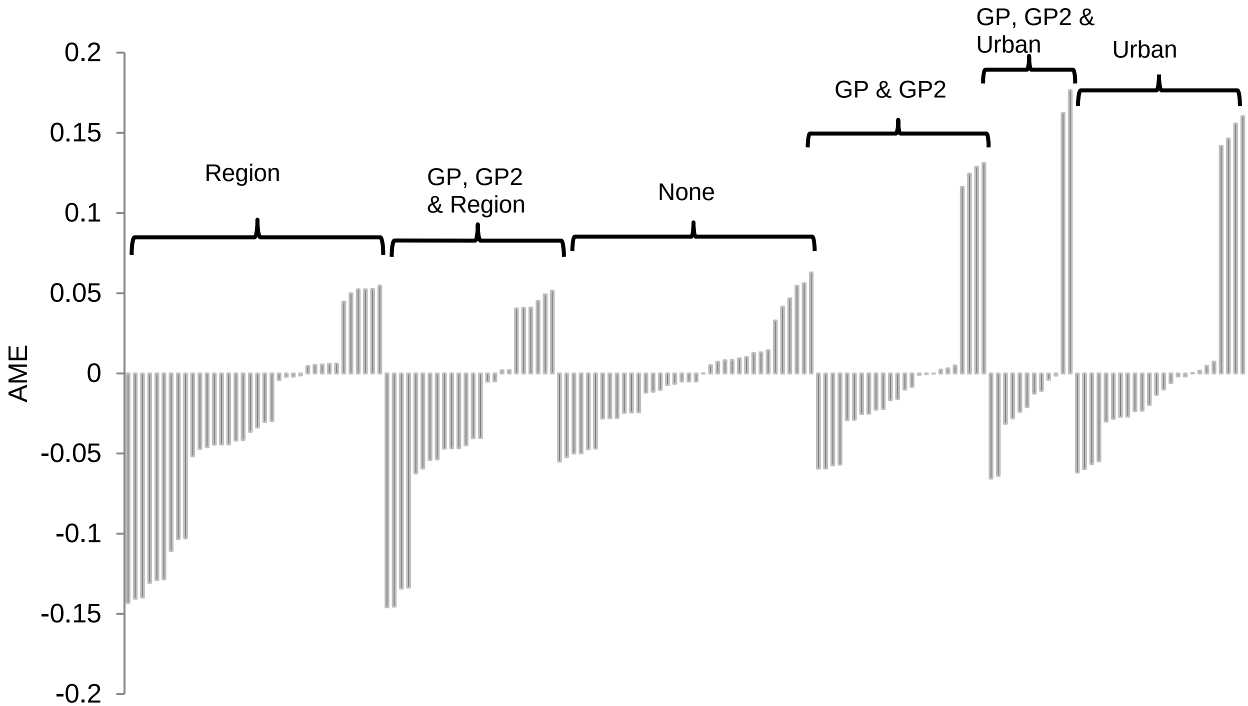
Note: Each bar represents a specific model's three-year average effect of introducing a shall-issue law, computed for all states using their characteristics in 1998-2000 (AME).

Figure 4B: Average Marginal Effects - By Law Specification - Property Crime



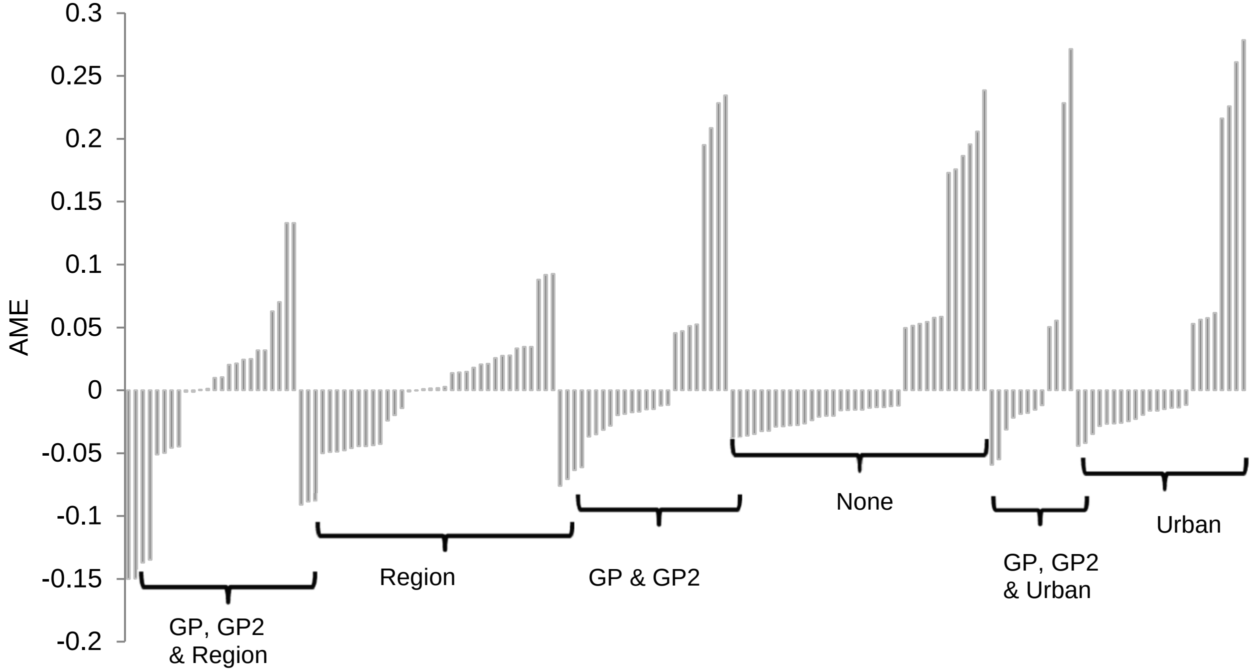
Note: Each bar represents a specific model's three-year average effect of introducing a shall-issue law, computed for all states using their characteristics in 1998-2000 (AME).

Figure 4C: Average Marginal Effects - Parameter Heterogeneity - Violent Crime



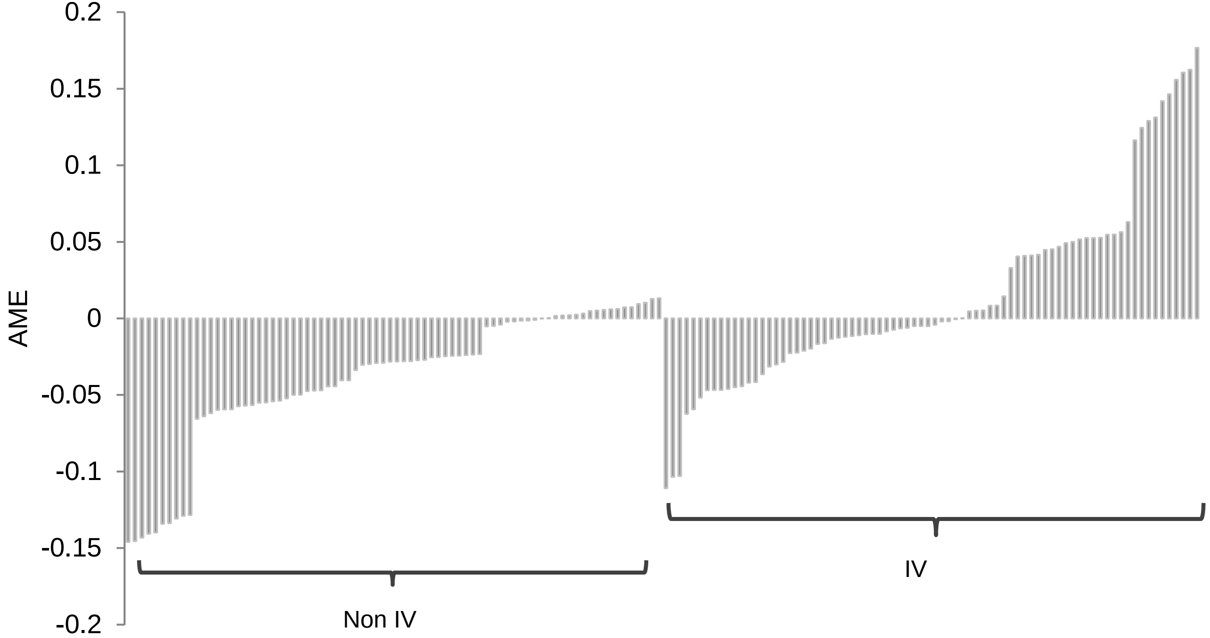
Note: Each bar represents a specific model's three-year average effect of introducing a shall-issue law, computed for all states using their characteristics in 1998-2000 (AME). GP and GP2 stand for gun prevalence and gun prevalence squared.

Figure 4D: Average Marginal Effects - Parameter Heterogeneity - Property Crime



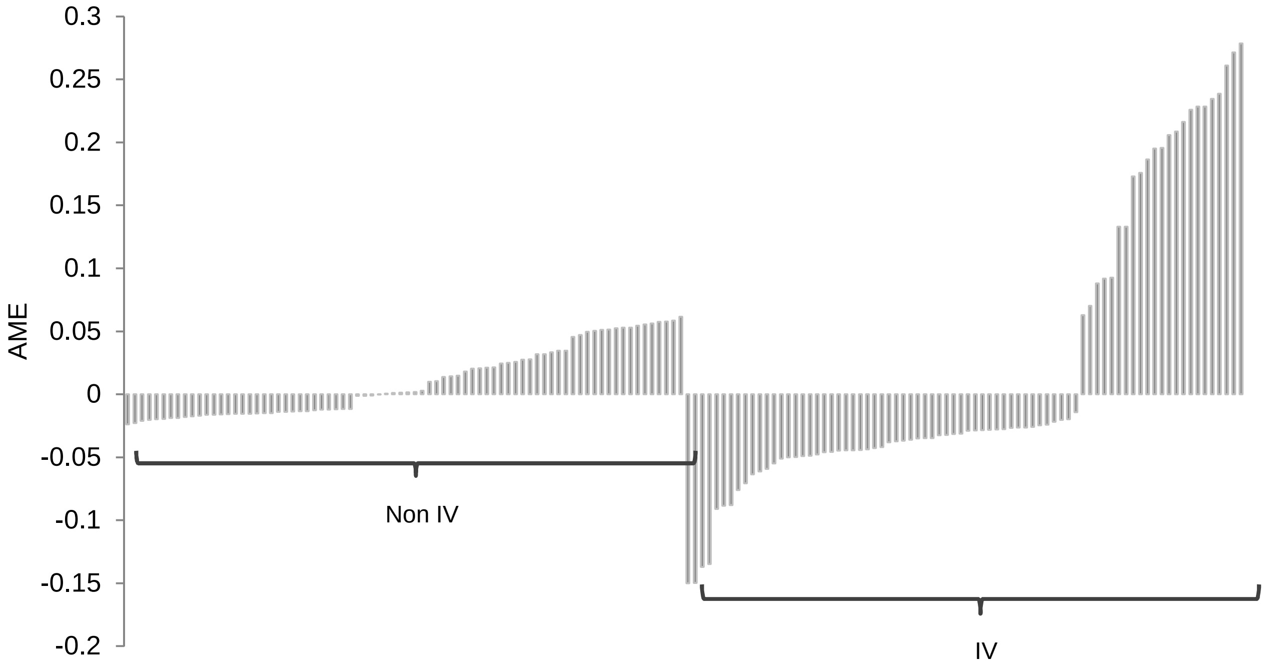
Note: Each bar represents a specific model's three-year average effect of introducing a shall-issue law, computed for all states using their characteristics in 1998-2000 (AME). GP and GP2 stand for gun prevalence and gun prevalence squared.

Figure 4E: Average Marginal Effects - IV versus Non-IV Specifications - Violent Crime



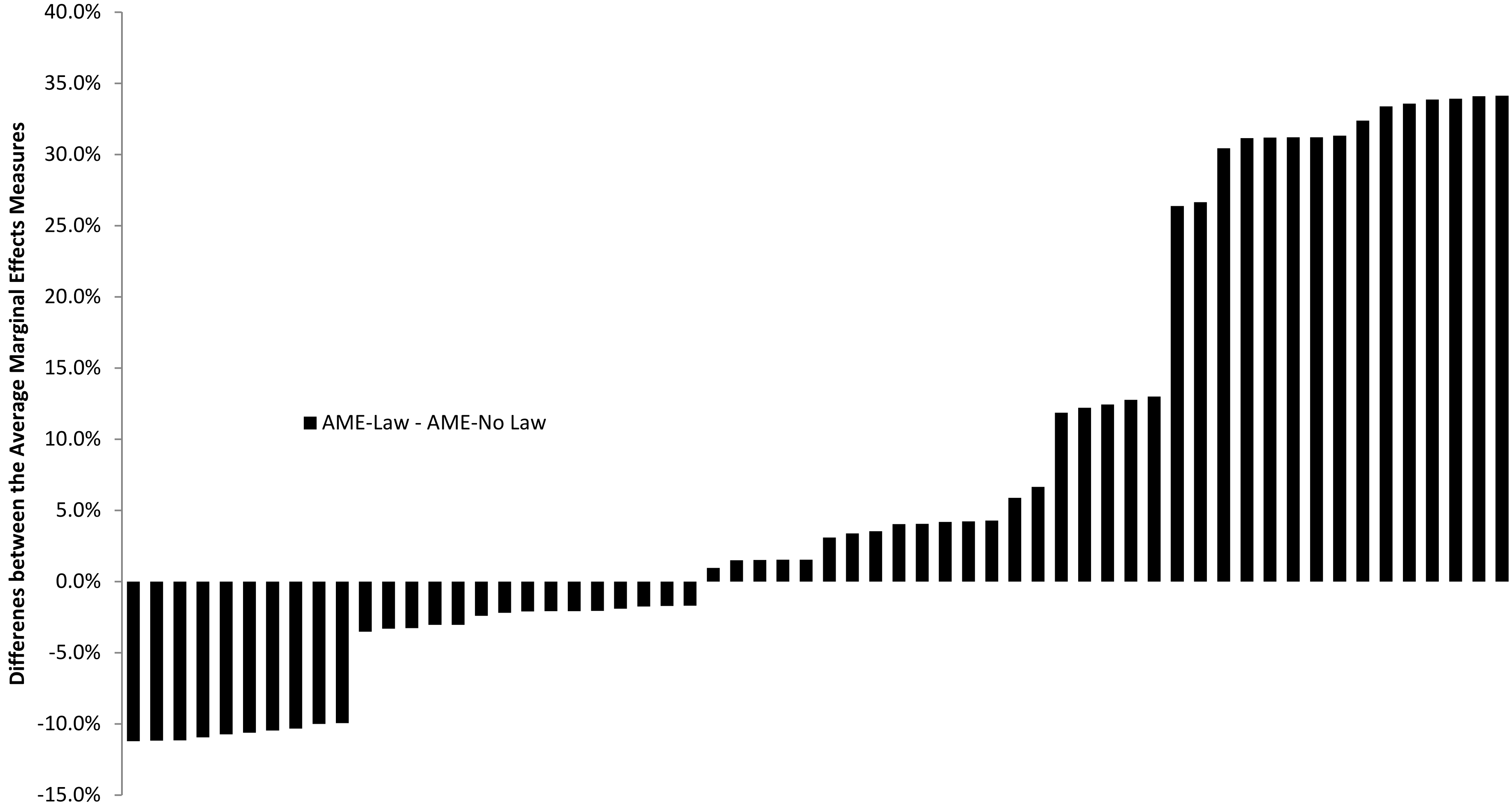
Note: Each bar represents a specific model's three-year average effect of introducing a shall-issue law, computed for all states using their characteristics in 1998-2000 (AME).

Figure 4F: Average Marginal Effects - IV versus Non-IV Specifications - Property Crime



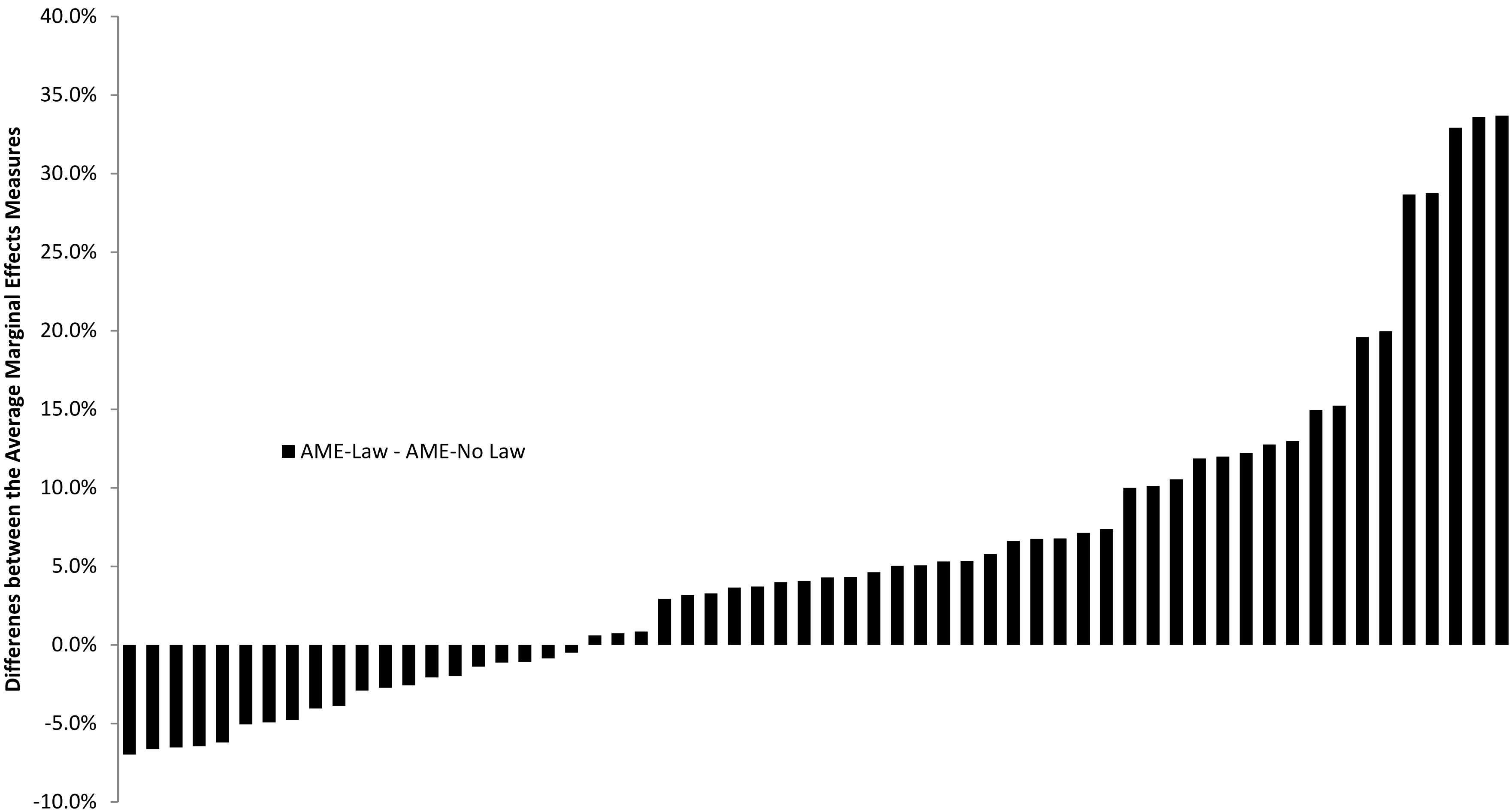
Note: Each bar represents a specific model's three-year average effect of introducing a shall-issue law, computed for all states using their characteristics in 1998-2000 (AME).

Figure 5A: Differences in the Average Marginal Effects: Models with Region Interactions - Violent Crime



Note: Each bar represents the difference between AME-Law and AME-No Law, for each model that contains interactions of the law with region dummies. AME-Law is the three-year average marginal effect for those states that implemented the law, evaluated using the characteristics of those states that implemented the law. AME-No Law is the same, except that it is evaluated for those states that did not implement the law.

Figure 5B: Differences in the Average Marginal Effects: Models with Region Interactions - Property Crime



Note: Each bar represents the difference between AME-Law and AME-No Law, for each model that contains interactions of the law with region dummies. AME-Law is the three-year average marginal effect for those states that implemented the law, evaluated using the characteristics of those states that implemented the law. AME-No Law is the same, except that it is evaluated for those states that did not implement the law.

Table 1: Average of AME Across Specifications

Model Dimension	Specification	Property	Violent
	Dummy	10.2%	0.2%
Formulation of the Law	Spline	-2.2%	-0.2%
	Hybrid	-3.6%	-3.7%
Demographics	All	1.5%	-1.0%
	Collapsed	1.5%	-1.4%
	None	1.7%	-0.9%
Additional Regressors	Urban	0.8%	-1.7%
	GP and GP2	1.8%	-1.3%
	None	2.6%	-0.4%
	GP and GP2	2.2%	0.3%
	Urban	3.4%	0.7%
Interactions	Region	-0.5%	-3.6%
	GP, GP2, and Urban	3.1%	0.6%
	GP, GP2, and Region	-0.8%	-3.3%
IV	Non IV	1.0%	-3.9%
	IV	1.9%	1.4%

Notes: Each panel of Table 1 corresponds to one dimension of the model space described in Figure 1. For a given panel, each row contains the simple unweighted average of the estimated marginal effects (in particular AME) of all models that include the model specification for that row. For example, consider the "Formulation of the Law" panel. The row labeled "Dummy" presents the average marginal effect across all models that use the dummy variable specification for the law, average across all other model dimension specifications.

AME is the three-year average marginal effect averaged across all states (including those that did not implement the law), using their characteristics in 1998-2000.

GP and GP2 stand for gun prevalence and gun prevalence squared.

Table 2: Model Averaging Results

Violent Crime

	Posterior Probability	AME Year 1	AME Year 2	AME Year 3	AME Average	AME-Law Year 1	AME-Law Year 2	AME-Law Year 3	AME-Law Average	AME-No Law Year 1	AME-No Law Year 2	AME-No Law Year 3	AME-No Law Average
Model A	98.4%	-10.29%	-5.36%	-0.63%	-5.43%	-8.79%	-3.89%	0.83%	-3.95%	-13.78%	-8.78%	-4.05%	-8.87%
		(4.88%)	*(4.60%)*	*(4.51%)*	*(4.57%)*	*(2.51%)*	*(2.28%)*	*(2.27%)*	*(2.24%)*	*(20.15%)*	*(16.49%)*	*(12.45%)*	*(16.50%)*
Model B	1.6%	-11.27%	-6.13%	-1.32%	-6.24%	-9.17%	-4.15%	0.56%	-4.25%	-16.20%	-10.75%	-5.73%	-10.89%
		(4.98%)	*(4.81%)*	*(4.74%)*	*(4.76%)*	*(2.46%)*	*(2.31%)*	*(2.30%)*	*(2.27%)*	*(22.71%)*	*(18.21%)*	*(13.41%)*	*(18.24%)*
Model-Averaged Results		-10.30%	-5.37%	-0.64%	-5.44%	-8.80%	-3.90%	0.82%	-3.96%	-13.82%	-8.81%	-4.08%	-8.90%
		(4.88%)	*(4.60%)*	*(4.51%)*	*(4.57%)*	*(2.51%)*	*(2.28%)*	*(2.27%)*	*(2.24%)*	*(20.20%)*	*(16.52%)*	*(12.46%)*	*(16.53%)*

Property Crime

	Posterior Probability	AME Year 1	AME Year 2	AME Year 3	AME Average	AME-Law Year 1	AME-Law Year 2	AME-Law Year 3	AME-Law Average	AME-No Law Year 1	AME-No Law Year 2	AME-No Law Year 3	AME-No Law Average
Model C	100.0%	-8.04%	-6.64%	-1.42%	-5.37%	-10.04%	-7.46%	-1.58%	-6.36%	-3.37%	-4.72%	-1.04%	-3.04%
		(1.77%)	*(1.24%)*	*(1.23%)*	*(1.31%)*	*(1.51%)*	*(1.12%)*	*(1.16%)*	*(1.15%)*	*(1.64%)*	*(1.18%)*	*(1.20%)*	*(1.23%)*

Notes: Model A is the hybrid specification that includes the full set of demographic controls, no additional covariates, time dummies, interactions with gun prevalence and region dummies, instrumental variables, and no population weights. Model B is the same as model A, except that it also includes urban dummies. Model C is the hybrid specification that includes the full set of demographic controls, no additional covariates, time dummies, interactions with gun prevalence and urban dummies, instrumental variables, and no population weights.

AME is the average marginal effect for all states using their characteristics in 1998-2000. AME-Law and AME-No Law are the average marginal effects for states that introduced shall-issue laws, and those that did not, during the sample period, also using their characteristics in 1998-2000.

Standard errors of the average marginal effects are given in italics and parentheses below the point estimates.

Appendix

Table A-1: Likelihood - Violent Crime

Model Specification					Log Likelihood
Formulation of the Laws	**Demographics**	**Additional Covariates**	**Parameter Heterogeneity**	**Instrumental Variables**	
Dummy	All	None	None	Non IV	-11,242.22
Dummy	All	None	None	IV	-11,224.72
Dummy	All	None	GP and GP squared	Non IV	-11,217.36
Dummy	All	None	GP and GP squared	IV	-11,174.96
Dummy	All	None	Urban	Non IV	-11,233.48
Dummy	All	None	Urban	IV	-11,204.32
Dummy	All	None	Region	Non IV	-11,176.49
Dummy	All	None	Region	IV	-11,130.16
Dummy	All	None	GP, GP squared and Region	Non IV	-11,206.28
Dummy	All	None	GP, GP squared and Region	IV	-11,141.55
Dummy	All	None	GP, GP Squared and Urban	Non IV	-11,151.07
Dummy	All	None	GP, GP Squared and Urban	IV	-11,101.91
Dummy	All	Urban	None	Non IV	-11,244.94
Dummy	All	Urban	None	IV	-11,227.50
Dummy	All	Urban	GP and GP squared	Non IV	-11,217.94
Dummy	All	Urban	GP and GP squared	IV	-11,175.55
Dummy	All	Urban	Region	Non IV	-11,177.45
Dummy	All	Urban	Region	IV	-11,134.22
Dummy	All	Urban	GP, GP Squared and Urban	Non IV	-11,150.61
Dummy	All	Urban	GP, GP Squared and Urban	IV	-11,097.53
Dummy	All	GP and GP Squared	None	Non IV	-11,222.65
Dummy	All	GP and GP Squared	None	IV	-11,198.33
Dummy	All	GP and GP Squared	Urban	Non IV	-11,204.10
Dummy	All	GP and GP Squared	Urban	IV	-11,174.17
Dummy	All	GP and GP Squared	Region	Non IV	-11,147.64
Dummy	All	GP and GP Squared	Region	IV	-11,108.19
Dummy	Subset	None	None	Non IV	-11,362.38
Dummy	Subset	None	None	IV	-11,336.19
Dummy	Subset	None	GP and GP squared	Non IV	-11,330.57
Dummy	Subset	None	GP and GP squared	IV	-11,285.21
Dummy	Subset	None	Urban	Non IV	-11,350.05
Dummy	Subset	None	Urban	IV	-11,309.89
Dummy	Subset	None	Region	Non IV	-11,276.61
Dummy	Subset	None	Region	IV	-11,220.49
Dummy	Subset	None	GP, GP squared and Region	Non IV	-11,321.74
Dummy	Subset	None	GP, GP squared and Region	IV	-11,263.81
Dummy	Subset	None	GP, GP Squared and Urban	Non IV	-11,249.56
Dummy	Subset	None	GP, GP Squared and Urban	IV	-11,199.36
Dummy	Subset	Urban	None	Non IV	-11,361.00
Dummy	Subset	Urban	None	IV	-11,335.35
Dummy	Subset	Urban	GP and GP squared	Non IV	-11,333.56
Dummy	Subset	Urban	GP and GP squared	IV	-11,282.91
Dummy	Subset	Urban	Region	Non IV	-11,271.77
Dummy	Subset	Urban	Region	IV	-11,223.64
Dummy	Subset	Urban	GP, GP Squared and Urban	Non IV	-11,244.30
Dummy	Subset	Urban	GP, GP Squared and Urban	IV	-11,197.51
Dummy	Subset	GP and GP Squared	None	Non IV	-11,333.37
Dummy	Subset	GP and GP Squared	None	IV	-11,308.00
Dummy	Subset	GP and GP Squared	Urban	Non IV	-11,319.82
Dummy	Subset	GP and GP Squared	Urban	IV	-11,287.86
Dummy	Subset	GP and GP Squared	Region	Non IV	-11,247.08
Dummy	Subset	GP and GP Squared	Region	IV	-11,194.02
Spline	All	None	None	Non IV	-11,201.21
Spline	All	None	None	IV	-11,191.88
Spline	All	None	GP and GP squared	Non IV	-11,162.99
Spline	All	None	GP and GP squared	IV	-11,154.59
Spline	All	None	Urban	Non IV	-11,190.34
Spline	All	None	Urban	IV	-11,177.13
Spline	All	None	Region	Non IV	-11,047.72
Spline	All	None	Region	IV	-11,031.52
Spline	All	None	GP, GP squared and Region	Non IV	-11,153.05
Spline	All	None	GP, GP squared and Region	IV	-11,145.71

Table A-1: Likelihood - Violent Crime

Model Specification					Log Likelihood
Formulation of the Laws	Demographics	Additional Covariates	Parameter Heterogeneity	Instrumental Variables	
Spline	All	None	GP, GP Squared and Urban	Non IV	-11,010.15
Spline	All	None	GP, GP Squared and Urban	IV	-10,989.16
Spline	All	Urban	None	Non IV	-11,207.88
Spline	All	Urban	None	IV	-11,189.44
Spline	All	Urban	GP and GP squared	Non IV	-11,164.09
Spline	All	Urban	GP and GP squared	IV	-11,147.95
Spline	All	Urban	Region	Non IV	-11,044.80
Spline	All	Urban	Region	IV	-11,024.41
Spline	All	Urban	GP, GP Squared and Urban	Non IV	-11,010.67
Spline	All	Urban	GP, GP Squared and Urban	IV	-10,986.65
Spline	All	GP and GP Squared	None	Non IV	-11,172.23
Spline	All	GP and GP Squared	None	IV	-11,170.06
Spline	All	GP and GP Squared	Urban	Non IV	-11,169.72
Spline	All	GP and GP Squared	Urban	IV	-11,149.68
Spline	All	GP and GP Squared	Region	Non IV	-11,023.77
Spline	All	GP and GP Squared	Region	IV	-11,000.20
Spline	Subset	None	None	Non IV	-11,315.77
Spline	Subset	None	None	IV	-11,309.07
Spline	Subset	None	GP and GP squared	Non IV	-11,276.66
Spline	Subset	None	GP and GP squared	IV	-11,265.98
Spline	Subset	None	Urban	Non IV	-11,307.73
Spline	Subset	None	Urban	IV	-11,301.93
Spline	Subset	None	Region	Non IV	-11,109.00
Spline	Subset	None	Region	IV	-11,089.59
Spline	Subset	None	GP, GP squared and Region	Non IV	-11,268.49
Spline	Subset	None	GP, GP squared and Region	IV	-11,261.69
Spline	Subset	None	GP, GP Squared and Urban	Non IV	-11,078.95
Spline	Subset	None	GP, GP Squared and Urban	IV	-11,053.04
Spline	Subset	Urban	None	Non IV	-11,315.58
Spline	Subset	Urban	None	IV	-11,308.66
Spline	Subset	Urban	GP and GP squared	Non IV	-11,278.49
Spline	Subset	Urban	GP and GP squared	IV	-11,264.29
Spline	Subset	Urban	Region	Non IV	-11,109.74
Spline	Subset	Urban	Region	IV	-11,091.26
Spline	Subset	Urban	GP, GP Squared and Urban	Non IV	-11,076.97
Spline	Subset	Urban	GP, GP Squared and Urban	IV	-11,058.96
Spline	Subset	GP and GP Squared	None	Non IV	-11,287.52
Spline	Subset	GP and GP Squared	None	IV	-11,272.11
Spline	Subset	GP and GP Squared	Urban	Non IV	-11,281.43
Spline	Subset	GP and GP Squared	Urban	IV	-11,273.62
Spline	Subset	GP and GP Squared	Region	Non IV	-11,087.27
Spline	Subset	GP and GP Squared	Region	IV	-11,066.30
Hybrid	All	None	None	Non IV	-11,199.55
Hybrid	All	None	None	IV	-11,176.68
Hybrid	All	None	GP and GP squared	Non IV	-11,164.71
Hybrid	All	None	GP and GP squared	IV	-11,130.04
Hybrid	All	None	Urban	Non IV	-11,184.51
Hybrid	All	None	Urban	IV	-11,114.25
Hybrid	All	None	Region	Non IV	-11,042.83
Hybrid	All	None	Region	IV	-10,958.30
Hybrid	All	None	GP, GP squared and Region	Non IV	-11,145.67
Hybrid	All	None	GP, GP squared and Region	IV	-11,070.62
Hybrid	All	None	GP, GP Squared and Urban	Non IV	-11,012.75
Hybrid	All	None	GP, GP Squared and Urban	IV	-10,939.45
Hybrid	All	Urban	None	Non IV	-11,201.32
Hybrid	All	Urban	None	IV	-11,186.59
Hybrid	All	Urban	GP and GP squared	Non IV	-11,163.99
Hybrid	All	Urban	GP and GP squared	IV	-11,134.94
Hybrid	All	Urban	Region	Non IV	-11,044.94
Hybrid	All	Urban	Region	IV	-10,960.88
Hybrid	All	Urban	GP, GP Squared and Urban	Non IV	-11,014.65
Hybrid	All	Urban	GP, GP Squared and Urban	IV	-10,937.17
Hybrid	All	GP and GP Squared	None	Non IV	-11,175.47
Hybrid	All	GP and GP Squared	None	IV	-11,151.44

Table A-1: Likelihood - Violent Crime

Model Specification					
Formulation of the Laws	Demographics	Additional Covariates	Parameter Heterogeneity	Instrumental Variables	Log Likelihood
Hybrid	All	GP and GP Squared	Urban	Non IV	-11,157.70
Hybrid	All	GP and GP Squared	Urban	IV	-11,087.99
Hybrid	All	GP and GP Squared	Region	Non IV	-11,022.16
Hybrid	All	GP and GP Squared	Region	IV	-10,945.36
Hybrid	Subset	None	None	Non IV	-11,310.76
Hybrid	Subset	None	None	IV	-11,294.48
Hybrid	Subset	None	GP and GP squared	Non IV	-11,275.16
Hybrid	Subset	None	GP and GP squared	IV	-11,240.00
Hybrid	Subset	None	Urban	Non IV	-11,301.91
Hybrid	Subset	None	Urban	IV	-11,226.82
Hybrid	Subset	None	Region	Non IV	-11,109.62
Hybrid	Subset	None	Region	IV	-11,037.95
Hybrid	Subset	None	GP, GP squared and Region	Non IV	-11,264.54
Hybrid	Subset	None	GP, GP squared and Region	IV	-11,179.98
Hybrid	Subset	None	GP, GP Squared and Urban	Non IV	-11,078.71
Hybrid	Subset	None	GP, GP Squared and Urban	IV	-11,009.37
Hybrid	Subset	Urban	None	Non IV	-11,311.57
Hybrid	Subset	Urban	None	IV	-11,292.76
Hybrid	Subset	Urban	GP and GP squared	Non IV	-11,277.29
Hybrid	Subset	Urban	GP and GP squared	IV	-11,241.19
Hybrid	Subset	Urban	Region	Non IV	-11,112.13
Hybrid	Subset	Urban	Region	IV	-11,041.68
Hybrid	Subset	Urban	GP, GP Squared and Urban	Non IV	-11,076.92
Hybrid	Subset	Urban	GP, GP Squared and Urban	IV	-11,011.94
Hybrid	Subset	GP and GP Squared	None	Non IV	-11,285.91
Hybrid	Subset	GP and GP Squared	None	IV	-11,269.28
Hybrid	Subset	GP and GP Squared	Urban	Non IV	-11,270.35
Hybrid	Subset	GP and GP Squared	Urban	IV	-11,203.07
Hybrid	Subset	GP and GP Squared	Region	Non IV	-11,089.35
Hybrid	Subset	GP and GP Squared	Region	IV	-11,018.35

Note: Each line corresponds to a specific model. Every specification includes time dummies instead of region trends and assumes errors are homoskedastic. GP stands for gun prevalence. For a description of the various modeling assumptions see Section 3.

Table A-2: Likelihood - Property Crime

Model Specification					Log Likelihood
Formulation of the Laws	Demographics	Additional Covariates	Parameter Heterogeneity	Instrumental Variables	
Dummy	All	None	None	Non IV	233.23
Dummy	All	None	None	IV	248.69
Dummy	All	None	GP and GP squared	Non IV	264.50
Dummy	All	None	GP and GP squared	IV	427.66
Dummy	All	None	Urban	Non IV	262.56
Dummy	All	None	Urban	IV	340.95
Dummy	All	None	Region	Non IV	239.77
Dummy	All	None	Region	IV	379.08
Dummy	All	None	GP, GP squared and Region	Non IV	296.97
Dummy	All	None	GP, GP squared and Region	IV	481.70
Dummy	All	None	GP, GP Squared and Urban	Non IV	265.53
Dummy	All	None	GP, GP Squared and Urban	IV	397.51
Dummy	All	Urban	None	Non IV	250.91
Dummy	All	Urban	None	IV	271.55
Dummy	All	Urban	GP and GP squared	Non IV	288.64
Dummy	All	Urban	GP and GP squared	IV	440.45
Dummy	All	Urban	Region	Non IV	259.11
Dummy	All	Urban	Region	IV	395.84
Dummy	All	Urban	GP, GP Squared and Urban	Non IV	287.17
Dummy	All	Urban	GP, GP Squared and Urban	IV	411.07
Dummy	All	GP and GP Squared	None	Non IV	265.23
Dummy	All	GP and GP Squared	None	IV	293.02
Dummy	All	GP and GP Squared	Urban	Non IV	297.79
Dummy	All	GP and GP Squared	Urban	IV	369.76
Dummy	All	GP and GP Squared	Region	Non IV	266.91
Dummy	All	GP and GP Squared	Region	IV	410.53
Dummy	Subset	None	None	Non IV	-49.76
Dummy	Subset	None	None	IV	-22.85
Dummy	Subset	None	GP and GP squared	Non IV	-9.29
Dummy	Subset	None	GP and GP squared	IV	200.22
Dummy	Subset	None	Urban	Non IV	-26.66
Dummy	Subset	None	Urban	IV	74.90
Dummy	Subset	None	Region	Non IV	-43.75
Dummy	Subset	None	Region	IV	148.68
Dummy	Subset	None	GP, GP squared and Region	Non IV	11.25
Dummy	Subset	None	GP, GP squared and Region	IV	265.68
Dummy	Subset	None	GP, GP Squared and Urban	Non IV	-8.96
Dummy	Subset	None	GP, GP Squared and Urban	IV	148.77
Dummy	Subset	Urban	None	Non IV	-35.13
Dummy	Subset	Urban	None	IV	-12.59
Dummy	Subset	Urban	GP and GP squared	Non IV	6.81
Dummy	Subset	Urban	GP and GP squared	IV	218.27
Dummy	Subset	Urban	Region	Non IV	-30.15
Dummy	Subset	Urban	Region	IV	164.70
Dummy	Subset	Urban	GP, GP Squared and Urban	Non IV	13.84
Dummy	Subset	Urban	GP, GP Squared and Urban	IV	160.42
Dummy	Subset	GP and GP Squared	None	Non IV	-10.51
Dummy	Subset	GP and GP Squared	None	IV	23.43
Dummy	Subset	GP and GP Squared	Urban	Non IV	10.33
Dummy	Subset	GP and GP Squared	Urban	IV	117.20
Dummy	Subset	GP and GP Squared	Region	Non IV	-0.99
Dummy	Subset	GP and GP Squared	Region	IV	191.19
Spline	All	None	None	Non IV	377.52
Spline	All	None	None	IV	383.46
Spline	All	None	GP and GP squared	Non IV	443.19
Spline	All	None	GP and GP squared	IV	443.79
Spline	All	None	Urban	Non IV	427.93
Spline	All	None	Urban	IV	431.15
Spline	All	None	Region	Non IV	418.52
Spline	All	None	Region	IV	446.14
Spline	All	None	GP, GP squared and Region	Non IV	487.61
Spline	All	None	GP, GP squared and Region	IV	497.47
Spline	All	None	GP, GP Squared and Urban	Non IV	471.59
Spline	All	None	GP, GP Squared and Urban	IV	502.93

Table A-2: Likelihood - Property Crime

Model Specification					Log Likelihood
Formulation of the Laws	Demographics	Additional Covariates	Parameter Heterogeneity	Instrumental Variables	
Spline	All	Urban	None	Non IV	396.97
Spline	All	Urban	None	IV	397.55
Spline	All	Urban	GP and GP squared	Non IV	456.96
Spline	All	Urban	GP and GP squared	IV	463.45
Spline	All	Urban	Region	Non IV	432.03
Spline	All	Urban	Region	IV	460.22
Spline	All	Urban	GP, GP Squared and Urban	Non IV	483.07
Spline	All	Urban	GP, GP Squared and Urban	IV	523.32
Spline	All	GP and GP Squared	None	Non IV	405.97
Spline	All	GP and GP Squared	None	IV	409.13
Spline	All	GP and GP Squared	Urban	Non IV	454.61
Spline	All	GP and GP Squared	Urban	IV	463.59
Spline	All	GP and GP Squared	Region	Non IV	445.17
Spline	All	GP and GP Squared	Region	IV	478.43
Spline	Subset	None	None	Non IV	136.21
Spline	Subset	None	None	IV	147.18
Spline	Subset	None	GP and GP squared	Non IV	213.20
Spline	Subset	None	GP and GP squared	IV	219.69
Spline	Subset	None	Urban	Non IV	177.40
Spline	Subset	None	Urban	IV	187.87
Spline	Subset	None	Region	Non IV	203.34
Spline	Subset	None	Region	IV	248.28
Spline	Subset	None	GP, GP squared and Region	Non IV	253.79
Spline	Subset	None	GP, GP squared and Region	IV	269.16
Spline	Subset	None	GP, GP Squared and Urban	Non IV	263.57
Spline	Subset	None	GP, GP Squared and Urban	IV	329.28
Spline	Subset	Urban	None	Non IV	148.28
Spline	Subset	Urban	None	IV	158.74
Spline	Subset	Urban	GP and GP squared	Non IV	222.91
Spline	Subset	Urban	GP and GP squared	IV	240.99
Spline	Subset	Urban	Region	Non IV	206.98
Spline	Subset	Urban	Region	IV	262.17
Spline	Subset	Urban	GP, GP Squared and Urban	Non IV	276.46
Spline	Subset	Urban	GP, GP Squared and Urban	IV	339.96
Spline	Subset	GP and GP Squared	None	Non IV	176.75
Spline	Subset	GP and GP Squared	None	IV	185.47
Spline	Subset	GP and GP Squared	Urban	Non IV	207.90
Spline	Subset	GP and GP Squared	Urban	IV	226.78
Spline	Subset	GP and GP Squared	Region	Non IV	230.26
Spline	Subset	GP and GP Squared	Region	IV	285.39
Hybrid	All	None	None	Non IV	381.07
Hybrid	All	None	None	IV	394.60
Hybrid	All	None	GP and GP squared	Non IV	446.83
Hybrid	All	None	GP and GP squared	IV	565.83
Hybrid	All	None	Urban	Non IV	421.54
Hybrid	All	None	Urban	IV	477.27
Hybrid	All	None	Region	Non IV	416.40
Hybrid	All	None	Region	IV	543.32
Hybrid	All	None	GP, GP squared and Region	Non IV	488.09
Hybrid	All	None	GP, GP squared and Region	IV	631.10
Hybrid	All	None	GP, GP Squared and Urban	Non IV	476.75
Hybrid	All	None	GP, GP Squared and Urban	IV	590.70
Hybrid	All	Urban	None	Non IV	389.60
Hybrid	All	Urban	None	IV	409.74
Hybrid	All	Urban	GP and GP squared	Non IV	460.52
Hybrid	All	Urban	GP and GP squared	IV	577.34
Hybrid	All	Urban	Region	Non IV	439.30
Hybrid	All	Urban	Region	IV	554.43
Hybrid	All	Urban	GP, GP Squared and Urban	Non IV	488.51
Hybrid	All	Urban	GP, GP Squared and Urban	IV	609.02
Hybrid	All	GP and GP Squared	None	Non IV	409.34
Hybrid	All	GP and GP Squared	None	IV	420.11
Hybrid	All	GP and GP Squared	Urban	Non IV	447.47
Hybrid	All	GP and GP Squared	Urban	IV	507.47

Table A-2: Likelihood - Property Crime

Model Specification					Log Likelihood
Formulation of the Laws	Demographics	Additional Covariates	Parameter Heterogeneity	Instrumental Variables	
Hybrid	All	GP and GP Squared	Region	Non IV	449.01
Hybrid	All	GP and GP Squared	Region	IV	566.61
Hybrid	Subset	None	None	Non IV	141.73
Hybrid	Subset	None	None	IV	159.32
Hybrid	Subset	None	GP and GP squared	Non IV	219.68
Hybrid	Subset	None	GP and GP squared	IV	359.91
Hybrid	Subset	None	Urban	Non IV	176.36
Hybrid	Subset	None	Urban	IV	244.31
Hybrid	Subset	None	Region	Non IV	210.91
Hybrid	Subset	None	Region	IV	336.90
Hybrid	Subset	None	GP, GP squared and Region	Non IV	258.89
Hybrid	Subset	None	GP, GP squared and Region	IV	421.51
Hybrid	Subset	None	GP, GP Squared and Urban	Non IV	282.32
Hybrid	Subset	None	GP, GP Squared and Urban	IV	398.91
Hybrid	Subset	Urban	None	Non IV	150.94
Hybrid	Subset	Urban	None	IV	167.23
Hybrid	Subset	Urban	GP and GP squared	Non IV	233.37
Hybrid	Subset	Urban	GP and GP squared	IV	368.08
Hybrid	Subset	Urban	Region	Non IV	221.06
Hybrid	Subset	Urban	Region	IV	341.10
Hybrid	Subset	Urban	GP, GP Squared and Urban	Non IV	288.24
Hybrid	Subset	Urban	GP, GP Squared and Urban	IV	407.67
Hybrid	Subset	GP and GP Squared	None	Non IV	175.73
Hybrid	Subset	GP and GP Squared	None	IV	198.18
Hybrid	Subset	GP and GP Squared	Urban	Non IV	217.74
Hybrid	Subset	GP and GP Squared	Urban	IV	280.53
Hybrid	Subset	GP and GP Squared	Region	Non IV	231.38
Hybrid	Subset	GP and GP Squared	Region	IV	371.20

Note: Each line corresponds to a specific model. Every specification includes time dummies instead of region trends and assumes errors are homoskedastic. GP stands for gun prevalence. For a description of the various modeling assumptions see Section 3.