# 11. Empirical analysis of regulation: the promise of field experiments in China

**Jonathan Klick**

## 1. INTRODUCTION

The "empirical revolution" has been the most exciting and important development in law and economics since the beginning of the twenty-first century. Law and economics went from being a field where theory and intuition dominated, to becoming a more mature social science where empirical testing is central in its scholarship. This approach to research holds the prospect of providing useful insights regarding the effects of policies and regulations. The combination of sophisticated legal and institutional knowledge with systematic empirical methods provides reliable guidance to policymakers.

The movement toward empirical work in law and economics was driven by improvements in research designs that allowed for more credible causal inferences to be made. These research designs rely primarily on so-called "natural experiments" to avoid the omitted variable bias problem that plagued earlier empirical work. Such methods, however, often require a stroke of luck as they exploit unplanned quasi-randomizations, leaving researchers with little to go on if proverbial lightning doesn't happen to strike in the policy domain where evaluations are needed. In these cases, we are back to theory, intuition, and poorly identified empirical work to provide policy advice that is often severely flawed.

Field experiments provide a promising alternative to the natural experiment approach, and China, with its large population, numerous jurisdictions, and active bureaucracy is well suited to implement systematic policy evaluations in the field. Such a program has the advantages of allowing for policy evaluations in real world settings without having to rely on the imperfect randomization that is studied in natural experiments. Additionally, field studies can be implemented on demand as opposed to relying on the serendipity that plays such a large role in quasi-experimental designs.

In this chapter I discuss the empirical revolution in law and economics, focusing on how it has made credible contributions to our understanding of the effects of regulations. I then discuss some of the shortcomings of the quasi-experimental approach that is at the core of modern empirical law and economics. Next I discuss how China could use field experiments to get an even more precise understanding of how regulations are likely to affect outcomes and behavior, providing a general roadmap of how such studies could be implemented.

## 2. THE EMPIRICAL REVOLUTION

While law and economics has been recognized as one of the most important innovations in legal scholarship during the second half of the twentieth century, for most of its history its tools were almost exclusively those of microeconomic theory. Posner's seminal *Economic Analysis of Law* had virtually no references to empirical work in its 1973 first edition; even today's eighth edition has a low ratio of empirical references relative to the theory works cited. The famous law and economics names – Coase, Posner, Calabresi, Shavell, Polinsky, Cooter, and so on – are all primarily theorists, and the standard textbooks in the field are basically silent regarding any data related to the topics discussed in each chapter. Relatively few law and economics scholars in US law schools prior to 2000 did empirical work. In discussing the roots of this absence of empiricists, Bill Landes (one of the few members of the older generation who does some empirical work) noted, "[M]ost law professors regard empirical work as a form of drudgery not worthy of first-class minds."[1]

Part of what motivated this institutional preference against empirical work was almost surely the fact that much empirical work that was being done was simply not very good. The practice of the time was to attempt to isolate causal relationships between laws and behavior or regulations and outcomes by trying to control for as many confounding factors as was convenient. This so-called control function approach, in principle, is perfectly reasonable. If it is literally the case that a researcher controls for every factor involved in generating an outcome in the regression, then all estimated coefficients can be interpreted causally. However, in practice this approach is doomed to fail. First, no researcher can know all of the factors that influence an outcome. Second, even if he did, many of those factors will be difficult, if not impossible, to quantify. Lastly,

1 Landes (2003).

even if the factors are known and quantifiable, data availability problems would often impede the researcher from being able to carry out this ideal regression.

Although everyone was aware of these problems, most empirical research was carried out with the implicit belief that effort was enough to overcome these problems. A good empirical paper was one that simply controlled for more stuff than the last paper written on the topic. While correlation wasn't causation, correlation conditional on a subset of the most important covariates (which was defined, roughly, as whatever data the researcher could find) was close enough for academic work. In such an environment, empirical papers were written, but it was wise not to believe their results. Recognizing this, Ed Leamer pleaded for increased transparency and humility, calling the standard practice a "con."[2] But since not much credence was given to the work anyway, con doesn't seem to be the right word. Perhaps more accurately (if metaphorically), empirical work was like masturbation: after a little work, and a little mess, you could see an outcome, even if it wasn't all that fulfilling – and it was probably better kept private. Given that, maybe it's not too surprising that law schools were not enamored with empirical work.

The situation began to change in the mid to late 1990s, riding the wave of what Josh Angrist has dubbed the "credibility revolution,"[3] empirical law and economics scholars started exploiting "natural" or "quasi" experiments to identify the causal effects of laws and regulations on behavioral outcomes. In addition to yielding more credible outcomes, this new approach is more intuitively accessible, combining to make the work and the scholars doing it more attractive to law school faculties.[4]

The fundamental problem of empirical studies using observational (that is, non-experimental) data is that they will contain some variety of the omitted variable bias.[5] This problem goes by many names – endogeneity, simultaneity, self-selection, reverse causality, and so on – but it all boils down to leaving out important variables that confound the estimation of the causal effect of interest (e.g., the effect of regulation x on behavioral outcome y). Formally, this problem arises when the research leaves out some variable (or variables) that helps determine the outcome (after all the effects of all of the included variables are controlled for) and the left-out

2 Leamer (1983).
3 Angrist and Pischke (2010).
4 For more on this aspect, see Klick (2011).
5 For a more detailed discussion of these issues, see Gelbach and Klick (2014).

variable (or variables) is correlated with the policy variable of interest (after controlling for the effects of all of the included variables).

This deceptively simple problem is ubiquitous. For virtually all outcomes, it is not clear a priori to the researcher which variables determine the outcome of interest, and it is equally unlikely that the researcher knows what subset of those variables is correlated with the policy of interest. Further, even if such knowledge existed, many of the relevant variables would be unquantifiable, or at least the researcher would not have data for all of the variables.

In the old approach, after controlling for whatever variables were at hand, researchers would suggest that because more could not be done, the estimated results were as good as it gets, and in any case, since the current approach controls for more stuff than the last paper on the topic, it is better. Unfortunately, there is an empirical analog to the theory of the second best.[6] If some but not all of the omitted variables leading to an omitted variable bias are controlled for, including more of those omitted variables can move the parameter estimate closer to or further away from the unbiased value of the parameter relative to the estimate that arises when fewer of the omitted variables are controlled for. What's worse, it is not possible to know whether the estimate has gotten better or worse.

Others in the old approach would begrudgingly admit the bias in their estimates, but then would provide some intuition for the direction, and perhaps even the magnitude, of the bias so as to be able to give a rough correction of the estimate. This too was largely folly, since any such correction requires knowledge of the conditional correlations between the omitted variables and the policy of interest, as well as the conditional correlations with the outcome being studied. Essentially this is question-begging, since, by and large, the entire reason the omitted variable bias arose in the first place was because the researcher did not have the relevant data. Thus, making claims about conditional correlations of non-existent data would seem to be simply making things up.

The newer, more credible[7] approach exploits a natural or quasi-randomizing shock which affects the treatment status (i.e., whether some entity is affected by a policy-relevant variable) of individuals (or firms, jurisdictions, etc.) and examines how some behavioral outcome changes

---

[6] Lipsey and Lancaster (1956).

[7] For an excellent and wide-ranging discussion of how much more credible the new work actually is, see the symposium "Con Out of Economics" in the Spring 2010 issue of the *Journal of Economics Perspectives*, which, in addition to the Angrist and Pischke piece, has important critiques from Ed Leamer and a number of others.

for those affected after the shock relative to their behavior prior to the shock. Further, to net out any background trend, the same before and after comparison is made for similar entities that are not affected by the shock. This comparison group is the analog to the "control" group in a true randomized experiment. Because there is no true randomization, special care must be taken to ensure that the chosen comparison group is indeed comparable.[8]

When the examined shock occurs just once, the study is often referred to as a differences-in-differences study. If various shocks occur over time, the general approach is referred to as a fixed effects model (which can be thought of as estimating the policy coefficient of interest as averaging over a series of differences-in-differences comparisons).

If the shock is plausibly random (relative to the behavioral outcome of interest) and if the comparison groups are plausibly comparable, the estimated treatment effect is plausible (or credible) as well. From a communication standpoint, this methodological approach is much more easily explained to a non-expert audience than discussing the control function approach. From the perspective of reliability, this approach relies on seemingly more plausible assumptions than the control function approach, or at least more transparent assumptions, as it is fairly simple for an audience to consider the assertion that the shock is random and that the comparison groups are sufficiently similar.

This approach has led to successes in a number of literatures that have, in turn, improved actual public policy. One such literature is the work on police and crime. While the old work in this topic generated crazy results that were widely dismissed,[9] the quasi-experimental work[10] has been very well received.

## 3. THE VALUE OF REAL EXPERIMENTS

However, despite the successes of the natural or quasi-experimental approach, there are at least two drawbacks. First, as discussed above,

[8] This is easier said than done. The necessary comparability applies both to observable and unobservable characteristics. In an actual randomized experiment, the randomization ensures this. In a natural or quasi-experimental design, this is necessarily an assumption, since comparability on unobservable characteristics cannot be verified.

[9] For a survey of the older work, see Cameron (1988).

[10] See, for example, Klick and Tabarrok (2005), MacDonald, Klick and Grunwald (2016) and Evans and Owens (2007).

the assumptions that the shock is truly conditionally random and that the comparison groups are sufficiently similar are untestable. While it is true that every research design relies on some assumptions, this surely limits confidence in this approach. Second, this approach relies on lucky happenstance in terms of finding a usable shock that is related to the research question. There will be some questions that simply do not benefit from the required shock, but need to be addressed anyway.

Actual experiments have the potential to address these shortcomings. In a truly randomized experiment, the first concern goes away since the policy treatment is randomized. Further, since the researcher himself is running the experiment, there is no longer the need to wait for lightning to strike – instead, the researcher throws the bolt down himself, and he can craft the shock to match whatever policy needs to be studied. The shortcoming of this kind of experiment, however, is one of external validity. How do we know that people in the lab act like people in the real world? The setting itself might make a difference. Further, if the people studied in the lab are different than the people to whom the policy will actually be applied, there is another concern about the relevance of the experimental results for guiding policy.

Field experiments have the potential to mitigate both of these external validity concerns. By randomizing the policy implementation (i.e., randomly choosing jurisdictions to serve as the treatment units and using the other jurisdictions as the comparison units) in the real world setting, one can get the best of both worlds – the internal validity (or reliability) of experiments and the external validity (or relevance) of the observational studies. John List has written extensively on the promise of field experiments and their implementation.[11] More generally, the value of learning from this kind of approach is highlighted by Abramowicz, Ayres, and Listokin, who advocate "randomizing law."[12]

China, in many ways, is uniquely suited to engage in systematic testing of regulations and policies. Its large population and numerous jurisdictional units make statistical power a foregone conclusion. Further, its well-functioning bureaucracy expedites implementation of and adherence to experimental designs. Taking such an approach to regulatory development could put China in a position where it can not only improve its own policies but also teach other countries about what works and what doesn't when it comes to policy design.

---

[11] See http://home.uchicago.edu/~jlist/research2/methodology.html.

[12] Abramowicz, Ayres and Listokin (2011).

## 4. CONCLUSION

Empirical policy analysis has improved substantially since the beginning of the twenty-first century. The use of natural or quasi-experimental research designs has improved the credibility and, therefore, usefulness of law and economics scholarship. However, these approaches are somewhat limited in that they rely on shocks that are outside of the researcher's control. This means that sometimes it is just not possible to study a particular policy of interest. Field experiments have the potential to remedy this shortcoming, and China may be uniquely suited to leverage field experiments to provide important insights for policymakers.

## REFERENCES

Abramowicz, Michael, Ian Ayres and Yair Listokin (2011), "Randomizing Law," *University of Pennsylvania Law Review*, 159(4), 929–1005.

Angrist, Joshua D. and Jörn-Steffen Pischke (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics," *Journal of Economic Perspectives*, 24(2), 3–30.

Cameron, Samuel (1988), "The Economics of Crime Deterrence: A Survey of Theory and Evidence," *Kyklos*, 41(2), 301–23.

Evans, William N. and Emily G. Owens (2007), "COPS and Crime," *Journal of Public Economics*, 91, 181–201.

Gelbach, Jonah B. and Jonathan Klick (2014), "Empirical Law and Economics" in Francesco Parisi (ed.), *Oxford Handbook of Law and Economics*, available at http://ssrn.com/abstract=2507324.

Klick, Jonathan (2011), *The Empirical Revolution in Law and Economics*, Inaugural Lecture for Erasmus Chair in Empirical Legal Studies (Erasmus Law Lectures), Eleven International Publishing.

Klick, Jonathan and Alexander Tabarrok (2005), "Using Terror Alert Levels to Estimate the Effect of Police on Crime," *Journal of Law & Economics*, 48(1), 267–79.

Landes, William M. (2003), "The Empirical Side of Law & Economics," *University of Chicago Law Review*, 70(1), 167–80.

Leamer, Edward E. (1983), "Let's Take the Con out of Econometrics," *American Economic Review*, 73(1), 31–43.

Lipsey, R.G. and Kelvin Lancaster (1956), "The General Theory of Second Best," *Review of Economic Studies*, 24(1), 11–32.

MacDonald, John M., Jonathan Klick and Ben Grunwald (2016), "The Effect of Private Police on Crime: Evidence from a Geographic Regression Discontinuity Design," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 179(3), 831–46.