

CHRISTOPH ENGEL*

Challenges in the Interdisciplinary Use of Comparative Law†

The world has more than 200 states. Many states are federations and hence consist of multiple jurisdictions. Seemingly there is thus ample room for a social science approach to comparative law. In this perspective, each legal order produces a data point. Variance in the solutions adopted by different legal orders is used as evidence that a certain legal design causes greater justice, better political stability, higher welfare, or more equity. The results could motivate the strife for legal betterment, by the way of legal transplants.

This Article cautions against the dangers inherent in this empirical enterprise. In a nutshell, the danger results from the fact that mere correlation (some jurisdictions are associated with some outcomes) is not causation (a difference in legal design is responsible for the difference in outcomes). Yet for choosing between alternative legal regimes, causation would be critical. The Article explains why comparative law is a conspicuously challenging source of empirical evidence. It discusses possible solutions.

INTRODUCTION

Interdisciplinarity is akin to foreign trade. If a country with a well-educated and equally motivated labor force imports food, this does not mean that no food could be grown in that country. Its population was not starving when international trade was less developed.

* Dr. Christoph Engel, Max Planck Institute for Research on Collective Goods, Bonn, Germany, and University of Bonn, as well as Erasmus University Rotterdam, engel@coll.mpg.de. Helpful comments by Holger Spamann and Amber Lynch are gratefully acknowledged.

† <https://doi.org/10.1093/ajcl/avab020>

But the makeup of its labor force gives the country a competitive advantage. If the labor force of a food-producing country is less advanced, the exporting and the importing country stand to gain from the exchange. The importing country may devote precious human capital to purposes that are more productive than growing plants. And the exporting country gets access to goods or services that require sophisticated input. This is the core of the theory of comparative advantage.¹

Arguably, comparative law and the social sciences are in the same situation. If a social scientist is interested in cross-country comparisons, she could, of course, start by building her own dataset. As most of social life is determined, shaped, or at least moderated by law, the quality of such data would benefit from taking legal factors into account. As law, to a large extent, is still a national affair, hardly any rule is uniform worldwide and there is interesting variance which the social scientist may want to exploit. Yet few social scientists have legal training and risk misinterpreting the legal factors of interest. Of course, individual social scientists could specialize in these factors and become better over time. But there is a rich field of comparative law that has already built this expertise. This expertise seems to be the comparative advantage of legal scholars.

On the other hand, legal scholars, and comparatists in particular, tend to be interested in the comparative performance of alternative doctrinal solutions. But the typical comparative lawyer has no social-science training, which makes it hard for her to assess the probative value of comparing different outcomes in countries with different legal solutions to the same normative problem. Of course, individual comparative lawyers might be able to acquire this expertise. But this would be time consuming and would divert their effort from what they are best at doing, i.e., from a painstaking analysis of minute, yet highly relevant, legal detail. To pick up the metaphor of international trade again, the translation of comparative legal work into data that lends itself to quantitative analysis, and the execution of that analysis, might be the comparative advantage of social scientists interested in cross-country comparisons.

Hence there is ample room for mutually beneficial cross-disciplinary trade. However, as others have documented, such trade is still conspicuously rare.² It would be a worthwhile exercise to search for “barriers to trade.”³ In the long run, institutional attempts at

1. The theory has originally been developed in DAVID RICARDO, *ON THE PRINCIPLES OF POLITICAL ECONOMY, AND TAXATION* (J. Milligan London, 1819); for a modern reprint, see DAVID RICARDO, *THE WORKS AND CORRESPONDENCE OF DAVID RICARDO* (Cambridge University Press 1981). For a modern treatment, see AVINASH DIXIT & VICTOR NORMAN, *THEORY OF INTERNATIONAL TRADE: A DUAL, GENERAL EQUILIBRIUM APPROACH* (1980).

2. Nuno Garoupa & Thomas Ulen, *Comparative Law and Economics: Aspirations and Hard Realities*, 69 AM. J. COMP. L. 664 (2021); Holger Spamann, *Empirical Comparative Law*, 11 ANN. REV. L. & SOC. SCI. 131 (2015).

3. For a repertoire of barriers to trade, see EUR. COMM’N, *REPORT FROM THE COMMISSION TO THE PARLIAMENT AND THE COUNCIL ON TRADE AND INVESTMENT BARRIERS 1 JANUARY 2019–31 DECEMBER 2019* (2020), trade.ec.europa.eu/doclib/html/158789.htm.

removing some of these barriers would be beneficial. Yet the aim of this Article is more limited. I want to explore the scope of the enterprise: How much room is there for the rigorous, in particular quantitative, analysis of legal institutions across jurisdictions?

I am bracketing theory. Learning about variance in legal institutions, or in legal discourse for that matter, may well inspire social scientists to develop new paradigms and to generate testable propositions.⁴ Yet for the purposes of this Article, I assume that a social scientist has already generated such a proposition. She is intrigued by the observation that legal rules, as well as legal practice applying them, seem to differ across countries in ways that could help her put a theoretical claim to the test.

In this Article I discuss why this seemingly obvious empirical approach is much harder than one might have thought.⁵ Let me stress at the outset that I do not claim that the enterprise is futile, and comparative lawyers and social scientists should stay in their respective silos. If the exchange between both camps is to gain momentum, the considerable methodological challenges should, however, not be swept under the carpet. There is much to be gained from collaboration. However, the social-science importers of comparative law must think hard about containing the risks, as must comparative lawyers, if they intend to exploit the power of social-science methods, and of quantitative analysis in particular. At the least, limitations must be acknowledged. And to the extent that the risks remain substantial, alternative research approaches are worth considering.

I. CAUSAL CLAIMS

At their core, the empirical social sciences test causal claims. In a prototypical design, some theory predicts that some input causes some output. A classic illustration is positive economic theory,⁶ or what political scientists and sociologists call the rational choice paradigm.⁷ The theory is consequentialist.⁸ It builds on the assumption that individuals choose between alternative actions anticipating their effects, and

4. For one such attempt, see Joseph A. Conti, *Relational Sociology and Comparative Law*, 69 AM. J. COMP. L. XX (2021).

5. The paper most similar in spirit, and equally cautious, is Spamann, *supra* note 2.

6. For a comparative lawyer interested in better understanding the paradigm, law and economics theory would be a good starting point. At its core, the theory embraces the paradigm (maybe, in more recent years, allowing for behavioral qualifications). The introductory texts to law and economics are written with the intention of engaging lawyers without modeling training. This makes these texts more accessible than treatises targeting economics students. And the paradigm is applied to legal applications. Good introductions include RICHARD A. POSNER, *ECONOMIC ANALYSIS OF LAW* (6th ed. 2014); ROBERT COOTER & THOMAS ULEN, *LAW AND ECONOMICS* (4th ed. 2012).

7. For a summary treatment, see THE HANDBOOK OF RATIONAL CHOICE SOCIAL RESEARCH (Rafael Wittek et al. eds., 2013).

8. For the history and the conceptual foundations of consequentialism, see JULIA DRIVER, *CONSEQUENTIALISM* (2011).

evaluating them according to the benefit they expect to accrue, net of the cost or risk involved. As the field of law and economics has shown over decades, this is by no means the only way, although in many contexts a thought-provoking one, of analyzing legal institutions.

A. *An Illustration: Does Tort Law Deter Rule Violations?*

To illustrate, one might be interested in the effect of liability rules on behavior, say knowingly damaging foreign property. Rational choice theorists would assume that the risk of being held liable is anticipated. If the expected loss resulting from having to compensate the victim exceeds the benefit of harming her, the potential tortfeasor would refrain from the action.⁹ Hence within the rational choice paradigm, a deterrence perspective on torts invites itself.¹⁰ In fact, the theory is even more precise. If the expected loss from having to compensate the victim is smaller than the benefit, the theory would predict that the prospect of liability does not affect the choice of the activity level, or of precautionary measures. This expectation is itself a theoretical construct. In the simplest model, the theorist additionally assumes that the decision maker does not worry about risk (is risk neutral). Then the amount of compensation can be multiplied with the probability of enforcement. This precision of the theory is desirable as it yields a clear prediction. There should be no difference in tortious behavior if either there is no liability or if the expected value of compensation is below the benefit of breaking the rule.

Now the tort law in different legal orders has embraced deterrence as one of the goals to a different degree.¹¹ In the United States, punitive damages have a long history.¹² They can be interpreted as a technique for increasing the severity of the sanction, and thereby, the deterrent effect of torts. By contrast, civil law jurisdictions have been much more hesitant to allow the amount of damages to exceed the harm suffered by the plaintiff.¹³ A social scientist interested in testing the deterrence effect of tort law may therefore be tempted to compare some measure for tortuous behavior between the United States and, say, Germany.

This is when interaction with a comparative legal scholar would be highly beneficial. In the United States, tort law is state law, and

9. For a thorough treatment, see URS SCHWEIZER, *SPIELTHEORIE UND SCHULDRECHT* (2015).

10. The canonical model on deterrence is Gary Stanley Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968).

11. See further André Tunc, *Introduction to 11 INTERNATIONAL ENCYCLOPEDIA OF COMPARATIVE LAW: TORTS* 154 (André Tunc ed., 1974); Gary T. Schwartz, *Reality in the Economics of Tort Law: Does Tort Law Really Deter?*, 42 UCLA L. REV. 377 (1994); Gary T. Schwartz, *Mixed Theories of Tort Law. Affirming Both Deterrence and Corrective Justice*, 75 TEX. L. REV. 1801 (1996).

12. See Theodore Eisenberg, *Measuring the Deterrent Effect of Punitive Damages*, 87 GEO. L.J. 347 (1998); Cass R. Sunstein et al., *Assessing Punitive Damages*, 107 YALE L.J. 2071 (1998).

13. HELMUT KOZIOL & VANESSA WILCOX, *PUNITIVE DAMAGES: COMMON LAW AND CIVIL LAW PERSPECTIVES* (2009).

differs between states. In Germany, the general clause in Article 823 of the German Civil Code (Bürgerliches Gesetzbuch) is complemented by a host of more specific rules, scattered over multiple statutes.¹⁴ Jurisprudence has allowed for exceptions to the general rule that compensation may not exceed the harm, for instance in copyright violations.¹⁵ Not infrequently, legal orders use very different doctrinal paths to reach similar decisions in similar cases.¹⁶ Closely interacting with a comparative lawyer may therefore help the social scientist to avoid comparing apples and oranges.

B. *Correlation Is Not Causation*

Unfortunately, the choice of the object of comparison is by far not the only hazardous part of the enterprise. Using evidence from multiple jurisdictions to test a theoretical claim almost epitomizes the panoply of concerns to be taken seriously when testing a causal claim.¹⁷ In the following, I will make the reader understand why the enterprise is fraught with risk. Some of these risks are routinely discussed in the social sciences, and should therefore encourage any social scientist to be cautious when using a difference in legal rules across jurisdictions as a source of variance. Other concerns are more specific to this very empirical strategy. Unfortunately, these more specific concerns carry even more weight.

Every empiricist worries about *reverse causality*. Her theory predicts that *A* causes *B*. But for the most part all she is able to observe is the cooccurrence of *A* and *B*. If there is a lot of *As*, there are also a lot of *Bs*. But correlation is not causation. The fact that there are so many *Bs* may be exogenous, and the presence of the *Bs* may actually have engendered many *As*. Let me use the tort example for illustration. Under section 4 of the Clayton Act,¹⁸ a private party who has been injured in her business due to a violation of antitrust rules may sue the violator for treble damages, i.e., three times the harm she has suffered. The German Gesetz gegen Wettbewerbsbeschränkungen has no comparable rule. The probability of being sued for an antitrust

14. BÜRGERLICHES GESETZBUCH [BGB] [CIVIL CODE], art. 823, www.gesetze-im-internet.de/englisch_bgb. For a comparative treatment in English, see CEES VAN DAM, EUROPEAN TORT LAW (2013).

15. Bundesgerichtshof [BGH] [Federal Court of Justice] Mar. 10, 1972, 59 ENTSCHIEDUNGEN DES BUNDESGERICHSHOFE IN ZIVILSACHEN [BGHZ] 286 (1972).

16. The doctrine of consideration is a classic. As such, it is not part of civil law. Yet civil law does not hold a promise against the promisor, if the promise is classified as a gift, and common law complements the doctrine with promissory estoppel. Consequently, there are only very few cases (like promising to pay a finder's fee for returning a lost dog) where outcomes would differ. See further DIETRICH ROTHOF, SYSTEM DER IRRTUMSLEHRE ALS METHODENFRAGE DER RECHTSVERGLEICHUNG: DARGESTELLT AM DEUTSCHEN UND ENGLISCHEN VERTRAGSRECHT (1968); FERDINAND FROMHOLZER, CONSIDERATION (1997).

17. See JOSHUA D. ANGRIST & JÖRN-STEFFEN PISCHKE, MOSTLY HARMLESS ECONOMETRICS: AN EMPIRICIST'S COMPANION (2008).

18. 15 U.S.C. § 15.

violation is certainly substantially below 100%. One might therefore expect that German law under-deters.

Let us assume, for the sake of the argument, that a researcher has had the good fortune of compiling a dataset quantifying the damage third parties suffer from violations of antitrust. Let us further assume that such damage is more frequent in Germany, or more severe, than in the US. Does that prove the hypothesis? Not necessarily. It could be that Germany has had fewer antitrust violations in the first place, so that the need for deterrence was less pronounced. This is another instance where collaboration with a comparative lawyer would be beneficial. This lawyer would probably suggest that the alternative explanation is unlikely to be true, given that the Clayton Act is much older than the pertinent provisions of German law. But another channel of reverse causality may be harder to dismiss. It has been argued that the United States is a more litigious society than Germany. Shareholder value is the dominant concept in corporate law.¹⁹ By contrast, in Germany, the corporation has always been understood as an enterprise that has to serve shareholders and stakeholders alike, including the workforce and the physical environment.²⁰ Hence business in the United States might have been in greater need of deterrence than comparable business in Germany.

The concern about *omitted variables* looms equally large. The following illustration stems from a different domain. Children used to be made to believe that babies were brought by storks. German data seems to provide evidence for this causal claim: storks are much more frequent in the eastern part of the country;²¹ this is also where the birth rate is highest.²² However this seeming proof is not a mystery: it points to an omitted variable. Before 1989, the eastern parts of the country were under communist rule. As communism was substantially less successful economically, storks had an easier time finding a habitat. On the other hand, the ideology of the ruling party put a high

19. Luca Enriques et al., *The Basic Governance Structure. Minority Shareholders and Non-shareholder Constituencies*, in *THE ANATOMY OF CORPORATE LAW: A COMPARATIVE AND FUNCTIONAL APPROACH* 79 (Reinier Kraakman ed., 2017).

20. DEUTSCHER CORPORATE GOVERNANCE KODEX [DCGK] [CORPORATE GOVERNANCE CODE], www.dcgk.de/files/dcgk/usercontent/en/download/code/191216_German_Corporate_Governance_Code.pdf.

21. See KAI-MICHAEL THOMSEN, NATURSCHUTZBUND DEUTSCHLAND (NABU), WHITE STORK POPULATIONS ACROSS THE WORLD WEISSSTÖRCHEN IN ALLER WELT (2013), https://bergenhusen.nabu.de/imperia/md/nabu/images/nabu/einrichtungen/bergenhusen/zensus_ergebnisse_2004.pdf (data about the distribution of white storks); *Weißstörche in aller Welt: Ergebnisse des 6. Internationalen Weißstorchzensus 2004/2005*, NABU: MICHAEL-OTTO-INSTITUT BERGENHUSEN, <https://bergenhusen.nabu.de/weisstorch/bestand-international.html> (last visited Nov. 27, 2020) (containing the results of an international stork census).

22. *Regionale Unterschiede in Geburten und Sterblichkeit*, BUNDESINSTITUT FÜR BEVÖLKERUNGSFORSCHUNG, www.bib.bund.de/DE/Fakten/Regional/Geburten-Sterblichkeit.html (last visited Nov. 27, 2020).

value on women being part of the labor force. This led to considerably better childcare, encouraging more women to have children.

Omitted variable bias must also be taken seriously when using a difference in some economic or societal outcome between two countries as evidence that a difference between the laws of the countries is causal. No two countries are the same. Often, some of these differences can be observed. For instance, one may have information about the distribution of income or about political stability. With this information, one can condition the effect of interest, i.e., of a difference between legal rules, on other, potentially relevant factors.²³ But it is not at all rare that one suspects further influences, but has no data to back them up. Then the difference in outcomes cannot be taken as support for the theoretical claim.

Let me again illustrate the concern with the question whether the obligation to compensate victims deters firms from breaking antitrust law. Let us assume that the difference in outcomes between the United States and Germany is too small to be meaningful.²⁴ Can the researcher infer that deterrence is not important in antitrust law? The prospect of having to compensate the victim would be non-deterrent if (i) there is at least a small probability that victims will not sue, or that they might lose in court, and (ii) the harm suffered by the suing victims is no larger than the benefit of breaking the law. Then the firm may still expect to make at least a small profit if it violates the law. If the firm maximizes profit, and cares exclusively about the risk of litigation, it breaks the law. Yet private litigation is not the only intervention to enforce antitrust rules. There is also public enforcement by antitrust authorities, and the violation of the rules can constitute a crime. The overall deterrent effect is a cumulative result of all enforcement activities.²⁵ Again the social scientist would be well advised to collaborate with a legal scholar who could inform her about these additional enforcement channels. And it would be necessary to quantify the deterrent effect achieved on these other channels to find out whether there is a residual effect of the prospect of private litigation.

A related concern is *selection*. The researcher wants to make a statement about the effect at the level of the population. However, she usually only has access to a sample. If this sample is randomly selected from the population, and sufficiently large, the researcher may be confident that the effect on the sample represents the effect on

23. In statistical jargon, one can “control for” these alternative explanations.

24. I.e., that it does not meet the conventional level for statistical significance. In the social sciences, this level is conventionally set at 5%. Technically, this is the probability of obtaining test results at least as extreme when redrawing an infinite number of samples from the same population, provided the null hypothesis is true. Hence this p-value defines the risk of a false positive result (one rejects the null hypothesis, although there is no support for the alternative hypothesis in the population).

25. For the United States, see Robert H. Lande & Joshua P. Davis, *Comparative Deterrence from Private Enforcement and Criminal Enforcement of the US Antitrust Laws*, 2011 BYU L. REV. 315.

the population at large. Yet this confidence breaks down if the sample is not representative of the population. The researcher is then led to believe in an effect, or the absence of an effect. But she actually can only make a statement about the subgroup of the population that she had a chance to study. Quite often, such a statement is not relevant. In fact, selection can be interpreted as omitted variable bias. The researcher has neglected to control for the fact that her sample has specific characteristics.

Selection is pervasive if data generated by the judiciary is used as evidence.²⁶ People do not sue at random. Among the already selective set of cases that are filed in the first place, those that make it to superior courts are again not randomly selected. Frequently, only the latter are reported and can be used as evidence. Litigation is strategic, on the part of the parties,²⁷ and sometimes even on the part of the courts.²⁸

Selection is of particular concern if the researcher wants to compare the effect of legal rules across jurisdictions. Jurisdictions have very different reporting standards.²⁹ What looks like a difference in substantive outcomes, and therefore like support for the hypothesis to be tested, may actually just be the result of different reporting cultures. In this respect, the assistance of a comparative lawyer is particularly urgent. This lawyer should have a good understanding of the judicial systems to be compared, in order to assess bias resulting from a difference in reporting cultures. Merely comparing the available decisions along dimensions covered by the available data will not suffice. Even if the available bodies of cases look sufficiently similar between the jurisdictions under comparison, the rulings may have been reported for very different reasons. For example, one needs to know that the judges in German higher courts used to earn additional money for handing in decisions to commercial reporters. They thus had a business interest in the composition of the body of published cases.

In its bluntest form, *measurement error* has featured prominently in the scholarly debate about legal differences as a cause for the differences in economic performance. The “legal origins” literature has claimed that common law is more responsive to the needs of the economy, and therefore has better protected shareholders, which in turn has helped the economies in common law countries to grow more steadily.³⁰ The underlying coding of legal orders was undertaken

26. George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1 (1984).

27. Michael Ramsden & Kris Gledhill, *Defining Strategic Litigation*, 4 CIV. JUST. Q. 407 (2019).

28. Daniel Klerman & Greg Reilly, *Forum Selling*, 89 S. CAL. L. REV. 241 (2015); Stefan Bechtold et al., *Forum Selling Abroad*, 92 S. CAL. L. REV. 487 (2018).

29. Cf. Peter Harris, *Ecology and Culture in the Communication of Precedent Among State Supreme Courts, 1870–1970*, 19 LAW & SOC'Y REV. 449 (1985).

30. Rafael La Porta et al., *The Economic Consequences of Legal Origins*, 46 J. ECON. LITERATURE 285 (2008). For an extensive review of this literature, see Garoupa & Ulen, *supra* note 2.

without much access to legal expertise. Once trained comparative lawyers stepped in, and redid the coding properly, the claimed effect largely vanished.³¹

It stands to reason that one cannot come to unbiased conclusions using biased data. It is less obvious why data analysts should be even concerned if there are no signs of bias. This is the situation that the statistical literature refers to when talking about measurement error.³² The explanatory variable of interest can be measured only imprecisely. In technical terms, the initial concern is noise, not bias. Statistical evidence routinely comes with a measure of confidence. In principle, if the explanatory variable can only be measured with error, the job of the statistician becomes only harder. The true effect must be bigger to meet the conventional criterion for confidence (the significance level). Seemingly, not much harm can be done. If the measurement has too much noise, the evidence does not support the theoretical claim. Empiricists typically do not worry too much about being too conservative.

But there is a catch. Very often, statistical analysis is not confined to assessing the effect of one explanatory variable on one outcome variable. Rather, one simultaneously assesses the explanatory power of multiple variables. This approach is particularly appealing in testing for a purported effect of the law. With this procedure, one may run a horse race between alternative (legal or extralegal) explanations. In such richer statistical models, the fact that one possible explanation is only measured with error can indeed cause bias. Because the first, noisy explanation cannot succeed in explaining the variation in the data, the alternative explanations carry more weight, and may well become significant. The analyst is led to believe that the data supports the alternative explanation, although this explanation would have remained insignificant had the first explanation been measured cleanly.

C. *Applying the Tricks of the Empirical Trade*

In a way, a comparative lawyer always compares apples and oranges. It requires little ingenuity to spot differences between the objects of comparison, i.e., the legal orders, which transcend the difference of interest. Ultimately, to prove causality, one would need the counterfactual. The counterfactual is unknown by definition. In the

31. John Armour et al., *How Do Legal Rules Evolve? Evidence from a Crosscountry Comparison of Shareholder, Creditor, and Worker Protection*, 57 AM. J. COMP. L. 579 (2009); John Armour et al., *Shareholder Protection and Stock Market Development: An Empirical Test of the Legal Origins Hypothesis*, 6 J. EMPIRICAL LEGAL STUD. 343 (2009). See contra Ralf Michaels, *Comparative Law by Numbers? Legal Origins Thesis, Doing Business Reports, and the Silence of Traditional Comparative Law*, 57 AM. J. COMP. L. 765 (2009); Holger Spamann, *The "Antidirector Rights Index" Revisited*, 23 REV. FIN. STUD. 467 (2010).

32. See Jonathan Gillard, *An Overview of Linear Structural Models in Errors in Variables Regression*, 8 REVSTAT 57 (2010).

language used in the statistical literature, one and the same case cannot simultaneously be “treated” and remain “untreated.”³³ The standard response of the empirical literature is not available when testing the effect of alternative legal solutions with the help of data from different jurisdictions. Legal rules are not *randomly assigned* to a random selection from the pool of jurisdictions.

An imperfect, but appealing, way out is adding a second dimension to the data. Rather than only considering outcomes from two (or more) jurisdictions at a single point in time, one observes the outcome of interest in all jurisdictions over an extended period of time. In other words, one goes beyond testing a “cross-section,” and exploits the information about development over time in a “panel.”³⁴ This empirical strategy is particularly appealing if, during the observation period, a given rule has changed in one jurisdiction, but not in the other. If the outcome variable changes in the “treated” jurisdiction, while remaining constant in the “untreated” jurisdiction, one may be reasonably confident that the change has indeed been caused by the intervention, i.e., by the entry into force of the rule in question.

Yet even such a “*difference in differences*” design has its limitations.³⁵ The design does not work if, in the first place, before the rule was changed in the “treated” jurisdiction, the outcome variable had exhibited a different development across the investigated jurisdictions. It is this “common trend” prior to the intervention that makes the approach credible. Yet even if the common trend can be established, one should be cautious about interpreting it. What one observes is the effect of a change in rules. Yet the question of normative interest usually refers to the presence or the absence of a rule.³⁶ Anecdotal evidence suggests that this difference matters in the legal domain. This concern is encountered, for example, in criminal law. Legislators have often reacted to a perceived rise in crime rates by making the punishment more severe. Even if, in the short run, the incidence of the crime in question goes down, it may stabilize at the original high level after the community of criminals has gotten used to the more severe rule.³⁷ The effect of changing the rule has not been sustainable.

In fact, the reality of legal reform frequently runs counter to estimating a causal effect in even deeper ways. Legal rules do not just happen to change; they are changed for a reason. Quite often, this reason is not merely an ideological difference between the previous

33. The logic is powerfully explained in ANGRIST & PISCHKE, *supra* note 17.

34. For background, see, e.g., MARNÓ VERBEEK, *A GUIDE TO MODERN ECONOMETRICS* (2008).

35. See further Richard Blundell & Monica Costa Dias, *Alternative Approaches to Evaluation in Empirical Microeconomics*, 44 J. HUM. RESOURCES 565 (2009).

36. For an experimental demonstration of the difference, see Christoph Engel, *A Random Shock Is Not Random Assignment*, 145 ECON. LETTERS 45 (2016).

37. Anthony N. Doob & Cheryl Marie Webster, *Sentence Severity and Crime: Accepting the Null Hypothesis*, 30 CRIME & JUST. 143 (2003).

and the present government. The change in rules is triggered by the perception of a social ill. The amended rule is expected to address the ill. From a social-science perspective, this constitutes reverse causality. The issue is most obvious if the jurisdiction where the rule has changed had been more deeply affected by the normative problem, or at least had been more sensitive to it. Then the treated and the untreated jurisdictions are not comparable. Assessing this source of bias requires fine-grained contextual knowledge, which a comparative lawyer will be much more competent to muster than a social scientist.

Moreover, legal rules are not designed, or changed for that matter, by a benevolent dictator. In a democracy, legislation is regulative policymaking.³⁸ Legal theorists may draw a line between rule generation and rule application. This line marks the difference between a debate *de lege ferenda* and *de lege lata*. If this divide can be taken for granted, the social scientist interested in gauging the effect of the rule need not bother about effects actually resulting from recent intervention, not from a deep difference between jurisdictions. But, in many contexts, one has reason to be skeptical. Some rules remain law on the books forever.³⁹ Others are designed with the intention of making implementation difficult,⁴⁰ for example, when the legislator wants to engage in merely symbolic policymaking.⁴¹ Yet other rules require that the law's subjects avail themselves of the new business opportunity created by the change in rules.⁴² Any of these qualifications make it difficult to pinpoint when, and to what degree, the outcome variable of interest would have to change if the theoretical claim of a causal effect is well founded. This is another instance that calls for the expertise of a comparative lawyer with full command not only of the legal doctrine but also of the implementation practice, in all jurisdictions under comparison—if such a scholar at all exists.

All of the foregoing concerns regarding the identification of a causal effect from observational data should be obvious to a well-trained empirical social scientist. After all, the social sciences have undergone their identification revolution.⁴³ Every graduate school teaches that correlation is not causation and, ideally, prepares the next generation of social scientists not only to discern and acknowledge the limitations of the analysis, but also to create designs that

38. ROLAND CZADA ET AL., *REGULATIVE POLITIK: ZÄHMUNGEN VON MARKT UND TECHNIK* (2013).

39. Roscoe Pound, *Law in Books and Law in Action*, 44 AM. L. REV. 12 (1910).

40. Renate Mayntz, *Implementation von regulativer Politik*, in *IMPLEMENTATION POLITISCHER PROGRAMME II*, at 50 (Renate Mayntz ed., 1983).

41. MURRAY J. EDELMAN, *THE SYMBOLIC USES OF POLITICS* (1964).

42. Christoph Engel et al., *Diffusion of Legal Innovations: The Case of Israeli Class Actions*, 15 J. EMPIRICAL LEGAL STUD. 708 (2018) (demonstrating that the newly extended opportunity to bring class actions in Israel for a long time did not lead to more class action cases).

43. Edward E. Leamer, *Let's Take the Con out of Econometrics*, 23 AM. ECON. REV. 31 (1983).

contain the concern regarding identification.⁴⁴ One powerful way of mitigating the problem is an indirect one.

One can describe the identification problem as a problem of endogeneity.⁴⁵ Ideally, one would want to cleanly separate cause and effect. Now for social reality, it would not be meaningful to postulate natural laws. Social life is too complex to be treated like the phenomenon of gravity. Social scientists take it for granted that they can only observe the causal relationship of interest with a certain degree of noise. This is why a social scientist is not happy to dismiss a causal claim because she has observed a single contradiction. The gold standard is not a perfect mapping between the explanatory variable and the dependent variable. A social scientist accepts a theoretical causal claim, at least for the time being, if it is sufficiently unlikely that the difference in outcomes between treated and untreated observations is the result of noise. She thus accepts residual error. Endogeneity is present if the unexplained part of the data is not just noisy, but informative about the explanatory variable.

This is where the solution starts. One may be in the position to replace the contaminated explanatory variable with another, sufficiently clean variable.⁴⁶ This strategy helps, if the replacement, called an *instrument* in statistics, is sufficiently correlated with the contaminated variable at hand. Of course, unless the correlation is near perfect, one loses explanatory power. Essentially, one only has the correlation between the two variables, rather than the variable of interest. But this is the price one has to pay for removing the potential bias resulting from endogeneity.

Technical as this may sound, this is one of the most productive ways for comparative lawyers and social scientists to engage in interdisciplinary interaction. Finding good instruments is hard. Either the instrument is also correlated to the error term, and then not valid. Or the instrument is truly exogenous, but the correlation between the instrument and the potentially contaminated explanatory variable of interest is too low, which is why the instrument is weak, and therefore useless.⁴⁷ In the social sciences, researchers are rightly rewarded with publications in top journals if they can come up with a convincing instrument that is neither endogenous nor weak.⁴⁸

44. Joshua D. Angrist & Jörn-Steffen Pischke, *The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics*, 24 J. ECON. PERSP. 3 (2010).

45. For background, see ANGRIST & PISCHKE, *supra* note 17.

46. For a very accessible introduction, see Joshua D. Angrist, *Instrumental Variables Methods in Experimental Criminological Research: What, Why and How*, 2 J. EXPERIMENTAL CRIM. 23 (2006).

47. John Bound et al., *Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak*, 90 J. AM. STAT. ASS'N 443 (1995).

48. Note, however, growing skepticism in economics about the ability of researchers to fulfil these conditions: see Abel Brodeur et al., *Methods Matter: P-Hacking and Publication Bias in Causal Analysis in Economics*, 110 AM. ECON. REV. 3643 (2020).

Once a comparative lawyer has understood what is needed for proper instrumentation, she can bring her knowledge to bear. Precisely because the discipline excels in painstaking analysis of minute detail, it is so precious to the social scientist. At the core of the legal-origins literature discussed above there is one such clever idea. Whether a country is classified as a common law or a civil law jurisdiction has been decided a long time ago: in the Middle Ages in Europe, by colonial powers in many other countries of the world. To the extent that true differences between common law and civil law persist to this day, one has indeed found an instrument. One may use the correlation between legal origin and one of the modern-day legal institutions to “instrument” the latter. Hardly any topic has been dissected that thoroughly by comparative lawyers as the differences, apparent and real, between common law and civil law. This past experience makes them well equipped to detect powerful instruments that a social scientist would have a hard time finding on her own.

D. Extrapolating from the Sample of Jurisdictions to the Population of Legal Orders

In the tradition of comparative law, there is great respect for detail. Even if, on the surface, the solutions found by two legal orders seem to differ profoundly, putting these doctrinal constructs into context may well reveal that the differences are negligible. And, conversely, the fact that one legal order seems to use the same doctrinal technique as another is insufficient to conclude that the two legal orders agree on the solution. The latter possibility looms particularly large in legal rules inherited from the colonial past⁴⁹ or introduced by way of a transplant.⁵⁰

From the perspective of quantitative analysis, the concern goes beyond misclassification. The very nature of the exercise is called into question. Statistical analysis requires a *population*. In the respect of interest, one must be able to argue that the population is homogeneous. Then observing a sufficiently large sample, once treated and once untreated, makes it possible to infer how other entities in the population react to the intervention in question. If every entity is different, there is no room for inference “out of sample.” Now, every legal order is a product of history. History does not repeat itself, either within or across countries. The standard critique in (comparative) law can therefore be substantiated in the language of statistics. Comparative law does not seem to be an appropriate object of quantitative analysis.

49. See, e.g., Maryam Kanna, *Furthering Decolonization: Judicial Review of Colonial Criminal Law*, 70 DUKE L.J. 411 (2020).

50. See, e.g., ALAN WATSON, *LEGAL TRANSPLANTS* (1993); Holger Spamann, *Contemporary Legal Transplants: Legal Families and the Diffusion of (Corporate) Law*, 2009 BYU L. REV. 1813.

Yet this radical statement throws out the baby with the bathwater. Statistical analysis has been productively employed in the analysis of individual behavior or of firms in the market. Now, each individual has its own history, as does each firm. To use a term that will be more familiar to comparative lawyers, the critical issue is abstraction. The statistical concept of population is a construct. The empiricist (in practice: implicitly) defines this construct when specifying the empirical model. As with theory, one needs to acknowledge that, as it has been famously argued, all models are wrong, but some are useful.⁵¹ Comparative lawyers should therefore not invest in demonstrating that the respective population is constructed; it always is. They should instead discuss whether the features of individual legal orders that the respective empirical model assumes away leave out elements that may not be neglected for the resulting evidence to be meaningful.

E. The Lack of Independence

Another feature of comparative law is of greater concern. For the interaction between social scientists and comparative lawyers to be productive, some outcome variable of interest must be compared at the level of legal orders. At first glance, a strong case for quantitative analysis can be made. There are more than 200 sovereign states. All of them have a legal order; federal states do even have multiple. Usually, statisticians feel quite comfortable with a few hundred observations. Most of the tests that they use to weed out spurious relationships require asymptotic theory. The tests rely on distributional assumptions. But by the central limit theorem, with enough observations the concern vanishes.⁵² The distribution mechanically approximates normality. If normality can be assumed, the standard statistical procedures for estimating central tendencies and standard errors are reliable.

Yet this line of argument requires a sufficient number of independent observations. If 100 randomly selected individuals receive some intervention, while another 100 individuals randomly selected from the same population are left without intervention, and if the outcomes of the former group differ systematically and sufficiently strongly from the outcomes of the latter group, one may confidently infer an effect at the population level. Yet if the intervention in question is the reaction to the severity of punishing robbery, and the treated group is a monastery, while the control group is a gang, one would not expect to draw a proper image of the intervention at the level of the

51. George E.P. Box, *Science and Statistics*, 71 J. AM. STAT. ASS'N 791 (1976).

52. Imre Bárány & Van Vu, *Central Limit Theorems for Gaussian Polytopes*, 35 ANNALS PROBABILITY 1593 (2007).

population at large. Monastic life impacts criminal behavior, as does gang membership.⁵³

While every empirical researcher would be alerted in such a striking case, dependence problems can be more subtle, and still be concerning. The most important source of dependence in the case of comparative law is what the empirical literature would call contamination.⁵⁴ Before deciding upon the act that is put to the statistical test, the entity in question has come under the influence of some other entity in the sample. For the very reason that legal orders are products of history, this is likely to be the case between jurisdictions. Legal rules are normally not designed from scratch. The standard *modus operandi* of the legislator is not design, but reform. As comparative lawyers have very elegantly worked out, the common origin, for instance in common or Roman law, has left discernible traces on many legal orders.⁵⁵

Moreover, legislators observe one another as well as the developments in the jurisprudence of other countries.⁵⁶ Sometimes they adopt en bloc pre-existing legislation of other jurisdictions. The spread of anti-trust legislation throughout the world, starting with the United States, the European Union, and Germany, is a case in point.⁵⁷ In other instances, the legislator of one country does not simply copy the solution of another country, but rather benefits from the experiences of other jurisdictions.⁵⁸ If the country in question is a democracy, the interplay between government and the opposition is a further channel of influence. Either may point to solutions abroad that seem more conducive to

53. This ties into the critique of universalism proffered by Tom R. Tyler, *The Role of Comparative Law in the Social Sciences: Understanding the Psychology of Social Order*, 69 AM. J. COMP. L. 748 (2021).

54. The topic is most frequently discussed with respect to repeated observations of groups of interacting individuals. From the second period of interaction on, choices are no longer independent. The classic reaction of statisticians is to estimate a mixed effects model. In such a model, the dependence is explicitly modeled: in the most straightforward case, by a random effect at the individual level, crossed with a further random effect at the group level. This specification is appropriate, if one argues that all group level effects are captured by the latter effect (the random effects are orthogonal), and that the group and individual effects are uncorrelated with the explanatory variables (there is no endogeneity). The approach, however, only works if one has sufficiently many, truly independent, units at the highest level of dependence. Otherwise the requirements from asymptotic theory for inference from the sample to the population are not met. See A. Colin Cameron et al., *Bootstrap-Based Improvements for Inference with Clustered Errors*, 90 REV. ECON. & STAT. 414 (2008).

55. See, e.g., KONRAD ZWIEGERT & HEIN KÖTZ, *INTRODUCTION TO COMPARATIVE LAW* (3d rev. ed. 1998).

56. See, e.g., Maximo Langer, *From Legal Transplants to Legal Translations: The Globalization of Plea Bargaining and the Americanization Thesis in Criminal Procedure*, 45 HARV. INT'L L.J. 1 (2004).

57. Keith N. Hylton & Fei Deng, *Antitrust Around the World: An Empirical Analysis of the Scope of Competition Laws and Their Effects*, 74 ANTITRUST L.J. 271 (2007).

58. Consumer protection is a pertinent illustration. See Stephen Weatherill, *Consumer Protection*, in ELGAR ENCYCLOPEDIA OF COMPARATIVE LAW 237 (Jan M. Smits ed., 2d ed. 2012).

furthering their own political convictions.⁵⁹ Yet another path of influence is business. Even if, originally, a country desires to protect its legislative autonomy, pressure from business actors may force it to align its solution with a prominent legislative trend.⁶⁰ As the debate over a California⁶¹ or Brussels effect⁶² has demonstrated, such pressure may even ultimately help establish a more stringent, rather than a more lenient, standard.

From the perspective of quantitative analysis, such cross-fertilization creates a challenge. Unless one can confidently argue that foreign influence has been negligible, the observations from different countries or jurisdictions are no longer independent from one another. As has been explained, this invalidates the established procedures for assessing confidence in quantitative results.

Note that the concern does not go away if the outcome variable is not legal in nature. One may, for instance, be interested in the effect of different inheritance rules on the birthrate, or in the effect of different techniques for consumer protection on the development of online trade. With such research questions, cross-fertilization between legal orders does not have a direct effect on the extra-legal dependent variable. Yet the project is motivated by the hypothesis that outcomes are, at least partly, determined by differences between legal rules. If, in turn, one suspects these rules to be influenced by one another, on this indirect channel there is dependence.

II. SOLUTIONS

No evidence whatsoever is almost certainly worse than evidence with limited probative value. The foregoing should therefore not be misread as a recommendation to refrain from using social science methods. Comparative law can benefit from quantitative methods, as can quantitative social science from the interaction with comparative lawyers. All one should derive from the previous Part is a call for modesty.⁶³ Observing a given outcome in two jurisdictions that differ with respect to legal solutions is not a panacea.

59. A good illustration is the European debate over the desirability of class action. See JÜRGEN G. BACKHAUS ET AL., *THE LAW AND ECONOMICS OF CLASS ACTIONS IN EUROPE: LESSONS FROM AMERICA* (2012).

60. See, for illustration, the debate over accounting standards: Ann Jorissen et al., *A Geographic Analysis of Constituents' Formal Participation in the Process of International Accounting Standard Setting: Do We Have a Level Playing Field?*, 32 *J. ACCT. & PUB. POL'Y* 237 (2013).

61. DAVID VOGEL, *TRADING UP: CONSUMER AND ENVIRONMENTAL REGULATION IN A GLOBAL ECONOMY* (1995).

62. ANU BRADFORD, *THE BRUSSELS EFFECT: HOW THE EUROPEAN UNION RULES THE WORLD* (2020).

63. In this respect, comparative law is just one branch of legal scholarship. Law engages with the lives of real people, and therefore usually cannot afford to artificially simplify the research question, and the data used for responding, in the interest of constructing a well-defined problem. At their core, legal problems are ill defined. This forces legal scholars to face the dilemma between doing justice to the normative

A. *Making Limitations Explicit*

One reaction to the methodological concerns that always exists is making the limitations explicit. Depending on the research question, the possibility of reverse causality may seem far-fetched, and measurement error may be unlikely as well. Dependence problems would probably still persist, and the constructed population may still appear somewhat contrived. But the more the result resonates with a plausible theory, the more one may feel justified pursuing this line of investigation further. Social scientists would call such projects descriptive or exploratory. Strictly speaking, the results would only generate, rather than test, hypotheses. But this may be of value, in particular in directing further research.

Moreover, not every research question requires comparing outcomes at the macro, or jurisdictional, level. If, for instance, one is interested in some determinant of judicial decisionmaking in the lower courts—say, the degree of professional experience required for assuming the office—one may try to construct a carefully selected dataset from two jurisdictions that differ in this respect, and may try to hold as many alternative explanations constant as possible.⁶⁴ This may in principle be done with the help of stratifying the sample in multiple dimensions. Stratification usually fails, however, as the number of observations is too small for which all control variables have the same expression: there are for instance too few individual judges with the exact same combination of parameters in multiple dimensions. This is when social scientists reweight the data,⁶⁵ to construct a synthetic control group.⁶⁶ Note, however, that this method can only control for what is observed. The researcher must be able to argue that further, unobserved differences are negligible. And, by design, the method cannot control for macro-level differences: all “treated” observations come from a jurisdiction with the same legal order, the same legal culture, the same political system, the same economic climate, and so forth.

B. *Experiments and Vignette Studies*

There are two more strategies worth mentioning: lab experiments and vignette studies. The core of the identification problem described above is the nature of the data. The researcher is interested in the

problem at hand and scientific standards, in ways conventionally defined by the social sciences. See further Christoph Engel, *Empirical Methods for the Law*, 174 J. INST. & THEORETICAL ECON. 5 (2018).

64. Lee Epstein, Urška Šadl & Keren Weinshall, *The Role of Comparative Law in the Analysis of Judicial Behavior*, 69 AM. J. COMP. L. XX (2021).

65. Alberto Abadie & Guido W Imbens, *Matching on the Estimated Propensity Score*, 84 ECONOMETRICA 781 (2016).

66. Alberto Abadie et al., *Comparative Politics and the Synthetic Control Method*, 59 AM. J. POL. SCI. 495 (2015).

effect of some legal solution on some (legal or extra-legal) outcome variable. She directly tests her theoretical claim with the object of inquiry. As I have explained above, all quantitative analysis is constructive. The specification of the statistical test relies on a model. No model is a complete mapping of some segment of reality. This is not a bug, but a feature. Only because the analyst zeroes in on the channel of interest can the claim be falsified. Falsification is, at its core, how quantitative analysis differs from storytelling.⁶⁷

Once one has understood the selective nature of the enterprise, it will become more acceptable to already replace the data. Rather than studying the relationship of interest with the data of interest, the researcher may focus on the former. Her theory tells her that some legal rule should have a defined effect on some outcome variable. This outcome variable will frequently be a decision by some discernible entity, for example, an individual. Then one may generate new data that directly tests the purported causal effect.

The social sciences have developed a whole panoply of methods for the purpose.⁶⁸ These methods differ by the degree of context that they require or accept. On the opposite end of the methodological spectrum is the laboratory experiment. In the interest of isolating the channel of influence, all context is deliberately stripped away. The typical experiment with a legal research question does not even talk about the ultimate goal of shedding light on the effect of the legal rule in question. Rather, the experiment is narrowed down to the one feature of the legal institution that theory predicts to do the trick. In that spirit, one may, for instance, want to understand whether the fact that judges are elected, rather than appointed, creates bias.⁶⁹

Ultimately, the researcher faces a tradeoff between internal validity (the relationship between cause and effect is beyond doubt) and external validity (the object of investigation is identical with, or at least very similar to, the object to which the theory refers).⁷⁰ There is an intermediate zone between maximum internal validity (laboratory experiment) and maximum external validity (observational data). In this set of intermediate methods, the researcher is willing to pay a price in terms of residual doubt about identification, in the interest of getting closer to the object of the inquiry. Or, the other way around, the researcher is willing to pay a price in terms of studying a phenomenon that is only analogous to the phenomenon of interest, in the interest of being more confident about causality. Field experiments

67. KARL POPPER, *THE LOGIC OF SCIENTIFIC DISCOVERY* (2005).

68. SIMEON YATES, *DOING SOCIAL SCIENCE RESEARCH* (2003).

69. Christoph Engel & Lilia Zhurakhovska, *You Are in Charge: Experimentally Testing the Motivating Power of Holding a Judicial Office*, 46 *J. LEGAL STUD.* 1 (2017).

70. Brian E. Roe & David R. Just, *Internal and External Validity in Economics Research: Tradeoffs Between Experiments, Field Experiments, Natural Experiments, and Field Data*, 91 *AM. J. AGRICULTURAL ECON.* 1266 (2009).

put more stress on internal validity. There is still random assignment to treatment. But treatment takes place under real-life conditions, which can never be standardized to the degree this can be done in the laboratory. Vignette studies put more stress on external validity. Participants are randomly exposed to context-rich scenarios, and are asked to give a hypothetical answer. The data is less credible, though, as responses are only hypothetical. But scenarios can otherwise be very realistic, and the method is also available if field experiments would be impractical or illegal. This will hold true for many research questions of interest when comparing jurisdictions. Take, for instance, family planning. No internal review board would clear a field experiment that randomly exposes families to contexts that make it less or more likely for them to want children. Asking hypothetically how they would react if a certain rule were in place will be much more acceptable.

C. *Description, Classification, and Prediction*

Finally, quantitative methods are not confined to the test of causal claims. The researcher may adopt an alternative epistemic goal. Actually, this research strategy has become very prominent lately, through the availability of large legal datasets, and high computing power. This is where comparative law may productively interact with machine learning.⁷¹

Computer scientists have developed a whole array of methods for organizing rich data. These methods are designed to detect patterns. A classic application in the area of comparative law is the claim that there are legal families. This claim can be directly translated into computer science. If the claim holds true, then it should be possible to find discernible patterns in the data from multiple jurisdictions. These patterns can be found at the level of rules, but also at the level of decisions taken on the basis of these rules. One can push the approach even a step further. Comparative law does not only claim that legal rules, and legal practice for that matter, exhibit systematic relationships among each other. Comparative law even identifies specific legal families. One may use the power of clustering algorithms to try and reconstruct these families. If the set of clusters found with the help of the computer differs from the comparative law tradition, this is the thought-provoking starting point for joint work between social scientists and comparative lawyers.⁷²

71. A very accessible introduction to the methods employed is GARETH JAMES ET AL., *AN INTRODUCTION TO STATISTICAL LEARNING* (2013).

72. Yun-chien Chang et al., *Drawing the Legal Family Tree: An Empirical Comparative Study of 170 Property Doctrines in 129 Jurisdictions*, 13 J. LEGAL ANALYSIS 231 (2021); Holger Spamann et al., *Judges in the Lab: No Precedent Effects, No Common/Civil Law Differences*, 13 J. LEGAL ANALYSIS 110 (2021).

CONCLUSION

Causal claims are tempting. Legal scholars would want to know whether a proposed reform would likely deliver on its promises. Causal claims are equally relevant in a backward-looking perspective. A classic application is the principle of proportionality. If constitutional jurisprudence adopts this principle, a legal rule will only be upheld if the case can be made that the rule serves some purpose of social betterment. The rule must be able to bring about a state of social affairs preferable to the state of affairs that would be expected should the rule be invalidated. Political scientists would call this an investigation into the effectiveness of regulative politics. Economists may wonder whether a society governed by a certain rule effectively experiences a clash between individual and social rationality. Sociologists may argue that a society embedded in a certain set of formal legal rules is more resilient to shocks than another. Psychologists may investigate whether the presence of some legal rule shapes the process by which individuals acquire information about some societal phenomenon of interest, say polarization around some leader. Jurisdictions are sovereign. This is why legal rules differ between jurisdictions. In terms of research questions, there is thus ample room for a productive collaboration and cross-fertilization between comparative lawyers and social scientists.

Yet testing causal claims is hard. The ideal setting is, by definition, impossible. One can never observe one and the same social setting, first with the respective legal rule in force, and then in its absence. The perfect counterfactual is imaginary. The social sciences have developed a series of proxies, and conventions about their acceptability. Unfortunately, comparative law is a particularly challenging type of evidence for the use of these proxies. These challenges have been worked out extensively by the empirical social sciences. For some challenges, in appropriate settings, the precise command that comparative lawyers tend to have of doctrinal detail may help construct clever workarounds. But the problem of dependence, in particular, is almost insurmountable. Legal rules in one jurisdiction do not develop independently of legal rules developed in other jurisdictions.

If the identification of a causal effect is not attainable, weaker empirical designs may still be feasible. One popular option is the construction of a synthetic control group (which, however, can only take known and understood intervening variables out of the equation). A safer option is generating new data, using some experimental method. Then identification can be achieved with the help of random assignment to treatment. Yet jurisdictions are not commonly willing to experiment with a rule borrowed from a different jurisdiction. Legislatures will not do that lightly. Vignette studies or lab experiments are easier to implement. But then what one studies is only analogous to what one wants to understand.

This Article should not be misunderstood as a plea against the use of social-science methods in comparative law, or against the use of comparative law as evidence in the social sciences. Collaboration between comparative lawyers and social scientists has high potential. There is merely no one size-fits-all design. All empirical methods have limitations. The researcher must choose carefully, and must explicitly justify why the limitations inherent in the chosen method are less concerning than the limitations arising in alternative approaches. Yet despite all these notes of caution: Not even trying, and just mustering common sense, is not likely to be the best choice.