ISSN 1936-5349 (print)
ISSN 1936-5357 (online)

HARVARD

JOHN M. OLIN CENTER FOR LAW, ECONOMICS, AND BUSINESS

EMPIRICAL COMPARATIVE LAW

Holger Spamann

Forthcoming in *Annual Review of Law and Social Science*, Vol. 11 (2015)

Discussion Paper No. 815

03/2015

Harvard Law School
Cambridge, MA 02138

# Empirical Comparative Law

**Holger Spamann**[*]

Abstract: I review the empirical comparative law literature with an emphasis on quantitative work. After situating the field and surveying its main applications to date, I turn to methodological issues. I discuss at length the obstacles to causal inference from comparative data, and caution against inappropriate use of instrumental variables and other techniques. Even if comparative data cannot identify any single causal theory, however, they are extremely important in narrowing down the set of plausible theories. I report progress in measurement design, and suggest improvements in data analysis and interpretation using techniques from other fields, particularly growth econometrics.

JEL codes C18, K00, P50

(forthcoming, 11 Annual Review of Law and Social Science (2015)

[*] Assistant professor, Harvard Law School. hspamann@law.harvard.edu. For very helpful comments, I am indebted to Jonathan Beauchamp, Martin Gelter, Tom Ginsburg, Yehonatan Givati, Dan Ho, Katerina Linos, Mila Versteeg, members of the Harvard Law School Empirical Legal Studies Group, and especially Ralf Michaels. Valerio Romano and Harin Song compiled very helpful lists of relevant articles. I gratefully acknowledge financial support from Harvard Law School's Summer Research Program.

# 1 Introduction

In this article, I review the literature that uses cross-country legal data to test causal theories in an explicit hypothesis-testing framework. This literature, which I call empirical comparative law (ECL), has grown tremendously in the last fifteen years. ECL's appeal is that the cross-country variation is large. In principle, this makes it easier to observe the variation's effects. In fact, cross-country variation is often the only variation available, as many characteristics, especially legal characteristics, are fixed within countries. On the downside, many problems afflicting empirical work in social science are particularly acute in cross-country data: samples are small (there is only one sample[1] of at most 200 countries), units (countries) are highly heterogeneous, and data is sparse (i.e., unavailable for many countries or variables). There are no experiments and few "natural experiments." As a consequence, comparative data can rarely if ever isolate any particular causal effect, although it can considerably reduce the number of plausible ones – or so I shall argue at length. I highlight the need for careful interpretation of results also under another angle, namely the insufficiency of standard significance tests for assessing the reliability of results. This issue is especially relevant for quantitative comparative studies because it is not possible to repeat a comparative test on a second, independent sample.

The vast majority of published ECL is quantitative. This does not imply that qualitative comparative evidence is not important – far from it. In particular, comparative evidence is most powerful when a single country provides a counter example to what might otherwise seem a necessary relationship. Partially because of this, however, most non-trivial yet credible hypotheses are probabilistic and hence lend themselves to quantitative tests (Spamann 2009). I will make only occasional mention of qualitative methods in this survey.

Part 2 situates ECL at the crossroads of empirical legal studies, comparative law, and sister empirical disciplines such as comparative politics. This part is not necessary for understanding the rest of article, but it may be helpful for readers wondering what, if anything, is new in ECL. Part 3 surveys the main applications of ECL to date. As ECL is a method rather than a body of knowledge, the survey's goal is not to be exhaustive but merely to illustrate the method's use. Part 4 and 5 form the substantive core of this article. Part 4 reviews the obstacles to causal inference from comparative data. I take a very skeptical view, but argue that ECL remains at least an important filter for causal theories and complements other empirical tests. Part 5 discusses practical issues of particular relevance for ECL in data collection (measurement), analysis, and interpretation. Part 6 concludes.

[1] Throughout the article, I refer to the collection of countries on earth as a sample rather than the population. In other words, I take the relevant population to be an abstractly defined set of possible countries rather than the set of presently or formerly existing countries on earth. The reason is that ECL aims or should aim to test theories that can predict what will happen in a country that was not yet in the sample, for example because it is new or because it implements a reform.

## 2 Relationship to Other Literatures

ECL is closely related to, and partially overlaps with, at least three literatures: empirical legal studies, comparative politics, and comparative law. The dividing lines are not sharp and may have as much to do with the people involved as with the questions and the methods.[2]

Substantively, ECL is a sub-field of empirical legal studies. The distinguishing feature of ECL is the use of cross-country data. This is not a fundamental distinction. For example, the comparison of national constitutions is not fundamentally different from the comparison of state constitutions (e.g., Chen and Malhotra 2007; Dixon and Holden 2014), and both can shed light on the same questions. That being said, cross-country data often offer more variation then sub-national data. Often, a particular question is decided at the national level, such that the only available variation is cross-country. The greater variability has both advantages and disadvantages. For example, an investigation of dictatorship effects would benefit from comparative data that contains dictatorships, while an investigation of details of democratic design against the backdrop of US-style mass media might be better served by a comparison of states. As a practical matter, cross-country data tend to be much more difficult to obtain, at least in a consistent format.

Comparative politics and other comparative social sciences do use cross-country data, and many of the papers surveyed below explicitly speak and may even belong to those disciplines. I have attempted to select papers that produce and/or use more and better comparative legal data than has traditionally been the case. In particular, the involvement of lawyers in ECL has considerably improved and broadened the legal data collection process.

ECL also has many points of contact with various branches of the heterodox field of comparative law.[3] The major difference is that comparative lawyers traditionally ask different, non-causal questions (Pistor 2010).[4] In particular, most comparative law has been devoted to understanding foreign legal systems (e.g., Lasser 2004), developing common concepts (e.g., Michaels 2006), and a comparative mapping of legal rules and institutions (e.g., Zweigert and Kötz 1998). Where comparative lawyers have tackled causal questions, as in the exploration of legal change (e.g., Glendon 1989), they have favored more exploratory, descriptive accounts over an explicit hypothesis-testing framework.[5] In spite or rather because of these differences, the opportunities for fruitful exchange are plentiful. Comparative law furnishes important concepts, knowledge of legal rules and institutions and their functioning in practice,

---

[2] Suchman and Mertz (2010) document a similar phenomenon in domestic empirical work. In particular, they distinguish "Empirical Legal Studies" from other empirical approaches. For brevity, I use the term "empirical legal studies" for any empirical work with legal data.

[3] On the heterodoxy of the field, see, e.g., Reimann (2002, pt. II), and compare the number of approaches surveyed in Reimann and Zimmermann (2006, pt. II).

[4] The dividing line has never been sharp. For example, Edouard Lambert (1905) speculated in his famous report for the Paris Congress of 1900 that comparative law and legal sociology are one and the same thing (35), with missions such as to reveal the "natural laws" of social life (32) or at least the effect of legal reforms in various countries (36).

[5] Some within the field even hesitate to attach the label "empirical" to comparative law, although comparative law is by definition empirical in the sense of learning from observation. Cf. Jansen (2006, 313) (contrasting comparative law's attempts to capture similarity and dissimilarity of legal systems to "the empirical sciences").

and hypotheses for ECL. ECL in turn generates data that can inform taxonomies of legal systems and other descriptive elements of comparative law.[6] In separate work, I also argue that comparative law would benefit from applying the hypothesis testing framework and empirical rigor to traditional comparative law questions such as legal families or traditions (Spamann forthcoming).

Comparative lawyers have often been quite critical of ECL, in particular the Law and Finance literature surveyed in section 3.2 below (see Michaels 2009 for a summary). I do not have space to address these explicitly, but my own criticisms from within the statistical-empirical framework nest most of them. For example, I would interpret the charges of neglect of functionally equivalent mechanisms, arbitrary selection and weighting of index components, or neglect of law in action as a problem of measurement validity (section 5.1) or low prior probability of the tested theory (section 5.3). Similarly, the criticism that empirical work has paid too little attention to local particularities is again a problem of measurement, or the problem of insufficient controls that I discuss at length in sections 4.1 and 5.2.

# 3 Examples

In this section, I review some particularly active areas of ECL. I make no attempt to be comprehensive. This would be pointless for a "field" identified by a method rather than an object of study, and it would be hopeless given the volume of the literature.[7] Rather, my goal is to illustrate the broad range of applications of ECL. For these merely illustrative purposes, I refrain from reporting statistical and economic magnitudes. I defer methodological discussions of measurement and interpretation to subsequent sections.

## 3.1 Constitutional law

There is a rich comparative empirical literature on constitutions in comparative politics and political economy (e.g., Ben-Bassat and Dahan 2008; for a survey, see, e.g., Landmann and Robinson 2009). In the last decade, however, this research has received a major boost through the Comparative Constitutions Project (CCP) directed by political scientists Zachary Elkins and James Melton and lawyer Tom Ginsburg, (comparativeconstitutionsproject.org). The CCP is a publicly available cross-national historical dataset of all written constitutions from 1789. It codes 668 constitutional characteristics, and tracks their development over time, including adoption, amendments, and suspensions.

[6] For example, Siems (forthcoming) performs a cluster analysis of the Doing Business investor protection data and finds no confirmation for standard taxonomies. Such work still remains too rare, even though the situation has improved since Reimann (2002, 686) wrote "[C]omparative law has still not acquired a solid empirical basis. We have ridiculously little statistical data about the legal systems we study and compare. Without such data, most of our conclusions rest on personal intuition, anecdotal information, or plain speculation, rather than on systematic observation of hard facts." See Siems (2014) for a textbook-length attempt to integrate "numerical" findings into comparative law.

[7] For example, Carrubba et al. (2014) count 154 articles just in comparative judicial politics from 1990-2009. Isolated examples of empirical comparative law can be found as far back as the 1960s (e.g., Schwartz and Miller 1964), and almost certainly before.

The CCP data reveal, inter alia, that ethnic divisions tend to be higher in countries with federalism and proportional electoral systems (Elkins and Sides 2007); that constitutions adopted during occupation display surprisingly little resemblance to the occupying power's constitution yet rarely survive for long (Elkins et al. 2008); that there is little relationship between judicial council design and judicial quality (Ginsburg and Garoupa 2009); that constitution-making processes (such as the involvement of a constitutional assembly, or the requirement of ratification by referendum) correlate with constitutional features such as the number of constitutionally guaranteed rights (Ginsburg, Elkins and Blount 2009); that constitutions last on average only 19 years but flexible ones last longer (Ginsburg, Melton and Elkins 2009), including flexibility provided by deferrals to legislation (Dixon and Ginsburg 2011); that constitutions have been growing in length, scope, and detail (Ginsburg 2010); and that executive term limits are overwhelmingly observed in established democracies (Ginsburg et al. 2011). It allows quantification of important phenomena such as the correlation of human rights on the books and in action (Law and Versteeg 2013), and sheds doubt on what used to be considered foundational distinctions such as between parliamentary and presidential democracies (Cheibub et al. 2014). It has inspired follow-up research, such as a book on comparative constitutional design (Ginsburg 2014). More fruitful research is sure to follow.

Other data remain in use, and are being developed. For example, Shulztiner and Carmi (2014) document the rise of "dignity"[8] with data collected directly from the constitutions of all 193 UN member states. They show both qualitatively and quantitatively that "dignity" is not necessarily associated with liberal practices, and may even be invoked for illiberal purposes. Dreher et al. (2010) use data combining human rights on the books and in action from humanrightsdata.com to show that terrorism tends to be followed by a significant decrease in the respect for human rights; using data derived from the US State Department and Amnesty International reports, Goderis and Versteeg (2012) show that such decreases are less in countries with stronger constitutional checks and balances (as measured by La Porta et al. 2004). Carrubba et al. (2014) construct a database on constitutional review conducted by high courts.

## 3.2 Law and Finance, Doing Business, and Legal Origins

Perhaps the largest literature in ECL to date is known as Law and Finance. In the eponymous paper, financial economists La Porta et al. (1998) introduced additive indicators of certain shareholder and creditor protection rules in 49 countries, respectively known as the anti-director rights index and the creditor rights index. They showed that the anti-director rights index positively correlated with equity market outcomes such as market capitalization and ownership dispersion (La Porta et al. 1997, 1998, 1999, 2000, 2002a, 2002b). Some of these early results subsequently yielded to better data (Spamann 2010c; Holderness forthcoming). But newer, more refined studies on even larger samples have upheld and refined other results and added new ones with new indices of public and private securities laws (La Porta et al. 2006), rules against managerial self-dealing (Djankov et al. 2008b), duration of and recovery in bankruptcy (Djankov et al. 2008a), and a revised creditor rights index (Djankov et al. 2007). A voluminous follow-up literature has documented correlations of these measures with various financial market outcomes, developed supporting theory, and tested corollary hypotheses on comparative and

[8] The CCP also provides this item.

domestic data. In a recent survey, La Porta et al. (2013, 450) conclude that this literature established that "better [legal] investor protection … is associated with improved financial development, better access to finance, and higher ownership dispersion." To be sure, several empirical studies dispute this claim (e.g., Armour et al. 2009b; Cheffins et al. 2013; Holderness forthcoming), including the regularity assumptions embedded in the quantitative methods used to support it (e.g., Milhaupt and Pistor 2008; Pistor 2013).

The main authors of Law and Finance soon exported their approach to other areas of law. They showed that, as measured by the legal indices they specifically designed and collected for these studies, procedural formalism "is associated with higher expected duration of judicial proceedings, less consistency, less honesty, less fairness in judicial decisions, and more corruption" (Djankov et al. 2003); "judicial independence and constitutional review are associated with greater freedom" (La Porta et al. 2004); "[c]ountries with heavier regulation of entry have higher corruption and larger unofficial economies, but not better quality of public or private goods" (Djankov et al. 2002); "[h]eavier regulation of labor is associated with lower labor force participation and higher unemployment, especially of the young" (Botero et al. 2004); and "[p]ublic disclosure [of politicians' financial and other conflicts], but not internal disclosure to parliament, is positively related to government quality, including lower corruption" (Djankov et al. 2010). A general and controversial theme of this literature is that, as measured by these studies, "interventionist" policies, such as government ownership of banks (La Porta et al. 2002a; Barth et al. 2006[9]), correlate with worse outcomes such as corruption.

The World Bank financed these and the later Law and Finance studies, and used them as foundation for its Doing Business project (doingbusiness.org) (World Bank 2014). Since 2004, Doing Business has been collecting annual legal data in ten or eleven areas of business law and regulation: starting a business, dealing with construction permits, getting electricity, registering property, paying taxes, trading across borders, getting credit, protecting minority investors, enforcing contracts, resolving insolvency, and labor market regulation. The standard template is to collect data on the cost, duration, and number of procedures involved in a paradigmatic case; some indicators instead code the presence of particular legal provisions. A dedicated team at the World Bank draws on a large network of respondent lawyers and other experts around the world to collect, verify, and improve its data. Over time, the World Bank has gradually refined its methodology and weeded out mistakes. As a result, some earlier results had to be corrected (e.g., Spamann 2010b). The method and the perceived neo-liberal bias of Doing Business have been extremely controversial (cf. Manuel et al. 2013). Used with care, however, Doing Business and related indicators[10] are extremely valuable pieces of information for empirical comparative research (cf. Davis 2014 and section 5.1 below).

[9] Barth et al. (2013) present an expanded and updated version of the underlying data on bank regulation and supervision in 180 countries from 1999 to 2011.

[10] For example, the Rule of Law Index from worldjusticeproject.org. This index has been developed by an alumnus of Doing Business (cf. http://worldjusticeproject.org/bio/staff/juan-carlos-botero, visited 1/31/2015) and resembles it in method and presentation.

Law and Finance and its progeny were successful in economics and finance and controversial in law in large part because of their solution to the endogeneity problem: does the law on the books cause the observed social phenomenon (e.g., financial market size), or is it just the other way around (see generally section 4 below)? La Porta et al. (1998) observed that their indices of investor protection were on average significantly higher in common law than in civil law systems. Arguing that membership in a legal family ("legal origin") was determined long before contemporary market outcomes and plausibly influences investor protection laws, La Porta et al. (1998) concluded that causation must run from the laws to the market outcomes. All the other studies cited above found the same correlation between common law and more "market-friendly" regulation. This raised the question why the common law countries were more market-friendly on average. Drawing on Merryman (1969) and other legal comparatists, Glaeser and Shleifer (2002) and others conjectured that the answer was to be found in the civil law's ostensible preference for statutes and less independent judges.

This conjecture drew a number of responses. First, the correlation between legal origins, laws on the books, and outcomes may be confounded by other factors. While the country-level data ruled out many (e.g., religion; La Porta et al. 2008), it could not address others. In particular, Klerman et al. (2011) pointed out that legal origin is almost perfectly correlated with colonial origin, which could have influenced subsequent developments through various other channels.[11] Bubb (2013) validated such concerns by zooming in to the local level. He showed that de facto property rights differ little between either side of the border separating Ghana and Côte d'Ivoire while other economic outcomes do; Michalopoulos and Papaioannou (2014) investigate similar phenomena across Africa. Second, even if laws at the country level were responsible for the diverging outcomes, intrinsic differences between common and civil law would not be the only plausible explanation. In particular, Spamann (2010a) showed that peripheral countries continue to copy legal materials from their origin countries, such that random variation in the origin countries England and France unrelated to the common/civil law distinction could generate the observed pattern. Third, and crucially, there was little direct evidence for the claim that differences in the role of case law (or for that matter any other phenomena traditionally linked to legal origins) explained the cross-country pattern of regulation.[12] For example, Roe (2006) pointed out that the legal rules in question were overwhelmingly statutory in all jurisdictions, not just the civil law countries, while Jackson and Roe (2009) showed that common law countries spent more money on public enforcement of securities laws. Similarly, in the one area where most modern comparative lawyers still see pronounced differences between common and civil law, civil procedure (Michaels 2009, 781), corrected data showed no performance differences between the two legal families (Spamann 2010b).

In response, La Porta et al. (2008, 286/308; 2013, 427/457) now characterize legal families generically as "style[s] of social control of economic life (and maybe of other aspects of life as well)," where "common

[11] Oto-Peralias and Romero-Avila (2014) argue that legal and colonial origin interacted, in particular because England pursued a strategy of "indirect rule" in thickly settled or otherwise hostile territories.

[12] The historical narrative in Glaeser and Shleifer (2002) has been questioned as well. See Klerman and Mahoney (2007) and Roe (2007).

law stands for the strategy of social control that seeks to support private market outcomes, whereas civil law seeks to replace such outcomes with state-desired allocations." More specifically, they conjecture that common and civil law differ in their "toolkits" or the "beliefs about how the law should deal with social problems … incorporated in legal rules, institutions, and education." This literature is currently lacking hypotheses that are both falsifiable and not obviously false, however, about which tools might be lacking in either family and how such beliefs might be hardwired in the legal system.[13]

## 3.3 Diffusion and Legal Transplants

Another active area of research has dealt explicitly with the extent to which, and the reasons why, law "diffuses" from one country to another.

One aspect of diffusion is the reception of formal legal materials from another jurisdiction, an idea popularized as "legal transplants" by Watson (1974).[14] Spamann (2010a) shows that periphery countries continued to import statutory texts and to cite treatises and other materials from core countries in the same legal family through the second half of the 20th century. In a study of ten European high courts, Gelter and Siems (2014) show, inter alia, that their citations to one another cluster within legal families as well.

Detecting outside influence is much harder when it does not involve literal copying or explicit references. In these cases, diffusion can only be inferred from the timing of reforms. In this spirit, Goderis and Versteeg (forthcoming) show that constitutional rights in 180 countries after World War II tend to track other countries with the same legal origin, particularly the former colonizer, but also countries with the same religion or aid donor. By contrast, Ginsburg and Versteeg (forthcoming) find no such pattern for the adoption of constitutional review, which correlates only with domestic political developments. There is a large literature on similar questions in political science. To give just one outstanding example, Linos (2013) examines "how health, family, and employment laws spread across countries," emphasizing the mechanism of foreign role models in democratic discourse.

A related but tougher question is whether countries influence each other substantively, with or without (visible) exchange of formal legal materials. For example, Greenhill et al. (2009) find that while formal labor laws in 90 developing countries tended to resemble those of their main export destinations in the years 1986-2002, their labor practices did not. In general, recipient systems seem to be less "functional" as measured by rates of change and flexibility of corporate law in ten countries (Pistor et al. 2003) and measures of effective legal institutions (absence of corruption etc.) in 49 countries (Berkowitz et al. 2003a, 2003b), at least where the local population had no prior experience with the foreign law.

---

[13] In particular, even proponents of legal origin differences affirm that civil law systems make extensive use of judicial precedent, i.e., case law (e.g., Merryman 1969, 48-49).

[14] There is also research on the more basic question whether (written) law in different countries is converging (e.g., Armour et al. 2009a, 2009b; Gahan et al. 2012).

## 3.4 Other examples

There are many more examples in virtually all areas of law. Most studies examine cross-sectional correlations, i.e., correlations of two or usually more features across countries at a given point in time.

For reasons discussed in section 4.2, however, some studies examine the correlation of *changes* across time and space (as in the diffusion studies mentioned in the previous subsection). For example, Armour and Cummings (2008) show that changes towards more "forgiving" bankruptcy laws tend to be followed by increases in self-employment ("entrepreneurship") in 15 Western countries in 1990-2005; Calderón and Chong (2009) show that increases in labor regulation are associated with decreases in income inequality in a large sample of countries in 1970–2000; Gonzalez and Viitanen (2009) show that the introduction of no-fault or unilateral divorce is associated with an increase in the divorce rate in European countries in the second half of the 20th century; and Klick et al. (2012) show that abortion liberalization is associated with changes in sexual behavior (as proxied by gonorrhea incidence) in 41 countries in 1980-2000. This list could easily be extended.[15]

# 4 Causation: The Identification Problem

In the previous section, I reported the results of the surveyed studies as mere correlations within a sample. The real interest for policy-makers and most authors, however, is to what extent these correlations provide evidence for a causal link between the phenomena under study. I argue here that comparative evidence by itself will hardly ever be sufficient to establish a causal claim, and that statistical methods that purport to do so are likely to do more harm than good in comparative settings. Even if comparative data cannot identify any single causal theory, however, they are extremely important in narrowing down the set of plausible theories. A thoughtful pursuit of this more modest agenda seems to me the most fruitful avenue for ECL.

## 4.1 The Idea of Causal Inference

The basic idea of inferring causation from comparative data is a *ceteris paribus* argument. If countries A and B are alike in all relevant respects except X and Y, and X precedes Y, then any difference in Y must be caused by the difference in X.[16] The large variation of interesting attributes across countries presents a distinct advantage for such an argument, in that many attributes such as judicial review differ only across countries, but not within. Unfortunately, this large variation is also comparative data's biggest disadvantage: there are never two countries that differ *only* in the attributes (X,Y) of interest. How then can comparative work identify the effect of the factor of interest (X) from among all others (X', X'', etc.), if at all?

[15] See, e.g., Moser (2005) and Lerner (2009) on patent law and innovation, and Haselmann et al. (2010) on creditor rights and lending.

[16] Of course one could (and should!) ask where the difference in X comes from. I hope that one can temporarily suppress this question for purposes of this thought experiment. In any real case, the origin of X is indeed a major issue, because whatever caused X might also have caused Y directly. Cf. the discussion in the subsequent paragraphs of the main text.

Qualitative and quantitative studies pursue fundamentally different answers to this question. Qualitative studies examine the countries in question in depth in an attempt to rule out meaningful differences on all other possibly relevant characteristics X', X'', etc., or at least that they had a confounding effect in these particular countries.[17] By contrast, quantitative work remains agnostic on the causes of outcome Y in any individual country, and focuses instead on average effects of X in groups of countries. The basic argument is familiar from randomized control trials, i.e., true experiments. For example, in a drug trial, it is accepted and in fact assumed that one cannot ascertain how much, if at all, the drug changed the health outcome for any individual patient (or would have changed it, in the case of a person receiving the placebo). Rather, one infers the average effect of the drug from the difference in average health outcomes between the treatment and control groups. This works because randomizing who receives the drug (or, generically, the treatment) makes the two groups identical in all other respects in expectation, and the probability of deviations from this expectation above a certain size can be estimated from the dispersion of individual outcomes (Holland 1986).

Social scientists generally cannot manipulate their study contexts and randomly assign a "treatment." Modern empirical social science increasingly attempts to approximate the experimental ideal, however, by exploiting natural or quasi experiments, i.e., settings where the key independent variable of interest (i.e., the candidate cause) is plausibly as-good-as-randomly assigned (Angrist and Pischke 2009, 2014; Ho and Rubin 2011; Dunning 2012; Imbens and Rubin 2015). Paradigmatic research exploits lotteries or sharp discontinuities at cutoff values of non-manipulable continuous variables. For example, researchers have estimated the effect of neighborhood quality on personal outcomes by comparing winners and losers in a housing lottery (nber.org/mtopublic), and the effect of elite schools on personal outcomes by comparing students just above and below the standardized test score required for admission (Dobbie and Fryer 2014).

Views differ on the availability of natural experiments in comparative settings. In my experience, outsiders to the comparative literature tend to be very skeptical, while insiders are unsurprisingly more upbeat.[18] All agree, however, that natural experiments are rare in comparative, and most research therefore pursues a different strategy. I will describe and assess that strategy first before returning to natural experiments.

## 4.2 Controlling

As in epidemiology and other non-experimental a/k/a observational fields, the standard way to isolate the factor of interest in ECL is to "control" for possible confounding factors. For example, the estimation of an effect of shareholder protection on equity market capitalization may allow for the possibility that

[17] Cf., e.g., King et al. (2001); Hirschl (2006); and Slater and Ziblatt (2013). The details, including the relationship to quantitative work, have been the subject of considerable controversy in political science, see, e.g., Geddes (2003); Brady and Collier (2004); Sekhon (2004); Mahoney and Goertz (2006).

[18] Tellingly, even political scientists and statisticians writing introductions to causal inference for lawyers do not mention comparative evidence. The only trace of comparative materials in Ho and Rubin (2011) is a cite to Bubb (2013). Epstein and King (2002, 103) mention cross-country research once, concluding a long list of possible sources of evidence with "*even* cross-country" (emphasis added).

the latter is also influenced by the country's majority religion or the country's size. This is also sometimes described as "holding constant" the confounding factors (e.g., religion, size). This description evokes the idea that the technique is supposed to create and then to compare observations that differ only in the factors of interest (e.g., shareholder protection), at least in expectation.

This idea is easiest to understand in an approach called matching. This approach first matches each treated observation to an untreated observation that is identical in terms of the pre-treatment (control) variables, or at least sufficiently close in some technical sense. It then estimates the treatment effect as the average difference in outcomes between the treated and their respective matched untreated observations. For example, one might find for each country with good shareholder protection a matching country of the same religion and similar size with bad shareholder protection (countries without a match are omitted). Without assuming anything about the interaction of religion and country size, one could then estimate the effect of shareholder protection on equity market size by calculating the average difference in equity market size between the matched countries. Of course, this will identify the causal effect of shareholder protection only if the matching variables (here, religion and size) include all confounding variables, which presupposes in particular that all confounders are observable in the first place. In any event, matching is rarely used in comparative studies because countries are too few and too diverse to find close matches on all but very short lists of variables.[19]

Most instead postulate a model of the interaction between the relevant variables, and estimates its coefficients. A typical model might be

$$\textit{(equity market size)} = a + b\cdot\textit{(investor protection)} + c\cdot\textit{(country size)} + \sum_{r \in religions} d_r \cdot r + \textit{(residual)},$$

where the residual is assumed to be uncorrelated with the other right-hand side variables and a, b, c, and the set of $d_r$ will be estimated with some technique such as linear regression. The benefit of this approach over matching is that one no longer requires treated and untreated observations to be identical in terms of the control variables (here country size and religion). The cost is that this only works if the variables truly interact only as postulated in the model.

In general, "controlling" is model-specific, and hence only as good as the model (e.g., Leamer 1983). In principle, the model need not be as simplistic as in the example above and can include all sorts of candidate interactions and non-linear effects. In comparative practice, however, there are far too few data points (countries) to estimate anything but simple linear models of the most obviously important observed variables. In addition, and as with matching, many relevant variables cannot be included because they are not observed or not measured, at least in most countries. As a result, the model can at best provide modest assurance that the relevant factors are even approximately "held constant."

---

[19] For example, one can never closely match the US on financial market size because the US has by far the largest financial market in the world. One would get a closer but still imperfect match on financial market size relative to GDP. The more variables one needs to match on, the fewer and poorer the matches will become. Matching on the estimated propensity score (e.g., Angrist and Pischke 2009, 3.3.2) can help but is only as good as the model for the propensity score (cf. next two paragraphs in the main text).

If there are at least two observations per country at different periods in time (panel data), a popular way to deal with unobserved cross-sectional heterogeneity is to remove it all using country fixed effects or similar methods (see examples in section 3.4).[20] This leads to a comparison of changes rather than levels, and for this reason is often referred to as differences-in-differences (DD). For example, one might estimate the relationship between shareholder protection and equity market size by examining the cross-country correlation not of these two variables at some point in time but of *changes* in these two variables between two points in time. DD identifies a causal effect if the so-called parallel trends assumption holds, i.e., if in expectation (and conditional on controls) all countries would have experienced identical changes in outcomes (e.g., equity market size) but for the change in the explanatory variable of interest (e.g., shareholder protection). Unfortunately, this is actually a rather strong assumption. Countries do not implement reforms randomly (cf. Rodrik 2012). Reforms often come in packages or react to changed circumstances unobserved by the researcher.[21] For example, many Asian countries improved shareholder protection in response to the Asian financial crisis of 1997/98 at the same time as they were recovering from a recession and implementing numerous other reforms. Unless such contemporaneous changes are carefully controlled for – and this is usually not possible in ECL for lack of data and contextual information –, DD estimates of the causal effect can easily be more biased than cross-sectional estimates.[22]

Painful experience in other disciplines has shown that these are not merely abstract concerns. In economics, Levine and Renelt (1992) revealed that almost all conclusions from cross-country growth regressions (as the literature then stood) were fragile to small changes in the selection of control variables. At the time, cross-country growth regressions followed the same template as most ECL today. Worse, LaLonde (1986) showed failure even of DD in a setting with much larger sample sizes and presumably less heterogeneity.[23] LaLonde compared experimental estimates of the earnings effect of a job-training program, which were significantly positive, with observational estimates for the same program, which were mostly negative for men and much higher than the experimental ones for

[20] With sufficiently long time series, DD can even remove biases from unobserved time-*variant* variables (Abadie et al. 2011 and forthcoming). There are other ways of using panel data that do not remove all cross-sectional variation; they are affected by a mixture of the problems discussed in this and the previous paragraph, and my sense is that they tend to obscure rather than ameliorate them.

[21] Diffusion studies present a special case of these problems (Spamann 2010a, IV.C). Diffusion studies infer foreign influence from an increase in the probability of adopting a measure after candidate leader countries have adopted the measure. The problem here is that the follower countries might just be reacting to the same common shocks to market organization, technology, security threats, and the like, without any role for interdependence specifically at the level of legal change.

[22] A separate problem is that DD also amplifies the effect of noise if and because the temporal variation in measurement error relative to its level is larger than that in the variables of interest. On a technical note, many studies in empirical comparative law do not account for the fact that repeated observations from the same country are not statistically independent. This omission may severely exaggerate the precision of DD estimates (Bertrand et al. 2004). The standard fix is clustering of the standard errors (by country), but other methods may be necessary if the number of changes is small (Cameron and Miller forthcoming).

[23] See Glazerman et al. (2003) for a systematic review of this and similar comparisons of non-experimental to experimental estimates in economics.

women.[24] Prominent cases in epidemiology include experimental refutation of observational claims that hormone replacement or certain vitamins reduce the risk of coronary heart disease and other ills (e.g., Smith and Ebrahim 2002; Lawlor et al. 2004a, 2004b; Hartz et al. 2013).[25] Under the impression of these and other failures, modern empiricists tend to be extremely skeptical of cross-country regressions. For example, Klick (2013, 908) comments on Djankov et al. (2003) that "[t]his kind of cross-sectional comparison has no chance of sorting out these issues, and conclusions based on this analysis are close to worthless in terms of having confidence in causality."

## 4.3 Natural Experiments and Instrumental Variables

Such disillusionment prompted what Angrist and Pischke (2010) call "the credibility revolution in empirical economics" (and, increasingly, political science): the search for natural experiments. Conceptually, natural experiments are controlling in reverse: rather than attempting to control for confounding factors during estimation, one uses knowledge of the "treatment" assignment mechanism (lottery, admission threshold, etc.) to argue that there are no confounding factors at work in the first place. This so-called unconfoundedness assumption is partially testable since it implies that the covariates should be balanced between treated and untreated groups.

Importantly, it is not necessary that the "treatment" X itself (e.g., good shareholder protection) be (quasi-)randomly assigned. It is sufficient if

1. First stage: some third variable Z monotonically affects X (this can be tested empirically), and
2. Exclusion restriction: Z is not in any way correlated with the outcome Y except through its effect on X; that is, Z is "exogenous" (this is an assumption).

Then the causal effect of X on Y can be estimated as the ratio of the estimated effect of Z on Y over the estimated effect of Z on X (e.g., Angrist and Pischke 2009 ch. 4). Here Z is called an instrumental variable (IV). For example, La Porta et al. (1998) initially introduced legal origin to the literature as an instrument for shareholder protection in an effort to estimate the latter's effect on equity market outcomes.

---

[24] One problem here is that of two people with identical observed earnings and other characteristics, the one to enter a training program tends to be the one whose job prospects are bleaker for some unobserved reason. A simple comparison of post-training earnings risks misattributing the effect of the initial bleak circumstances to the training program. The candidate "effect" (wages) in fact influences the candidate "cause" (enrolling in a training program). This problem is known as "endogeneity," but it can also be cast as an omitted variable problem because job prospects are not observable to the researcher. An example of an equivalent problem in comparative law is the possibility that high latent crime triggers harsh criminal law, leading to a positive correlation between crime and punishment even though punishment does reduce crime, everything else being equal.

[25] Grodstein et al. (2003) point out that many other epidemiological estimates were confirmed by experimental evidence. ECL works with far less data than epidemiology, so one should not expect ECL estimates to have the same success rate. That being said, I share the view that observational data remain useful, see subsection 4.4 below.

IV is the only type of natural experiment that has found wide application in ECL.[26] There are three mutually reinforcing reasons, however, to be very skeptical about IV estimates in ECL.[27] First, the IV estimator is notoriously unreliable – in particular, biased away from zero – in small samples such as those of ECL (Bound et al. 1995). Second, the exclusion restriction will rarely hold, if ever, in comparative applications. Country-level factors do not cleanly affect let alone correlate with only one variable of interest. For example, legal origin turned out to be correlated with multiple policy measures and outcomes, disqualifying it as an instrument for any one of them (La Porta et al. 2008). The same problem was discovered in many variables that were initially used as instruments in the cross-country growth literature (Durlauf et al. 2005; Bazzi and Clemens 2013). This has pushed researchers to search for ever more far-fetched instruments. Far-fetched instruments, however, are almost by definition weak instruments, i.e., they only have a weak first stage effect. This third problem feeds back to the first and the second, as the small sample bias and the bias from any violation of the exclusion restriction are inversely related to the strength of the instrument (Bound et al. 1995). If an a priori weak instrument appears to be strong in the data, it is probably a false positive (cf. section 5.3 below).

### 4.4 Summary: A more modest agenda

In summary, comparative data can rarely and perhaps never answer non-trivial causal questions by themselves. Attempts at causal inference using DD or IV will be grossly misleading if the treacherous conditions of these methods are not met, and can thus do more harm than good.

That being said, comparative data remain important for assessing causal claims. They may not affirmatively pin down any particular cause, but they can considerably reduce the set of plausible ones. Comparative patterns are more consistent with some theories than others (Mankiw 1995; Durlauf 2009). While some bias is always possible, some biases are less plausible than others.[28]

In particular, comparative estimates can be an important complement to estimates of causal effects from real or natural experiments. Experimental estimates cleanly identify causal effects in a particular setting (internal validity). But they cannot by themselves establish that the effect would be comparable in a different setting (external validity) (cf. Rodrik 2009; Sims 2010). Comparative data can help triangulate the generalizability of the experimental finding.

For example, cross-sectional comparative data cannot directly identify the effect of US mass incarceration on crime because the effect of punishment on crime is confounded with the reverse effect

---

[26] For example, Licht et al. (2007), Givati and Troiano (2012), and Dari-Mattiacci and Guerrero (2014) use language as instruments for culture in order to tease out the latter's effect on law. Besides IV, the other type of natural experiment is to exploit discontinuities around a cutoff, such as an international border (cf. Keele and Titiunik, forthcoming). The only example of this in ECL is Bubb (2013). The problem for ECL is that many legal rules change simultaneously at the border, such that the discontinuity in outcomes, if there is one, does not identify the effect of any one of the legal changes. The point of Bubb (2013) was to show that there was *no* discontinuous change in outcomes at the border, providing evidence against any effect of law.

[27] Consistent with this conjecture, Albouy (2012) argues that the most famous comparative result using an instrumental variable (settler mortality; Acemoglu et al. 2001) was an artifact of measurement and specification error.

[28] For a method of inferring this from the data, see Oster (2014).

of (latent) crime on punishment. By flexibly controlling for all known and measured confounders, however, one can construct joint bounds on the size of the effect and the size of any omitted confounder, opening the way for a more targeted discussion of potential explanations (Spamann 2015).

# 5 Maximizing Comparative Information

I now review certain features of empirical research that assume particular importance in ECL, whether or not causal inference is attempted. They divide into collecting, analyzing, and interpreting comparative data. All three depend on the hypothesis under investigation, and all are connected. In particular, the better the measurement and the controls, the stronger will be the conclusions that can be drawn from the data.

## 5.1 Data Collection: Measurement

Comparative work in general and comparative legal work in particular face special difficulties in designing and collecting consistent measurements.[29] Modern communications technology has considerably eased the problem of access to foreign raw information such as statutes or case law.[30] The real difficulty, however, is to distill the raw information into a measure that achieves a close fit between the facts and the concept (validity) in a reproducible, consistent manner (reliability)

Earlier failures have demonstrated that reliable measurement of alien legal institutions requires a very detailed coding protocol and usually also the involvement of lawyers in collecting and coding the data (Spamann 2010b, 2010c).[31] It is possible that certain legal institutions are straightforward enough for lay coding, as is done in the CCP. For more complex questions such as the resolution of a particular case or the legality of a transaction, however, it is hard to imagine that lay coders could correctly combine or even locate all relevant materials.

The validity of comparative measures may be compromised by inconsistency of meaning or importance of certain features across countries. For example, some statutory provision may be very important for shareholder protection in one country but irrelevant in another because of the absence or presence of certain other rules or institutions (Black et al. forthcoming). Similarly, some institution may be important in one country but redundant in another because of the presence of a functional equivalent.[32] Whether or not the measure should take into account such functional equivalents etc. is determined by the measured concept, and thus ultimately by the hypothesis under investigation. For example, the hypothesis may be specifically about the effect of *statutory* shareholder protection, perhaps because this is the only concept under policymakers' direct control. In this case, taking into account case law or legal practice in measurement would diminish rather than increase the measure's validity, however important case law and practice would be for broader concepts of shareholder protection. Conceptual

[29] For an example outside of law, see, e.g., the OECD's harmonized unemployment rates.
[30] Such technology now includes crowdsourcing sites like nomography.wustl.edu and participedia.net.
[31] To enable others to verify and replicate the measurement, it is also advisable to post the raw data and coding protocol online, as I did in Spamann (2008).
[32] On the notion of functional equivalence in comparative law, see Michaels (2006).

clarity is thus a precondition for valid measurement. Broad concepts like “shareholder protection” may require further refinement before one can even begin to discuss validity (Bebchuk and Hamdani 2009; see generally Adcock and Collier 2001).

Much progress has been made in legal measurement design. Measuring law is not limited to counting the presence or absence of certain statutory rights (as in, e.g., La Porta et al. 1998, Djankov et al. 2008b). One possible improvement is to determine the weight of individual components through factor analysis (Rosenthal and Voeten 2007). A more fundamental improvement is to account for the interaction of different rules by coding not the rules themselves but the treatment of a paradigmatic case (cf. Djankov et al. 2008a; World Bank 2014) or better several cases.[33] Non-legal phenomena can serve as an indirect measure if they are plausibly directly and strongly related to the legal aspect of interest. For example, the average discount at which minority shares trade relative to control blocks (Dyck and Zingales 2004) is arguably a direct function of minority shareholder protection, albeit not only legal shareholder protection. When more than one measure is available, they can be synthesized into one superior measure (Pemstein et al. 2010).

In principle, the quality of measurement can be explicitly validated empirically. In particular, one can verify that the measurement correlates with other measurements of the same concept, or with other variables that the concept is known to be correlated with (Adcock and Collier 2001).[34] Unfortunately, such opportunities are rarely available in ECL. The consequence is that studies tend to test *joint* hypotheses: the substantive hypothesis, *and* the hypothesis that the measurement of the relevant concept is valid (or as many such hypotheses as there are concepts involved). I will return to this problem in subsection 5.3.

## 5.2 Data Analysis: Controlling Revisited

In data analysis, it is important to control as comprehensively, flexibly, and transparently as possible precisely because candidate causes (“treatments”) are not randomly assigned in ECL. Without random assignment, other variables may be systematically correlated with the treatment and bias the estimate of the effect of interest. While it is arguably pointless to try to identify any particular causal effect (see section 4.2 above), sensible controlling will help a great deal in limiting the set of plausible biases and, ultimately, plausible causal relationships. In this respect, ECL has much to learn from modern growth empirics (cf. Durlauf 2001, 2009; Durlauf et al. 2005).

Current practice in ECL is to select a few controls ad hoc. The reason is that cross-country samples are small. In small samples, precise estimation of many parameters is impossible with classical methods. Ad hoc selection merely gives a semblance of precision, however, by neglecting model uncertainty. Three improvements are available. First, missing data for individual observations can be imputed to increase sample size. Imputation not only avoids wasting information; it also reduces selection bias (Little and

[33] This would resemble the common core approach in classical comparative law (common-core.org) (Michaels 2009; Spamann 2009).

[34] For an example involving a legal measure, see the investigation of measures of judicial independence by Ríos-Figueroa and Staton (2012).

Rubin 2002; Honaker and King 2010). Second, model-averaging techniques can explicitly account for model uncertainty (e.g., Magnus et al. 2010). Third, in some applications, principled selection among controls is possible even when the number of possible controls is larger than the sample size (Belloni et al. 2014). The latter assumes that the number of truly relevant factors is ultimately small. This so-called sparsity assumption is strong. If one considers it false, however, then one should arguably abandon comparative research because there is no hope of identifying complex connections with few data points.[35]

At the same time, there is also a danger of controlling too much. To be more precise, the use of certain controls implies assumptions that may not be plausible and may change the interpretation of the results. In particular, use of a variable as a control in a regression interpreted as an approximation of a causal relationship implicitly assumes that the variable is exogenous, i.e., not itself affected by the outcome variable. If it were so affected (i.e., endogenous), then the estimates for all independent variables would be biased in generally unknown ways. While this is well known in the abstract, the consequences are not always fully appreciated. For example, most studies, including all of the Law and Finance and Legal Origins literature, control for GDP per capita. The resulting estimates are unbiased only if GDP is exogenous, i.e., if the outcome variables, such as financial market size or the quality of judicial procedures, have no effect on GDP. This is possible, but it would make the estimates much less policy-relevant.

## 5.3 Interpretation

Last not least, it is important to interpret results sensibly in light of prior information, including the study design itself. I already discussed at length the obstacles to causal inference (section 4) and the steps that should be taken to at least reduce the number of alternative causal interpretations (section 5.2). That discussion was mostly concerned with bias, i.e., the possibility that the estimate would *systematically* be higher or lower than the true effect because of confounding with other effects such as selection. I only hinted at the issue of spurious findings, i.e., the possibility that the estimate on a particular sample is *fortuitously* higher or lower than the true effect because of sampling error. For example, the treatment group in a drug trial might fortuitously contain a disproportionate number of subjects with hidden health problems, which would make the drug's efficacy appear less than it truly is. Unlike bias, sampling error will differ from sample to sample. In principle, replication on a new, independent sample can therefore address suspicions that the finding is spurious. In ECL, this is cold comfort. Usually, there is only one sample, which is the set of existing countries on earth, or perhaps some relevant subset thereof.[36] This is why spurious findings deserve particular attention in this survey.

The standard way of dealing with sampling error is to derive an estimate of the sampling variation – the standard error – from the data in order to calculate the probability of (erroneously) estimating an effect

[35] This is the "bet on sparsity" principle coined by Hastie et al. (2009, 611): "Use a procedure that does well in sparse problems, since no procedure does well in dense problems." But see Gelman (2011), who argues that sparsity is inapposite in social science.
[36] Cf. note 1 above.

of equal or greater size under the null hypothesis of no effect – the p-value.[37] A p-value of 10% or perhaps 5% is commonly considered statistically significant and tends to be required for publication. As is well known in theory and increasingly appreciated in practice in other disciplines[38], however, low reported p-values are insufficient to address spuriousness.[39] There are two reasons for this.

First, because of multi-testing, the true probability of falsely rejecting the null hypothesis tends to be much higher than the reported p-value. It is common for individual researchers to try many variables and specifications and report only the “successful” ones. In any event, researchers collectively try many more variables and specifications, and only the “successful” researchers publish their findings. The problem here is not the multi-testing per se, as extensive testing and even filtering of promising results is surely desirable. Rather, the problem is that the reported p-values are grossly understated.[40] Reported p-values assume that only a single study was performed. But the greater the number of (unreported) studies, the greater the probability of finding a spurious result above a certain size.

Second, by definition, p-values are not equal to the probability that the null hypothesis is correct, nor is one minus the p-value equal to the probability that the alternative hypothesis is correct. Rather, the odds for the alternative hypothesis after seeing the data – the posterior odds – are equal to the odds prior to seeing the data multiplied by the Bayes factor, which is the ratio of the prior probabilities of observing the data under the alternative and the null hypotheses, respectively (e.g., Kass and Raftery 1995). Of this formula’s two factors, only the Bayes factor is loosely related to the p-value (e.g., Strnad 2007, 2.2). The other factor – the prior odds – means that prior plausibility matters even after seeing the data. A wildly implausible theory may become less implausible after seeing the data, but unless the result is extremely strong, the theory will remain implausible. Importantly, multi-testing presumably implies that any of the tested models/theories has a low prior probability of being true, or else fewer models/theories would have been tested (Cox 2006, 88).

The formula for the posterior odds also emphasizes that a test can be informative only to the extent the predictions of the null and the alternative differ. At first sight, this may not seem very important because a particular point estimate is naturally much more likely to arise if the true effect is equal to or close to the point estimate (the usual interpretation) than if the true effect is zero. But an effect of that size may not be plausible, and an effect of plausible size may not yield very different predictions from the null (cf. Gelman and Karlin 2011). In particular, when measurement is known to be very noisy, the expected estimate under the alternative hypothesis will be strongly biased towards zero. Besides, the alternative hypothesis is rarely if ever specified as a precise number, let alone the one actually later estimated. This issue would require a longer detour into statistics and is beyond the scope of this

---

[37] Estimating standard errors can be tricky. One issue of particular importance to comparative studies is that no country is literally independent from all others, as would be required for standard methods of calculating standard errors. This issue has not received attention in ECL, presumably on the assumption that it is minor (which may be in conflict with the diffusion research).

[38] See, e.g., Pashler and Wagenmakers (2012) on the “replicability crisis” in psychology.

[39] An additional problem is that an exclusive focus on statistical significance does not take into account the respective consequences of erring on one side or another. See, e.g., Ziliak and McCloskey (2007).

[40] On the importance of priors, see the next paragraph in the main text.

paper.[41] The important takeaway, however, is that limitations of the data and data analysis remain important for interpretation of the results even if the latter are “statistically significant.” An approximate litmus test is to what extent one would consider an estimated coefficient of zero as evidence against the alternative hypothesis. The less this is the case, the less the predictions of the alternative hypothesis given measurement error etc. differ from the null, and hence the less one can learn from the evidence.

As a practical matter, the foregoing means that one cannot credibly test effects that must reasonably be small relative to the noise. In particular, one cannot sensibly test with comparative data the effect of technical rules on big picture outcomes such as GDP growth that are the product of a large number of factors. Instead, one should focus on the technical rules’ effects on less distant outcomes. For example, to test the effect of culture on property protection, Dari-Mattiacci and Guerriero (2014) specifically collect data on one directly pertinent and easily measurable variable, namely the number of years after which an illegally dispossessed owner of a moveable good loses her property rights to a bona fide purchaser, if ever.

Since comparative evidence is limited, it is imperative to test the theory’s assumptions or implications also in other, domestic settings. For example, Linos (2013) uses survey evidence from the US to bolster her claim that foreign and international models legitimate policy options and hence diffuse; this evidence is particularly powerful because the size, geopolitical dominance, and geographic isolation of the US made US voters least likely to be so influenced (this is known as most difficult case logic; Hirschl 2006). Similarly, Cassar et al. (2014) bolster claims that well-functioning legal institutions increase trust with experimental evidence. On the other hand, the claim that legal origin mattered became much less plausible when domestic evidence of archetypical differences between common and civil law, such as reliance on case law, and their relationship to the tested outcomes was not forthcoming (for example, the driving force of US investor protection turns out to be statutes, not case law).

# 6 Conclusion

Comparative information is important to assess causal claims. At the same time, this article cautioned against drawing overly strong conclusions from comparative data alone. From an individual researcher’s perspective, it is tempting to brush aside these concerns and “just do it.” After all, there is no risk of being shown wrong by a controlled experiment as in other disciplines. In fact, there is not even a risk that another researcher will obtain different results on a different sample – there is only one planet earth. But from the profession's perspective, the inability to weed out errors through replication is all the more reason to look critically at empirical findings. Otherwise erroneous findings will pile up and blur our vision, and the incentive to publish such findings will divert attention from higher value targets.

[41] See, e.g., Strnad (2007, 2.2). Bayesian statistics formally integrates data and prior beliefs, including about aspects of the study design (e.g., Gelman et al. 2013). It allows precise treatment of, e.g., doubts about the strength and exogeneity of instruments discussed in section 4.3 (Conley et al. 2012).

We will do better if we are clear about the strengths and weaknesses of comparative data. Comparative data will rarely if ever sort out causal questions by themselves. That being said, they can be an extremely important piece in a broader empirical and theoretical analysis. Theories gain strength if they fit the comparative facts, and lose if they do not.[42] Establishing such facts through high quality data collection should be the first priority. [43]

[42] For an example of such use of comparative data, see., e.g., Givati (2014).
[43] For example, data on the number, size, and budget of courts have become available only recently and only for the member states of the Council of Europe, and even that only with significant qualifications regarding the comparability of the data (European Commission for the Efficiency of Justice 2014).

## References


1. Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. Journal of the American Statistical Association 105:493-505
2. Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. Forthcoming. Comparative Politics and the Synthetic Control Method. American Journal of Political Science.
3. Acemoglu, Daron, Simon Johnson, and James Robinson. 2001. The Colonial Origins of Comparative Development: An Empirical Investigation. American Economic Review 91:1369-1401.
4. Albouy, David. 2012. The Colonial Origins of Comparative Development: An Empirical Investigation: Comment. American Economics Review 102:3059-3076.
5. Angrist, Joshua, and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics. Princeton and Oxford: Princeton University Press.
6. Angrist, Joshua, 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.
7. Angrist, Joshua, and Jörn-Steffen Pischke. 2014. Mastering 'Metrics. Princeton and Oxford: Princeton University Press.
8. Armour, John, and Douglas Cumming. 2008. Bankruptcy Law and Entrepreneurship. American Law and Economics Review 10:303-350.
9. Armour, John, Simon Deakin, Prabirjit Sarkar, Siems, Mathias, and Ajit Singh. 2009b. Shareholder Protection and Stock Market Development: An Empirical Test of the Legal Origins Hypothesis. Journal of Empirical Legal Studies 6:343-380.
10. Armour, John, Simon Deakin, Priya Lele, and Mathias Siems. 2009a. How Do Legal Rules Evolve? Evidence from a Cross-Country Comparison of Shareholder, Creditor, and Worker Protection. American Journal of Comparative Law 57:579-629.
11. Barth, James, Gerard Caprio, and Ross Levine. 2006. Rethinking Bank Regulation: Till Angels Govern. Cambridge University Press.
12. Barth, James, Gerard Caprio, and Ross Levine. 2013. Bank Regulation and Supervision in 180 Countries from 1999 to 2011. NBER Working Paper No. 18733.
13. Bazzi, Samuel, and Michael Clemens. 2013. Blunt Instruments: Avoiding Common Pitfalls in Identifying the Causes of Economic Growth. American Economic Journal: Macroeconomics 5(2):152–186.
14. Bebchuk, Lucian, and Assaf Hamdani. 2009. The Elusive Quest for Global Corporate Governance Standards. University of Pennsylvania Law Review 157:1263-1317.
15. Becht, Marco, Colin Mayer, and Hannes Wagner. 2008. Where do firms incorporate? Deregulation and the cost of entry. Journal of Corporate Finance 14:241-256.
16. Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. High-Dimensional Methods and Inference on Structural and Treatment Effects. Journal of Economic Perspectives 28(2):1-23.
17. Ben-Bassat, Avi, and Momi Dahan. 2008. Social rights in the constitution and in practice. Journal of Comparative Economics 36:103–119.
18. Benjamini, Yoav, and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society (Series B) 57:289-300.
19. Berkowitz, Daniel, Katharina Pistor, and Jean-Francois Richard. 2003a. Economic development, legality, and the transplant effect. European Economic Review 47:165-195.
20. Berkowitz, Daniel, Katharina Pistor, and Jean-Francois Richard. 2003b. The Transplant Effect. American Journal of Comparative Law 51:163-203.
21. Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. How much should we trust Differences-in-Differences Estimates? Quarterly Journal of Economics 119:249-275.
22. Black, Bernard, Gledson de Carvalho, Vikramaditya Khanna, Woochan Kim, and Burcin Yurtoglu. Forthcoming. Methods for Multicountry Studies of Corporate Governance: Evidence from the BRIKT Countries. Journal of Econometrics.

23. Botero, Juan, Simeon Djankov, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. The Regulation of Labor. Quarterly Journal of Economics 119:1339.
24. Bound, John, David Jaeger, and Regina Baker. 1995. Problems with Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak. Journal of the American Statistical Association 90:443-450.
25. Brady, Henry, and David Collier eds. 2004. Rethinking Social Inquiry: Diverse Tools, Shared Standards.
26. Bubb, Ryan. 2013. The Evolution of Property Rights: State Law or Informal Norms? Journal of Law and Economics 56:555-594.
27. Calderón, César, and Alberto Chong. 2009. Labor market institutions and income inequality: an empirical exploration. Public Choice 138:65-81.
28. Cameron, Colin, and Douglas Miller. Forthcoming. A Practitioner's Guide to Cluster-Robust Inference. Journal of Human Resources.
29. Carrubba, Clifford, Matthew Gabel, Gretchen Helmke, Andrew Martin, and Jeffery Staton. 2014. The Comparative Law Project. Center for Empirical Research in the Law, Washington University in St. Louis.
30. Cassar, Alessandra, Giovanna d'Adda, and Pauline Grosjean. 2014. Institutional Quality, Culture, and Norms of Cooperation: Evidence from Behavioral Field Experiments. Journal of Law and Economics 57:821-863.
31. Cheffins, Brian, Steven Bank, and Harwell Wells. 2013. Questioning 'law and finance': US stock market development, 1930–70. Business History 55:601-619
32. Cheibub, José Antonio, Zachary Elkins, and Tom Ginsburg. 2013. Beyond Presidentialism and Parliamentarism. British Journal of Political Science.
33. Chen, Jowei, and Neil Malhotra. 2007. The Law of k/n: The Effect of Chamber Size on Government Spending in Bicameral Legislatures. American Political Science Review 101:657-676.
34. Conley, Timothy, Christian Hansen, and Peter Rossi. 2012. Plausibly Exogenous. Review of Economics and Statistics 94:260-272.
35. Cox, D.R. 2006. Principles of Statistical Inference. Cambridge University Press.
36. Dari-Mattiacci, Giuseppe, and Carmine Guerriero. 2014. Law and Culture: A Theory of Comparative Variation in Bona Fide Purchase Rules. Amsterdam Law School Legal Studies Research Paper No. 2014-57.
37. Davis, Kevin. 2014. Legal Indicators: The Power of Quantitative Measures of Law. Annual Review of Law and Social Science 10:37-52.
38. Dixon, Rosalind, and Tom Ginsburg. 2011. Deciding not to decide: Deferral in constitutional design. International Journal of Constitutional Law 9:636-672.
39. Dixon, Rosalind, and Richard Holden. 2014. Constitutional Amendment Rules: The Denominator Problem. In Ginsburg (2014), 195.
40. Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer. 2007. Private credit in 129 countries. Journal of Financial Economics 84:299.
41. Djankov, Simeon, Oliver Hart, Caralee McLiesh, and Andrei Shleifer. 2008a. Debt Enforcement around the World. Journal of Political Economy 116:1105.
42. Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2010. Disclosure by Politicians. American Economic Journal: Applied Economics 2:179-209.
43. Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008b. The law and economics of self-dealing. Journal of Financial Economics 88:430.
44. Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. Courts. Quarterly Journal of Economics 118:453.
45. Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2002. The Regulation of Entry. Quarterly Journal of Economics 117:1.
46. Dobbie, Will, and Roland Fryer. 2014. The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools. American Economic Journal: Applied Economics 6(3):58-75.
47. Dreher, Axel, Martin Gassebner, and Lars Siemers. 2010. Does Terrorism Threaten Human Rights? Evidence from Panel Data. Journal of Law and Economics 53:65-93.
48. Durlauf, Steven 2009. The Rise and Fall of Cross-Country Growth Regressions. History of Political Economy 41:315-333.

49. Durlauf, Steven, Paul Johnson, and Jonathan Temple. 2005. Growth Econometrics. In Handbook of Economic Growth 1A:555-677.
50. Durlauf, Steven. 2001. Manifesto for a Growth Econometrics. Journal of Econometrics 100:65-69.
51. Dunning, Thad. 2012. Natural Experiments in the Social Sciences: A Design-Based Approach. Cambridge University Press.
52. Dyck, Alexander, and Luigi Zingales. 2004. Private Benefits of Control: An International Comparison. Journal of Finance 59:537-600.
53. Elkins, Zachary, and John Sides. 2007. Can Institutions Build Unity in Multiethnic States? American Political Science Review 101:693-708.
54. Elkins, Zachary, Tom Ginsburg, and James Melton. 2008. Baghdad, Tokyo, Kabul …: Constitution Making in Occupied States. William and Mary Law Review 49:1139-1178.
55. European Commission for the Efficiency of Justice (CEPEJ). 2014. European judicial systems – Edition 2014 (2012 data): efficiency and quality of justice.
56. Gahan, Peter, Richard Mitchell, Sean Cooney, Andrew Stewart, and Brian Cooper. 2012. Economic Globalization and Convergence in Labor Market Regulation: An Empirical Assessment. American Journal of Comparative Law 60:703-741.
57. Garoupa, Nuno, and Tom Ginsburg. 2009. Guarding the Guardians: Judicial Councils and Judicial Independence. American Journal of Comparative Law 57:103-134.
58. Gelman, Andrew, and John Carlin. 2011. Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. Perspectives on Psychological Science 9:641-651.
59. Geddes, Barbara. 2003. Paradigms and Sand Castles: Theory Building and Research Design in Comparative Politics. University of Michigan Press.
60. Gelman, Andrew. 2011. Causality and Statistical Learning. American Journal of Sociology 117:955-966.
61. Gelman, Andrew, John Carlin, Hal Stern, David Dunson, Aki Vehtari, and Donald Rubin. 2013. Bayesian Data Analysis (3rd ed.).
62. Gelter, Martin, and Mathias Siems. 2014. Citations to Foreign Courts – Illegitimate and Superfluous, or Unavoidable? Evidence from Europe. American Journal of Comparative Law 62:35-85.
63. Glendon, Mary Ann. 1989. The Transformation of Family Law. Chicago and London: University of Chicago Press.
64. Ginsburg, Tom (ed.). 2014. Comparative Constitutional Design. Cambridge: Cambridge University Press.
65. Ginsburg, Tom, James Melton, and Zachary Elkins. 2009. The Endurance of National Constitutions. Cambridge: Cambridge University Press.
66. Ginsburg, Tom, James Melton, and Zachary Elkins. 2011. On the Evasion of Executive Term Limits. William and Mary Law Review 52:1807-1873.
67. Ginsburg, Tom, Zachary Elkins, and Justin Blount. 2009. Does the Process of Constitution-Making Matter? Annual Review of Law and Social Science 5:209-223.
68. Ginsburg, Tom. 2010. Constitutional Specificity, Unwritten Understandings and Constitutional Agreement. In Constitutional Topography: Values and Constitutions, 69-93 (A. Sajo and R. Uitz eds.).
69. Ginsburg, Tom, and Mila Versteeg. Forthcoming. Why Do Countries Adopt Constitutional Review? Journal of Law, Economics, and Organization.
70. Giuseppe Dari-Mattiacci and Carmine Guerriero. 2014. Law and Culture: A Theory of Comparative Variation in Bona Fide Purchase Rules. Working paper, University of Amsterdam.
71. Givati, Yehonatan. 2014. Legal Institutions and Social Values: Theory and Evidence from Plea Bargaining Regimes. Journal of Empirical Legal Studies 11:867-893.
72. Givati, Yehonatan, and Ugo Troiano. 2012. Law, Economics, and Culture: Theory of Mandated Benefits and Evidence from Maternity Leave Policies. Journal of Law and Economics 55:339-364.
73. Glaeser, Edward, and Andrei Shleifer. 2002. Legal Origins. Quarterly Journal of Economics 117:1193-1229.
74. Glazerman, Steven, Dan Levy, and David Myers. 2003. Nonexperimental versus Experimental Estimates of Earnings Impacts. Annals of the American Academy of Political and Social Science 589:63-93.
75. Goderis, Benedikt, and Mila Versteeg. 2012. Human Rights Violations after 9/11 and the Role of Constitutional Constraints. Journal of Legal Studies 41:131-164.
76. Goderis, Benedikt, and Mila Versteeg. Forthcoming. The Diffusion of Constitutional Rights. International Review of Law and Economics.

77. Gonzalez, Libertad, and Tarja Viitanen. 2009. The effect of divorce laws on divorce rates in Europe. European Economic Review 53:127-138.
78. Greenhill, Brian, Layna Mosley, and Aseem Prakash. 2009. Trade-based Diffusion of Labor Rights: A Panel Study, 1986–2002. American Political Science Review 103:669-690.
79. Grodstein, Francine, Thomas Clarkson, and JoAnn Manson. 2003. Understanding the Divergent Data on Postmenopausal Hormone Therapy. New England Journal of Medicine 348:645-650.
80. Hartz, Arthur, Tao He, Robert Wallace, and John Powers. 2013. Comparing hormone therapy effects in two RCTs and two large observational studies that used similar methods for comprehensive data collection and outcome assessment. BMJ Open 2013;3:e002556. doi:10.1136/bmjopen-2013-002556.
81. Haselmann, Rainer, Katharina Pistor, and Vikrant Vig. 2010. How Law Affects Lending. Review of Financial Studies 23:549-580.
82. Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. The Elements of Statistical Learning (2nd ed., 10th printing 2013). Available at http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf.
83. Hirschl, Ran. 2006. The Question of Case Selection in Comparative Constitutional Law. American Journal of Comparative Law 53:125-155.
84. Ho, Daniel, and Donald Rubin. 2011. Credible Causal Inference for Empirical Legal Studies. Annual Review of Law and Social Science 7:17-40.
85. Holderness, Clifford. Forthcoming. Law and Ownership Reexamined. Critical Finance Review.
86. Holland, Paul. 1986. Statistics and Causal Inference. Journal of the American Statistical Association 81:945-960.
87. Honaker, James, and Gary King. 2010. What to Do about Missing Values in Time-Series Cross-Section Data. American Journal of Political Science 54:561-581.
88. Imbens, Guido, and Donald Rubin. 2015. Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press.
89. Jackson, Howell, and Mark Roe. 2009. Public and private enforcement of securities laws: Resource-based evidence. Journal of Financial Economics 93:207-238.
90. Jansen, Nils. 2006. Comparative Law and Comparative Knowledge. In Reiman and Zimmermann (2006), 305-337.
91. Kass, Robert, and Adrian Raftery. 1995. Bayes Factors. Journal of the American Statistical Association 90:773-795.
92. Keele, Luke, and Rocio Titiunik. Forthcoming. Geographic Boundaries as Regression Discontinuities. Political Analysis.
93. King, Gary, Robert Keohane, and Sidney Verba. 2001. Designing Social Inquiry: Scientific Inference in Qualitative Research. Princeton University Press.
94. Klerman, Daniel, and Paul Mahoney. 2007. Legal Origin? Journal of Comparative Economics 35:278-293.
95. Klerman, Daniel, Paul Mahoney, Holger Spamann, and Mark Weinstein. 2011. Legal Origin or Colonial History? Journal of Legal Analysis 3:379-409.
96. Klick, Jonathan. 2013. Shleifer’s Failure. Texas Law Review 91:899-909.
97. Klick, Jonathan, Sven Neelsen, and Thomas Stratmann. 2012. The Relationship between Abortion Liberalization and Sexual Behavior: International Evidence. American Law and Economics Review 14:457-487.
98. La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2002a. Government Ownership of Banks. Journal of Finance 57:265.
99. La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2013. Law and Finance after a Decade of Research. In Handbook of the Economics of Finance, vol. 2A (George Constantinides, Milton Harris, and Rene M. Stulz eds), 425-491. Elsevier.
100. La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2008. The Economic Consequences of Legal Origin. Journal of Economic Literature 46:285.
101. La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2006. What Works in Securities Laws? Journal of Finance 61:1.
102. La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 1999. Corporate Ownership Around the World. Journal of Finance 54:471.

103. La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1998. Law and Finance. Journal of Political Economy 106:113.
104. La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 1997. Legal Determinants of External Finance. Journal of Finance 52:1131.
105. La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 2002b. Investor Protection and Corporate Valuation. Journal of Finance 57:1147.
106. La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny. 2000. Agency Problems and Dividend Policies around the World. Journal of Finance 55:1.
107. La Porta, Rafael, Florencio Lopez-de-Silanes, Cristian Pop-Eleches, and Andrei Shleifer. 2004. Judicial Checks and Balances. Journal of Political Economy 112:445.
108. LaLonde, Robert. 1986. Evaluating the Econometric Evaluations of Training Programs with Experimental Data. American Economic Review 76:604-620.
109. Lambert, Edouard. 1905. Conception générale, définition, méthode et histoire du droit comparé. Le droit comparé et l'enseignement du droit. In Congrès international de droit comparé, tenu à Paris du 31 juillet au 4 août 1900. Procès verbaux des séances et documents, vol. I, 26-61. Paris: LGDJ.
110. Landmann, Todd, and Neil Robinson. 2009. The SAGE Handbook of Comparative Politics. Russell Sage.
111. Lasser, Mitchel. 2004. Judicial Deliberations: A Comparative Analysis of Judicial Transparency and Legitimacy. Oxford University Press.
112. Law, David, and Mila Versteeg. 2013. Sham Constitutions. California Law Review 101:863-952.
113. Lawlor, Debbie, George Davey Smith, Richard Bruckdorfer, Devi Kundu, and Shah Ebrahim. 2004a. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? Lancet 363(9422):1724-1727.
114. Lawlor, Debbie, George Davey Smith, and Shah Ebrahim. 2004b. The hormone replacement–coronary heart disease conundrum: is this the death of observational epidemiology? International Journal of Epidemiology 33:464–467.
115. Leamer, Edward. 1983. Let's Take the Con Out of Econometrics. American Economic Review 73:31-43.
116. Lerner, Josh. 2009. The Empirical Impact of Intellectual Property Rights on Innovation: Puzzles and Clues. American Economic Review: Papers and Proceedings 2009, 99(2):343–348.
117. Levine, Ross, and David Renelt. 1992. A Sensitivity Analysis of Cross-Country Growth Regressions. American Economic Review 82:942-963.
118. Licht, Amir, Chanan Goldschmidt, and Shalom Schwartz. 2007. Culture rules: The foundations of the rule of law and other norms of governance. Journal of Comparative Economics 35:659-688.
119. Linos, Katerina. 2013. The Democratic Foundations of Policy Diffusion. Oxford and New York: Oxford University Press.
120. Little, Roderick, and Donald Rubin. 2002. Statistical Analysis with Missing Data (2nd ed.). Hoboken: Wiley.
121. Magnus, Jan, Owen Powell, and Patricia Prüfer. 2010. A comparison of two model averaging techniques with an application to growth empirics. Journal of Econometrics 154:139-153.
122. Mahoney, James, and Gary Goertz. 2006. A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research. Political Analysis 14:227-249.
123. Mankiw, N. Gregory. 1995. The Growth of Nations. Brookings Papers on Economic Activity 1:275-326.
124. Manuel, Trevor, Carlos Arruda, Jihad Azour, Chong-en Bai, Timothy Besley, Dong-Sung Cho, Sergei Guriev, Huguette Labelle, Jean Pierre Landau, Arun Maira, and Hendrik Wolf. 2013. Independent Panel Review of the Doing Business report. Washington, D.C. http://www.dbrpanel.org/sites/dbrpanel/files/doing-business-review-panel-report.pdf
125. Merryman, John. 1969. The Civil Law Tradition. Stanford University Press.
126. Michaels, Ralf. 2006. The Functional Method of Comparative Law. In Reiman and Zimmermann (2006), 339-382.
127. Michaels, Ralf. 2009. Comparative Law by Numbers? Legal Origins Thesis, Doing Business Reports, and the Silence of Traditional Comparative Law. American Journal of Comparative Law 57:765-795.
128. Michalopoulos, Stelios, and Elias Papaioannou. 2014. National Institutions and Subnational Development in Africa. Quarterly Journal of Economics 129:151-213.
129. Milhaupt, Curtis, and Katharina Pistor. 2008. Law and Capitalism: What Corporate Crises Reveal About Legal Systems and Economic Development Around the World. Chicago: University of Chicago Press.

130. Moser, Petra. 2005. How Do Patent Laws Influence Innovation? Evidence from Nineteenth-Century World's Fairs. American Economic Review 95:1214-1236.
131. Mulligan, Casey, and Andrei Shleifer. 2005. Conscription as Regulation. American Law and Economics Review 7:85.
132. Ochsen, Carsten. 2010. Crime and labor market policy in Europe. International Review of Law and Economics 30:52-61.
133. Oster, Emily. 2014. Unobservable Selection and Coefficient Stability: Theory and Validation. Working paper, University of Chicago Booth School of Business, available at http://faculty.chicagobooth.edu/emily.oster/papers/selection.pdf.
134. Oto-Peralías, Daniel, and Diego Romero-Avila. 2014. The Distribution of Legal Traditions Around the World: A Contribution to the Legal-Origins Theory. Journal of Law and Economics 57:561-628.
135. Pashler, Harold, and Eric–Jan Wagenmakers. 2012. Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? Perspectives on Psychological Science 7:528-530.
136. Pistor, Katharina, Yoram Keinan, Jan Kleinheisterkamp, and Mark West. 2003. Innovation in corporate law. Journal of Comparative Economics 31:676-694.
137. Pistor, Katharina. 2010. Statistische Rechtsvergleichung: Eine Kritische Bestandsaufnahme. Zeitschrift für Vergleichende Rechtswissenschaft 109:348-361.
138. Pistor, Katharina. 2013. A legal theory of finance. Journal of Comparative Economics 41:315-330.
139. Reiman, Mathias. 2002. The Progress and Failure of Comparative Law in the Second Half of the Twentieth Century. American Journal of Comparative Law 50:671-700.
140. Reimann, Mathias, and Reinhard Zimmermann. 2006. The Oxford Handbook of Comparative Law. Oxford University Press.
141. Ríos-Figueroa, Julio, and Jeffrey Staton. 2012. An Evaluation of Cross-National Measures of Judicial Independence. Journal of Law, Economics, and Organization 30:104-137.
142. Rodrik, Dani. 2009. The New Development Economics: We shall experiment, but how shall we learn? In What Works in Development? Thinking Big and Thinking Small (Jessica Cohen and William Easterly eds.), 24-47. Brookings Institution Press.
143. Rodrik, Dani. 2012. Why We Learn Nothing from Regressing Economic Growth on Policies. Seoul Journal of Economics 25:137-151.
144. Roe, Mark. 2006. Legal Origins, Politics, and Modern Stock Markets. Harvard Law Review 120:460-527.
145. Roe, Mark. 2007. Juries and the political economy of legal origin. Journal of Comparative Economics 35:294-308.
146. Rosenthal, Howard, and Erik Voeten. 2007. Measuring Legal Systems. Journal of Comparative Economics 35:711-728.
147. Schwartz, Richard, and James Miller. 1964. Legal Evolution and Societal Complexity. American Journal of Sociology 70:159-169.
148. Sekhon, Jasjeet. 2004. Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals. Perspectives on Politics 2:281-293.
149. Shulztiner, Doron, and Guy Carmi. 2014. Human Dignity in National Constitutions: Functions, Promises and Dangers. American Journal of Comparative Law 62:461-491.
150. Siems, Mathias. 2014. Comparative Law. Cambridge: Cambridge University Press.
151. Siems, Mathias. Forthcoming. Taxonomies and Leximetrics. In The Oxford Handbook of Corporate Law and Governance (Jeffrey Gordon and Wolf-Georg Ringe, eds.). Oxford University Press.
152. Sims, Christopher. 2010. But Economics is Not an Experimental Science. Journal of Economic Perspectives 24(2):47-68.
153. Slater, Dan, and Daniel Ziblatt. 2013. The Enduring Indispensability of the Controlled Comparison. Comparative Political Studies 46:1301-1327.
154. Smith, George Davey, and Shah Ebrahim. 2002. Data dredging, bias, or confounding: They can all get you into the BMJ and the Friday papers. British Medical Journal 325:1437-1438.
155. Spamann, Holger. 2009. Large–Sample, Quantitative Research Designs for Comparative Law? American Journal of Comparative Law 57:797-810.

156. Spamann, Holger. 2010a. Contemporary Legal Transplants – Legal Families and the Diffusion of (Corporate) Law. Brigham Young University Law Review 2009:1813-1877.
157. Spamann, Holger. 2010b. Legal Origin, Civil Procedure, and the Quality of Contract Enforcement. Journal of Institutional and Theoretical Economics 166:149-165.
158. Spamann, Holger. 2010c. The 'Anti-Director Rights Index' Revisited. Review of Financial Studies 23:467-486.
159. Spamann, Holger. 2008. Appendix to "The 'Anti-Director Rights Index' Revisited." Available at http://spamann.net/ASSETS/Spamann_ADRI_revisited_extrafiles.ZIP?id=374.
160. Spamann, Holger. 2015. The US Crime Puzzle: A Comparative Perspective on US Crime and Punishment. ssrn.com/abstract=2463720.
161. Spamann, Holger. Forthcoming. [On Civil and Common Law]
162. Strnad, Jeff. 2007. Should Legal Empiricists Go Bayesian? American Law and Economics Review 7:195-303.
163. Suchman, Mark, and Elizabeth Mertz. 2010. Toward a New Legal Empiricism: Empirical Legal Studies and New Legal Realism. Annual Review of Law and Social Science 6:555-579.
164. Watson, Alan. 1974. Legal Transplants.
165. World Bank. 2014. Doing Business 2015: Going Beyond Efficiency. Washington, D.C.: International Bank for Reconstruction and Development / The World Bank.
166. Ziliak, Stephen, and Deirdre McCloskey. 2007. The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives. University of Michigan Press.
167. Zweigert, Konrad, and Hein Kötz. 1998 [1996]. Introduction to Comparative Law (3rd ed., Tony Weir trans.).