

An Evaluation of Cross-National Measures of Judicial Independence

Julio Ríos-Figueroa
CIDE, División de Estudios Políticos

Jeffrey K. Staton*
Department of Political Science, Emory University

We provide a conceptual map of judicial independence and evaluate the content, construct, and convergent validity of 13 cross-national measures. There is evidence suggesting the validity of extant *de facto* measures, though their proper use requires attention to correlated patterns of measurement error and missing data. The evidence for the validity of extant *de jure* measures is weaker. Among other findings, we do not observe a strong and direct link between the rules that allegedly promote judicial independence and independent behavior. The results suggest that while the measurement of both *de jure* and *de facto* judicial independence requires a careful strategy for measuring latent concepts, the way that scholars should address this issue depends on whether they are targeting the incentives for independent behavior induced by formal rules or independent behavior itself. (*JEL* C19, C80, O43)

1. Introduction

Over the last quarter century, scholars have converged on an approach to institutional reform, in which judicial independence is viewed as critical to the promotion and maintenance of many aspects of human welfare. Modern institutional research suggests that independent judiciaries, which constrain arbitrary state power, ensure that state promises to respect individual rights are perceived credible (e.g., North and Weingast 1989). In turn, credibility breeds efficient investment, state solvency, growth and development (e.g., Barro 1997; Acemoglu et al. 2001; Frye 2004). Scholars also suggest that independent courts help stabilize democratic regimes by generally lowering the stakes of holding power (North et al. 2000). Researchers also have found an inverse relationship between judicial independence and human rights violations (Cross 1999; Keith 2002b; Hathaway 2007; Powell and Staton 2009).¹ More broadly, an

*Department of Political Science, Emory University, 1555 Dickey Drive 327 Tarbutton Hall, 30322, Atlanta, GA, USA. Email: jeffrey.staton@emory.edu.

1. Judicial independence is of course regarded well on normative grounds, as well. Most obviously, it is considered a fundamental component of the rule of law (e.g., Raz 1997: 198).

independent judiciary is commonly considered a necessary condition for a rights revolution (Epp 1998).

NGOs, states, and international organizations have promoted vigorously the construction of judicial independence around the globe; and more broadly, the rule of law (see Carothers 2006). To validate recommendations, to track reform performance, and to evaluate theoretical propositions, researchers have constructed a number of judicial independence indicators in recent years. We have identified 13 cross-national indicators, which cover a wide array of states and a substantial number of years. Clearly, we do not lack data. Yet, scholars have expressed concerns about the validity of cross-national judicial independence scores (e.g., Hammergren 2006: 14). Stasavage (2002: 51, fn. 17) has gone so far as to exclude the judiciary in his analysis of the effects of political institutions on private investment because “no accurate cross-country data are available to determine when and where the judiciary acts as a veto player with respect to policies which matter for investors.”

In this article, we evaluate the validity of extant cross-national measures of judicial independence. We find little evidence of a direct and strong relationship between (a) the formal, typically constitutional rules that scholars claim promote independent judging and (b) independent judicial behavior. Put simply, indicators of *de jure* and *de facto* independence are at best weakly correlated—in some cases, they are negatively related. Different *de jure* indicators are not even strongly correlated with each other. Since studies that make use of *de jure* measures assume that rules create strong incentives for independent behavior, at a minimum, our findings raise questions about the proper interpretation of empirical results connecting formal rules of judicial independence to, say, economic freedom (e.g., La Porta et al. 2004). This is not to deny associations scholars have reported in prior research. It would certainly appear that long judicial tenure may be associated with indicators of economic freedom, but it is unclear that it is because these rules directly incentivize judges to behave independently of the government. The stakes of developing measures that are tightly linked to a careful theory of incentives are clear. Until we can measure properly the rules that promote independent judging, it is unclear what to make of the scholarship suggesting powerful effects of these rules. Theoretical accounts linking judicial institutions to judicial behavior need to be far more sensitive to the particular incentives rules likely to establish and to the political and social factors that likely condition the effects of these institutions.

We also find evidence suggesting that *de facto* judicial independence indicators are capturing the concept they purport to measure. This is true even though scholars often use proxy indicators that include surprising content in light of the targeted concept. Having said that, measurement error would seem particularly large in these indicators, and nonrandom missing data are common. The stakes of addressing these problems are clear, as well. Patterns of measurement error and missing

data are related in ways that undermine typical approaches to evaluating robustness. Measurement error is larger and data are more likely to be missing in underdeveloped states. For this reason, when we evaluate the robustness of our findings to alternative measures without addressing the missing data problem, we are more likely to conduct our analysis on the set of states for which our measures are most likely to agree. The consequence is that we are likely to be more confident in the robustness of our results to alternative indicators than we should be.

Reflecting on these findings, it is helpful to keep in mind that whether we are thinking about the incentives rules create for independent behavior or independent behavior itself, we have in mind concepts that are not directly observable. They are latent, and for this reason, measurement requires inference (Treier and Jackman 2008). By implication, we can be assured that our concepts are measured with error. The continued use of existing *de facto* indicators without a coherent strategy for addressing measurement error and missing data risks a variety of biases. In this context, a measurement model would seem to offer a productive way forward, especially with respect to measuring *de facto* independence, where it would appear that existing efforts to capture the concept are, roughly speaking, on target.

We divide the remainder of the article as follows. Section 2 provides a conceptual map of judicial independence and reviews the availability of the indicators around the world and over time, both activities we hope can serve as a guide to both users of existing measures and potential creators of new ones. Section 3 reports the results of our validation analysis of the extant indicators. Section 4 concludes.

2. A Conceptual Map of Judicial Independence

Important conceptual differences characterize the literature on judicial independence (Russell and O'Brien 2001; Burbank and Friedman 2002).² Still, we can distinguish between a limited number of ways in which scholars use the term, and in this way offer a conceptual map that is useful for dividing the available measures. In this section, we define terms, raise two special conceptual concerns, and summarize the availability of indicators of these concepts across the world and over time.

2.1 *De Jure* and *De Facto* Judicial Independence

A familiar distinction in the literature involves the difference between *de facto* and *de jure* concepts (e.g., Feld and Voigt 2003). The latter deals with formal rules designed to insulate judges from undue pressure, either from

2. There may be subtle but important differences regarding independence concepts depending on to whom independence refers (e.g., the highest court, a court, a judge, the entire judiciary), and independence from whom (e.g., other political branches, superior in the judicial hierarchy, parts in a trial). In this article, however, we focus on the broader conceptual ideas.

outside the judiciary or from within. Institutions like fixed tenure, multi-lateral appointment procedures, budgetary autonomy, and judicial councils are thought to provide such insulation and thus influence behavior by promoting judicial *autonomy* or *influence* subconcepts that we develop below (e.g., Rosenberg 1991).

The concept of *de facto* judicial independence is behavioral and can be further differentiated between two subconcepts. The first demands that judges be the “authors of their own opinions” (Kornhauser 2002: 42–55). On this account, a judge is independent when she does not respond to undue pressures to resolve cases in particular ways.³ In other words, a judge is independent when her decisions reflect her preferences (Becker 1970: 1–8). Judicial independence in this sense reflects autonomy. What judges think sincerely about the record controls the outcomes of their cases. A second concept of *de facto* judicial independence recognizes that lacking financial or physical means of coercion, courts depend on the assistance of other political authorities to enforce their decisions (Hamilton et al. 2009). Under this second concept, the argument is that it makes little sense to call a judge independent if her decisions are routinely ignored or poorly implemented. Judicial independence requires not only that judges resolve cases in ways that reflect their sincere preferences, but also that these decisions are enforced in practice even when political actors would rather not comply. Judicial independence in this sense reflects influence—that there is a causal relationship between what how judges think the underlying conflict they are adjudicating should be resolved and how it is resolved in practice (Cameron 2002).

It is important to stress that although these concepts are distinct, they are obviously related. An influential judge is autonomous. If she were not, she would not say that there is a causal relationship between judicial preferences and outcomes. Yet, an autonomous judge need not be influential. Decisions may reflect sincere judicial preferences but nevertheless be incompletely implemented or completely ignored. The two subconcepts of independence emerge in familiar arguments. For example, an argument in which independent courts serve as veto players assumes that decisions are final (Ferejohn and Shipan 1990). Arguments in which independent courts allegedly enforce international agreements, and by doing so raise the costs of ratification to some states, must assume compliance (Hathaway 2007). These types of arguments suggest a concept of *de facto* judicial independence understood as influence. In contrast, a model like Weingast (1997), in which citizens can make use of constitutional signals to help coordinate their responses to a government

3. Of course, political authorities always apply some pressure when they are parties to a case or when they file briefs as interested parties (e.g., *amicus curiae* briefs and other similar institutions). The key here is that this pressure is perfectly acceptable within the rules of the legal system.

that has exceeded its constitutional limits arguably might only require a court-like actor issuing decisions that reflect its sincere evaluation of the record, i.e., autonomy.

2.2 Challenges of Measuring *De Jure* and *De Facto* Independence

The primary measurement challenge scholars confront is that judicial independence, however defined, is not directly observable. This might appear a strange claim to make about *de jure* independence, where it would seem that scholars are looking to measure constitutional rules that are directly observable in, well, constitutions. Yet, this is not quite what scholars are after. They are not looking to measure formal rules *per se*. Rather they are looking to measure the incentives rules are likely to induce, and these incentives are very much unobservable. Research teams have confronted this challenge with a reasonable approach. They aggregate information via an index of multiple institutions (e.g., length of tenure and methods of appointment and removal), each of which are believed to create the “right” incentives (e.g., Keith 2002b; Feld and Voigt 2003; La Porta et al. 2004). This is perfectly sensible when we understand well the connections between particular rules and behaviors; however, it is unclear that the literature on judicial institutions provides the kind of theoretical and empirical support necessary to justify an indexing of this sort. In fact, there are arguments in the field suggesting that the same judicial institution can have contradictory effects on judges’ incentives (Helmke and Staton 2011) and provide different incentives depending on political context (Pozas-Loyo and Rios-Figueroa 2006).

Even more problematic are scenarios in which scholars wish not so much to measure incentives but actual behavior, and nevertheless depend on a *de jure* indicator. In such cases, researchers depend critically on a coherent theory of behavior produced by judicial institutions. For instance, La Porta et al. (2004: 447) argues that judicial independence promotes both economic and political freedom “the former by resisting the state’s attempts to take property, the latter by resisting its attempts to suppress dissent.” To proxy this *de facto* judicial independence concept, they use a *de jure* index. This kind of design depends critically on a direct link between rules and independent behavior.

Like *de jure* independence, the measurement of *de facto* judicial independence involves saying something about behaviors that are not directly observable. Fortunately, the field possesses a far more developed literature, suggesting ways in which the concept ought to manifest. Moreover, even in the absence of a careful theoretical argument, observing contradictory decisions by the same judge on remarkably similar issues or observing political officials simply ignoring judicial decisions provide relatively clear signals about autonomy and influence. Although most scholars continue to construct single proxy indicators, others have attempted to

combine multiple sources of information via a deterministic index (e.g., Henisz 2000; Feld and Voigt 2003).⁴

The theoretical literature raises another, particularly difficult challenge for the measurement of *de facto* independence. Whether one adopts the influence or autonomy concept, it is worth considering whether and how our measurement strategies address the behavioral implications of strategic models of judicial decision-making. Simmons (2009: 22), writing about the advantage of litigation strategies to advance human rights norms, provides a representative of the kind of behavior we have in mind. She writes:

One of the most important conditions for litigation to be a potentially useful strategy to enforce rights is judicial independence. For courts to play an important enforcement role, they must be at least somewhat independent from political control. The government or one of its agencies, representatives or allies is likely to be the defendant in rights cases, and unless local courts have the necessary insulation from politics, *they are unlikely to agree to hear and even less likely to rule against their political benefactors*. Anticipating futility, individuals or groups may decide to avoid the courts altogether.

The kind of strategic judicial deference Simmons alludes to is anticipated by common models of judicial–government interaction (e.g., Helmke 2005; Vanberg 2005), and has received considerable empirical support in studies around the globe (Ginsburg 2003; Herron and Randazzo 2003; Helmke 2005). If these models are correct, then nonsystematic observation of judicial decision-making, litigation strategies, and compliance as a means of developing measures of independence can be significantly misleading. Indeed, a court that offers little constraint on government can appear to be highly constraining if it chooses its cases wisely. It certainly can ensure systematic compliance, which makes inferring judicial influence from the outcomes of legal conflicts difficult.

Once we recognize that judicial decision-making can be strategic, the core measurement challenge is that direct observation of judicial behavior without estimating how judicial decision-making is influenced by political concerns can be misleading. A court that is neither autonomous nor influential could appear to be both, say if it strategically selects and decides cases so as to minimize conflict (e.g., Ginsburg 2003). Decision-making may appear autonomous if we fail to consider case selection, which removes controversial cases from the sample. And a court can appear influential as evidenced by a very high rate of compliance, but only because it

4. Given our findings on *de facto* measures, we would encourage scholars to build on the logic of aggregation evident in the Feld and Voigt and Henisz approaches, by say the development of a probabilistic measurement model.

does not ask governments to do what they would not. Carrubba (2005) has discussed this problem in the European context, and Kastellec and Lax (2008) have shown in the US context that not taking into account judicial strategy can result in biased measures of lower court's compliance with judicial decisions.

For this reason, a precise measure of independence derived from court behavior requires a wealth of case-specific data, which allows for the systematic estimation of the extent to which judicial decisions respond to external, political pressures, and the extent to which judicial decisions are properly implemented.⁵ For a worldwide sample, comparative judicial politics has yet to produce such data.⁶ In light of this problem, scholars might look to measures of autonomy or influence that capture the types of behaviors we should observe if the judicial system functions as a genuine constraint on the state. This is not to say that court-specific measures are invalid. Rather, the point is that given central findings in judicial politics on strategic judicial behavior, we might consider being creative about our measurement choices until we can systematically estimate judicial effectiveness around the world.

One possible solution involves seeking a measure of the behavioral or attitudinal consequences of autonomy or influence, which are not the subject of the analysis. These manifestations ought to follow from a theoretical argument. For example, consider Staton's (2010) model of judicial independence, in which observers of judicial–government interactions develop beliefs about judicial autonomy by considering whether decisions reflect plausible legal interpretations or are best understood through the prism of overt political jurisprudence. The key prediction from the model is that in political contexts that induce little strategic judicial deference, there should be a strong, positive relationship between awareness with a court and perceptions of autonomy; however, in political contexts that frequently induce strategic deference, the effect should be attenuated, potentially negative. Under this kind of model, the estimated relationship between awareness and perceptions of judicial impartiality as found in public opinion data could proxy judicial autonomy. Clague et al.'s (1999) Contract Intensive Money (CIM) score suggests a similar approach for measuring influence. The CIM returns “the ratio of non-currency money to the total money supply” (Clague et al. 1999: 188). Conceptually, high values of CIM reflect a society's belief that property rights institutions constitute credible constraints on state predation. In so far as the judiciary is one salient element of a state's cluster of property rights

5. Even in countries with good case databases, there are potential biases if one considers only published and not all judicial decisions (Keele et al. 2009). Judges many times are strategic with regards to which decisions they publicize (Staton 2010), which would also bias measures based on the decisions available on the court's web pages, for example.

6. Tate et al. (2006) and Martin et al. (2006) have begun to develop a broadly comparative data set on high court decisions; however, at present it covers only 11 states. Thus, we are a long way away from having worldwide data on judicial systems.

institutions, we might entertain the possibility that it returns information on judicial influence. We do not necessarily advocate either approach here, and believe that this information might be more fruitfully leveraged as one of the many manifestations of influence or autonomy in a measurement model rather than as single proxies (though we will evaluate the validity of the CIM below). Our point is that creative, theoretically driven approaches to measurement are essential when we believe that our concept is not only latent but also subject to a variety of strategic pressures.

2.3 Data Availability

We review 13 cross-national measures of judicial independence, which social scientists have used in empirical research. Table 1 provides basic information on the concept that each measure tries to capture, the type of sources from which the measures were created, and the coverage of each measure across time and space. The appendix contains detailed information on the construction of the measures summarized in Table 1.

The table summarizes a number of salient features of the measurement landscape. Conceptually, it is clear that many more research teams have attempted to measure *de facto* independence than *de jure* independence; and, within the former category, scholars largely seem to have targeted the influence concept. Second, extant measures limit the time frame, which can be evaluated empirically. There are only two measures available that provide data prior to the 1980s, and neither are unambiguously designed to capture judicial independence only (see Appendix A).⁷ The Polity IV measure (XCONST) captures, in part, legislative independence. The Political Risk Service's (PRS) measure aims to capture order in the society, as well as the quality of a state's legal system. The Henisz (HEN) measure is derived from XCONST and PRS, so that it inherits their composite approach. And finally, the Clague et al. measure (CIM) is a proxy score, which more broadly targets the extant to which property rights are protected. Third, there is considerable variation in the breadth of coverage offered by each measure. For example, La Porta et al.'s (LAP) *de jure* score, the measure with the narrowest coverage, provides a score for only 69 states. In contrast, Tate and Keith's (T&K) *de facto* measure covers 191 countries—very nearly the entire set of autonomous states in the world in 2004. It is important to note, however, that the number of countries covered varies by year. For instance, the XCONST measure covers 106 countries in 1960 and 158 in 2008, and the Global Competitiveness Report's (GCR's) measure covers 56 countries in 1995, yet 114 in 2005.

7. There are three measures that are cross-sectional only: Feld and Voigt's measures of *de jure* and *de facto* independence (F&V-J and F&V-F, respectively), and LAP's measure of *de jure* independence. Feld and Voigt's (2003: 503) measures are based on a questionnaire sent to country experts in which they were asked to comment on how independent the judiciary has been since 1960, so it could be understood as a 40-year average and let it be constant across all years. In the analyses that follow, we take its value for 2003, the year in which the article was published.

Table 1. Thirteen Indicators of Judicial Independence

Measure	Concept			Source	States	Years	
	<i>De Jure</i> concept	<i>De Facto</i> concept					
		Autonomy	Influence				Unclear
BTI			X		Country experts reports	119	2006; 2008
CIM			X		National economic data	169	1960–2000
CIRI			X		US State Department country reports	190	1981–2009
F&V-F			X		Expert Survey	85	Cross-sectional ^a
F&V-J	X				Expert Survey	85	Cross-sectional ^a
GCR				X	CEO's Survey	114	1995; 2000–05
HEN			X		Combination of PRS and XCONST	170	1980–04
H&C		X			US State Department country reports	176	1992–99
KEITH	X				Constitutions coded by author	176	1976–96
LAP	X				Constitutions coded by author	69	Cross-sectional ^a
T&K			X		US State Department country reports	191	1990–2004
PRS			X		Staff coding of several sources	140	1984–2009
XCONST				X	Polity IV coding of several sources	163	1800–2009

Lists of the following 13 measures BTI measure of Judicial Independence (2008); Measure of CIM by Clague et al. CIM (1999); CIRI = Cingranelli & Richards (2008) measure of Judicial Independence; F&V-F and F&V-J = Feld and Voigt's (2003) measure of *De Jure* and *De Facto* Judicial Independence; GCR = measure of "judicial independence"; Wittold Henisz's (2000) measure of judicial independence; H&C (2004) measure of judicial independence; KEITH = Keith's (2002) measure of judicial independence; La Porta et al (2004) measure "Judicial Checks" and Balances"; T&K = Tate & Keith (2007) measure of judicial independence; and, XCONST = Polity measure on "Constraints on the Executive" (Gurr 1990). The number of countries covered varies per year for time series measures. Table 1 reports coverage in the latest available year.

^aFeld and Voigt's measures are based on a questionnaire sent to country experts in which they were asked to comment on how independent the judiciary has been since 1960 (2003: 503). Their scores can be understood as a 40 years average. We take its value for 2003, the year in which the paper was published.

Table 1 also shows the primary sources used for the construction of each measure, revealing a few interesting issues. Although the three *de jure* measures use national constitutions as their primary source, one of them is based on a survey of experts who were asked about features of their constitutions (Feld and Voigt 2003). Also, the *de jure* teams do not focus on the same institutional features. While La Porta et al. (2004) focuses largely on the length of the tenure and the constitutional review powers of the judges, Keith (2002, KEITH) coded a more comprehensive list of *de jure* features including the ban against exceptional or military courts. As we show below, these choices are consequential. Notice also that three of the *de facto* measures—T&K, Howard & Carey (H&C), and Cingranelli and Richards (CIRI)—share the same primary source: the US State Department’s yearly human rights country reports. Nevertheless, the teams neither did intend to measure the same concept nor did they operationalize their measures identically.

3. Validation Analysis

We now turn to an analysis of the validity of the 13 measures we have identified. We begin with the content of the measures. We then consider the extent to which they provide similar information about concepts of interest and whether they explain outcome variables that are theoretically connected to judicial independence.

3.1 Content Validation

Do the indicators contain content reflective of the targeted concepts? We highlight typical content and raise two obvious concerns about content validity.

3.1.1 De Jure Measures. In general, the content of *de jure* independence measures appears sensible in the light of general understandings and beliefs about the kinds of institutions that are likely to incentivize independent judicial behavior. Consider the three indexes that we evaluate. Feld and Voigt (2003) produces an interval-level measure of *de jure* independence derived from surveys of country experts. Their scores aggregate information on 12 formal rules that govern the constitutional status of a state’s highest court, its appointment and retention rules, and its judicial review authority (see Appendix A). La Porta et al. (2004) combine the length of judicial tenure on the highest ordinary and administrative courts and whether the constitution states that judicial decisions are a source of law. Keith (2002a) produces seven ordinal measures based on the United Nations principles of judicial independence. The scores indicate whether a state’s constitution formally guarantees the tenure of high court judges; ensures the finality of judicial decisions; grants judges exclusive authority over their jurisdiction; bans special or military jurisdiction for civilian crimes; financial autonomy; a separation of powers system; and

specifically enumerates appointment qualifications. Apodaca (2004) creates an additive index from these seven items, which we evaluate.

Unsurprisingly, each of the indicators include features of judicial tenure. Two of the three, Feld and Voigt (2003) and Keith (2002a), include content related to judicial salaries or other resources that might be manipulated politically. Still, it is important to stress that the indicators are far from identical. Only Feld and Voigt specifically evaluate the process by which judges are appointed, treating appointment procedures that involve less political selection (e.g., where professional organizations are involved) as more independent than procedures in which appointment is carried out entirely by the political branches. Only Keith includes information on whether judges have exclusive control over their jurisdiction. And while Keith's scores are available for a reasonable time series, both the Feld and Voigt and LAP measures are available cross-sectionally.

3.1.2 De Facto Measures. As Table 1 suggests, most *de facto* measures seem to target influence. Yet, the Howard and Carey (2004) (H&C) measure offers a good example of a *de facto* independence measure aimed at the concept of autonomy. Howard and Carey (2004: 286) define judicial independence as, "The extent to which a court may adjudicate free from institutional controls, incentives, and impediments imposed or intimidated by force, money, or extralegal, corrupt methods by individuals or institutions outside the judiciary, whether within or outside government." This measure, which is derived from US State Department country reports, is a three-category ordinal scale, which indicates whether the high court of a state in a particular year is fully independent, partially independent or dependent. According to Howard and Carey (2004: 287–88), a state has a fully independent judiciary if the high court functions in practice:

- independently of the executive and legislature, and is
- relatively free from corruption and bribery and provides basic criminal due process protections to criminal defendants.

A state has a partially independent judiciary if its high court either satisfies the first or the second condition or partially satisfies both. A state has a dependent judiciary if its high court satisfies neither condition. In general, the measure is reasonably well connected to the concept. That said, Tate and Keith (2009), who also make use of the State Department's country reports to generate their measure, question whether H&C's focus on criminal due process rules moves the measure away from standard concepts of judicial independence. If H&C have the autonomy concept in mind, then this concern is certainly apt. It is possible for two courts to be equally the "authors of their opinions" while the first provides searching review of due process allegations and the second is instead highly deferential in this context.

Henisz (2000) provides a representative example of an influence, *de facto* independent measure. HEN wishes to measure the extent to which a state's judiciary serves as a constraint on government. To that end, the HEN measure incorporates information from two other indicators of the influence, *de facto* independence concept: the Polity IV Executive Constraints (XCONST) measure and the PRS law and order measure. The XCONST indicates the extent to which a state's chief executive faces constraints on its decision-making process. According to the team, one (but not the only) constraint on this authority is an independent judiciary, and for this reason, states receive a higher XCONST score when their judiciaries are deemed independent by the Polity coders. The PRS measure combines a law component, intended to measure the "strength and impartiality of the legal system," with an order component, intended to measure popular observance of formal legal rules. HEN assigns a score of 1 to states that are coded sufficiently high on both the XCONST and PRS scores and a 0 otherwise (see Appendix A for coding rules).

3.1.3 Questions of Content Validity. The content of these indicators raise two kinds of flags about validity. The first concerns indicators where it is clear that the research team was aiming directly at one of the concepts of judicial independence we summarize above, yet included seemingly unconnected content. Consider the following list of representative examples.

- Feld and Voigt's *de jure* index includes information on the accessibility of the judiciary,
- LAP's *de jure* index includes information on whether the constitution states that judicial decisions are a source of law, and
- H&C's *de facto* measure asks whether there is evidence that the courts provide basic due process rights to the criminally accused.

It would seem inappropriate to include accessibility rules in a scale of independence, especially when these rules serve to measure well the concept of "access." Similarly, information about the extent to which judicial decisions are considered sources of law (or perhaps even an indicator for the existence of constitutional review authority) seems closer to a *de jure* measure of judicial power or formal authority than of judicial independence (Rios-Figueroa 2011). And Tate and Keith (2009) themselves have questioned whether H&C's focus on criminal due process is really linked tightly to independent behavior. The idea is that it would seem possible for two courts to be equally autonomous even if the first requires a searching review of criminal procedures, yet the second is highly deferential in this context.

A second flag concerns indicators where either the original research team did not target judicial independence explicitly (even if their measures

are used as proxies) or where the original team explicitly combined two very different concepts of judicial independence. The most obvious example of the first scenario is the CIM that returns no direct information on the judiciary. Henisz's *de facto* measure that makes use of XCONST and PRS provides another representative example. First and foremost, XCONST was designed by the Polity IV project to capture "constraints on the executive." An independent judiciary is one such constraint, but XCONST itself contains information on alternative mechanisms by which chief executives may be constrained, such as an independent legislature that is ideologically distinct from the executive. Second, since the PRS measure attempts to pick up information on both social order and law, the HEN measure risks treating orderly societies as those with independent courts. The PRS measure itself can be relatively high in highly ordered societies, even if there is little evidence of judicial independence. This is true, for example, of Cuba.

The Cingranelli and Richards (2008) measure provides another illustrative example. The team is attempting to measure a *de facto* concept, yet they include *de jure* information. Like H&C and T&K, CIRI is derived from the US State Department's Human Rights Reports. It provides a three-category ordinal independence score. Judiciaries that face "active government interference or corruption" may be coded as partially independent or dependent, depending on the degree of external pressures or corruption. However, to be coded as fully independent, a state's judiciary must satisfy the following criteria:

- it has the right to rule on the constitutionality of legislative acts and executive decrees;
- judges at the highest level of courts have a minimum of seven-year tenure;
- the President or Minister of Justice cannot directly appoint or remove judges and the removal of judges is restricted (e.g., allowed for criminal misconduct);
- actions of the executive and legislative branch can be challenged in the courts;
- all court hearings are public; and
- judgeships are held by professionals.

Notice that these criteria all provide *de jure* information. Although the distinction between partial and no judicial independence is based on *de facto* criteria, CIRIs appear to reserve the highest category for states that satisfy this set of formal rules.

3.1.4 Summary. Each *de jure* measure we review contains some content that would seem directly connected to independence, yet some content that seems better connected to other concepts, like access or formal

judicial power. And though nearly all of the *de facto* indicators contain directly connected content, others were originally designed with related concepts in mind or combine elements of *de jure* and *de facto* independence. It is tempting to conclude that many of the extant measures, perhaps all, have low content validity. This is transparent in the case of a proxy like the CIM. But it is no less true of indicators like Feld and Voigt's *de facto* index that treats accessibility as a feature of judicial independence or LAP's and Keith's *de jure* indexes that treat the legal status of judicial decisions and judicial powers as rules that incentivize independent judging.

It is worth remembering, however, that every research team is wrestling with a latent concept, and so teams depend critically on theories, beliefs at least, about how the concept is likely to manifest. Feld and Voigt's rationale for including information about accessibility is that courts that are hard to access will be less effective (i.e., influential) in constraining governments—perhaps they even might be less likely to try. A similar logic could support including information on judicial review in a *de jure* index. Even content that seems completely reasonable depends on an underlying theory. Observed judicial tenure (or the turnover rate for judges) provides a clear example. The logic for including this kind of information is simple. Higher turnover rates reflect judiciaries that are being politically tampered with. But we might observe relatively low rates of turnover when judges are highly sensitive to government desires, and for that reason do not create incentives for purges.

Clearly, we prefer content that is connected to a clear concept and a coherent theory linking the concept to its observable manifestations. In many cases, however, scholars will reasonably disagree about precisely the right content to include. Even if they share conceptual definitions, they may not share theoretical understandings about how the concept manifests. This lack of theoretical consensus reinforces a central theme of this article—we must expect to err in measurement.

3.2 Convergent Validation

Convergent validation considers whether different indicators of the same concept return similar information about that concept (Adcock and David Collier 2001: 540). Ideally, we would like a benchmark indicator of each judicial independence concept against which we could compare new measures. Unfortunately, there is no benchmark measure of either *de facto* or *de jure* judicial independence. Still, we can evaluate the extent to which existing measures are positively associated with each other. Further, since formal rules are supposed to create strong incentives for independent behavior and since a measure of independence designed to uncover influence should also capture autonomy, we should observe reasonable convergence among all of the indicators.

Figure 1 displays correlation coefficients for each of our measures, first with respect to the XCONST measure and then for the T&K

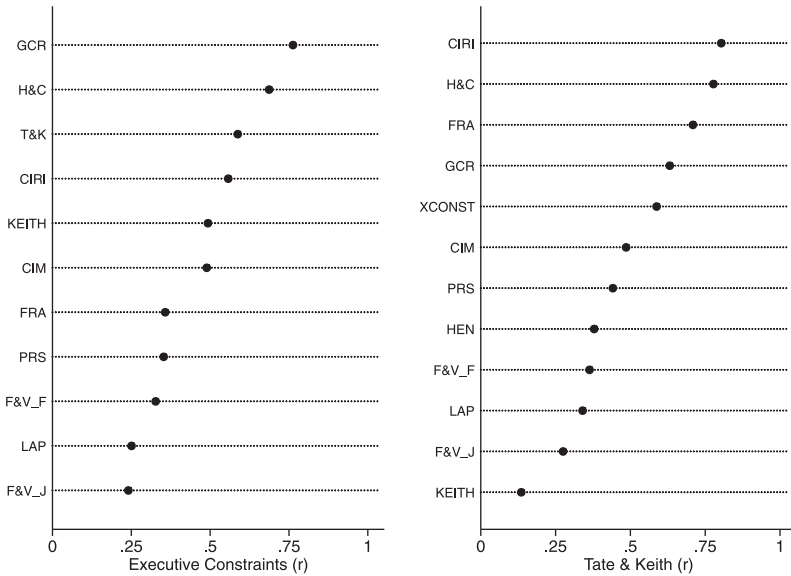


Figure 1. *Convergence of Judicial Independence Measures*. Dot plots of correlation coefficients for Polity IV and T&K. Results for *de jure* measures are circled.

measure.⁸ The figure suggests mixed results for scholars in need of measures of *de facto* independence. The correlations between the *de facto* measures are all positive, and some are quite large.⁹ Importantly, however, and as Haggard et al. (2008) have found in their work on the rule of law, there is considerable variation in the degrees of association.¹⁰ Figure 1 also suggests that the *de jure* and *de facto* measures do not converge. With

8. Results are expressed in Pearson's r , though some of our measures are ordinal. We have also calculated the Tau b coefficients for each of the ordinal measures, and the results are substantially similar. Table 6 contains a complete correlation matrix for these indicators.

9. The HEN measure is entirely determined by the values of the Polity IV and PRS law and order measures. For this reason, it is unclear that a positive correlation between HEN and the measures from which it derives would say anything useful about the validity of these scores. Thus, we exclude HEN measure from the left panel of Figure 1 and from the top panel of Figure 2.

10. We should also note that some of the strongest relationships are found between measures derived from the same primary source (e.g., US Department of State reports), yet it is important to recall that only two of these measures intended to capture the influence concept (T&K and CIRI), whereas the third aimed at autonomy (H&C). And clearly, the teams used different rules for operationalizing the concept. As it turns out, H&C, the autonomy measure, is more strongly associated with both State Department influence measures than they are with each other. More importantly, the core point of this analysis is that there is considerable variation in the degrees of association, and we observe this variation even among measures that are derived from completely independent sources. For example, compare the correlation between T&K and GCR (0.710) with the correlation between Polity IV's XCONST and Feld and Voigt's *de facto* scores (0.33).

one exception, in the left panel, the correlations between the *de jure* and *de facto* measures are relatively small.

Figure 2 highlights a different concern. The *de jure* measures are not strongly correlated with each other. Even the smallest correlation between all of the *de facto* measures and the Polity IV indicator in this figure is larger than the largest correlation between LAP and the other *de jure* measures. The *de jure* research teams have not focused on precisely the same institutional features, and we find no evidence suggesting that there is a bundle of good institutions of judicial insulation from which designers are picking. If there was, then focusing on different pieces of each state's suite of judicial institutions would result in roughly the same measure of *de jure* independence. This is not what we observe.

To summarize, although there is no benchmark judicial independence measure, there is some evidence of convergence among the *de facto* measures. Yet, we would not wish to conclude that these measures are perfect substitutes for each other. They seem to return related, but not identical information. The same cannot be said for the *de jure* measures. Under the convergence criterion, at least, it appears that our measures may be tapping into different configurations of rules that allegedly promote judicial insulation.

3.3 Construct Validation

Construct validation asks whether measures of a concept are related to measures of alternative concepts in ways anticipated by theoretical models (Adcock and David Collier 2001: 537). In this section, we summarize two construct validity tests in which we replicate well-known studies in the literature on law and comparative politics: Acemoglu et al.'s (2001) analysis of institutions and economic development and La Porta et al.'s (2004) analysis of judicial independence and political rights.

Acemoglu, Johnson and Robinson (AJR) evaluate the claim that legal protections for property rights are associated positively with economic development. To measure institutional protections, the authors use the 10-year average (1985–95) of the PRS index of protection against expropriation (Acemoglu et al. 2001: 1377–8). Since independent judiciaries are typically viewed as core components of the institutional architecture for the protection of property rights, AJR's analysis offers a straightforward opportunity to evaluate the construct validity of the extant independence measures. AJR's core contribution to the literature involves their strategy for addressing a familiar endogeneity concern in rule of law models of economic development. If higher economic development permits states to craft quality legal institutions, simply regressing measures of development on measures of institutions will result in biased institutional estimates—unaddressed endogeneity will inflate coefficients. AJR instrument for institutional protections with a measure of mortality rates among colonial settlers in the 19th century. The logic behind the instrument is that where Europeans faced hospitable environments, they built institutions that

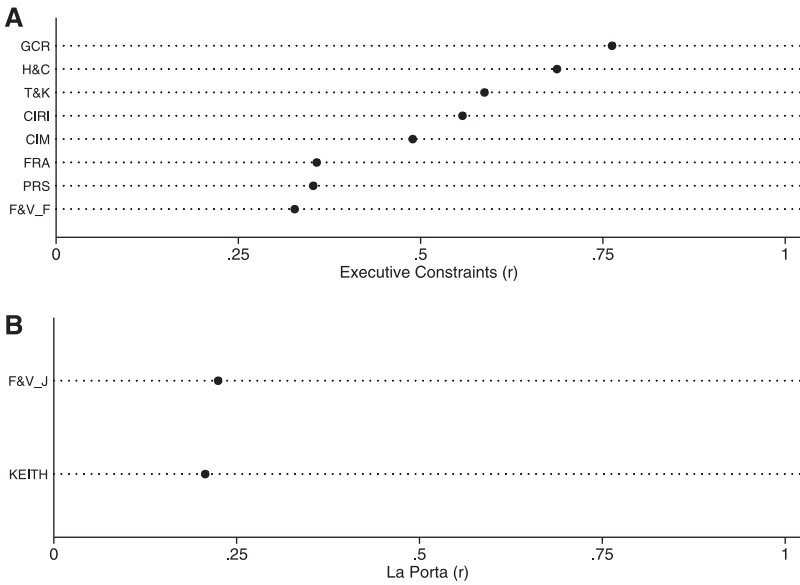


Figure 2. *De facto* & *De jure* Convergence. Dot plots of correlation coefficients for *de facto* (A) and *de jure* (B) measures of judicial independence.

reflected those at home, anticipating future migration. Where settlers faced relatively inhospitable environments, colonizer states did not construct such institutions, much less allow significant local rule. Finally, original institutional designs have been sticky, so that once constructed they tended to persevere over time.

Table 2 summarizes the results of two-stage least squares (2SLS) models of economic development, replicating AJR's models for all but two of our measures.¹¹ Following AJR, we take the 10-year average of each of our judicial independence measures. To ease the construction of a comparable average measure of judicial independence over the 1985–95 period, we

11. We exclude BTI from the analysis because it was not available until 2006. The LAP and *de jure* measures were collected after 1998; however, in so far as the institutions the teams evaluated do not change greatly over short periods of time, we included them in the AJR replications nonetheless. This is consistent with how LAP use their measure. Also, in constructing their *de facto* index, Feld and Voigt ask experts to comment on their judiciary's behavior since 1960. For this reason, it is probably best thought of as a 40-year average score for each state. We also exclude the KEITH results from our summary. Like LAP, we did not find significant results for KEITH. This suggests further evidence that the *de facto* measures are tapping into a distinct conceptual dimension than the *de jure* measures. That said, the 2SLS estimates for KEITH were two orders of magnitude larger than those obtained from the single equation OLS estimation. Diagnostic checks suggested that core model assumptions were significantly violated in this case. In order to keep the models, we estimate absolutely consistent with AJR and yet not report results that we do not believe, we exclude KEITH from the table.

Table 2. AJR Replication

	First stage: legal Institutions		Second stage: log GDP/cap, 1995		N
	β (SE)	p	β (SE)	p	
Original study					
AER	-0.34 (0.18)	0.07	1.10 (0.44)	0.07	64
<i>De facto</i> measures					
CIM	-0.03 (0.01)	0.01	10.69 (3.62)	<0.01	77
CIRI	-0.07 (0.02)	0.01	4.96 (1.89)	0.01	78
F&V-F	-0.04 (0.03)	0.19	8.73 (6.34)	0.17	42
GCR	-1.14 (0.31)	<0.01	0.36 (0.12)	0.17	24
HEN	-0.08 (0.05)	0.09	4.11 (0.08)	0.08	78
H&C	-0.11 (0.05)	0.05	3.26 (1.57)	0.04	77
PRS	-0.23 (0.15)	0.11	1.65 (0.86)	0.05	68
T&K	-0.08 (0.05)	0.11	4.57 (2.67)	0.09	77
XCONST	-0.38 (0.17)	0.03	0.81 (0.37)	0.03	74
<i>De jure</i> measures					
F&V-J	-0.05 (0.02)	0.05	9.00 (4.19)	0.03	74
LAP	-0.07 (0.06)	0.24	6.04 (4.84)	0.21	77

Replications two-stage least squares models of economic development, presented in Table 4, Column 8 of Acemoglu et al. (2001), using a state's AER. Consistent with the original models, all institutional measures reflect the average score for each indicator from 1985 to 1995. When this was impossible, we took the score closest to 1995. We used binary versions of the State Department measures, coded 1 for "fully independent" and 0 otherwise. Thus, the average reflects the proportion of years in which the judiciary was coded as fully independent. Control variables include average latitude and continent dummies, which are not reported.

transformed the State Department measures into binary indicators of judicial independence. The average of these measures thus reflects the proportion of years between 1985 and 1995 that states possessed a fully independent judiciary.

The table displays the estimates for the European settler mortality measure in the first-stage equation in the left column and the 2SLS estimates for the institutional measures in the right column.¹² The signs of both the first-stage estimates of settler mortality and the 2SLS estimates of judicial independence all reflect the AJR results. Four of the 2SLS estimates fail to reach statistical significance at the 0.10 level, yet this is true for two of the three *de jure* indicators.¹³ The substantive effects of the measures also reflect what AJR find. The original model indicates that a 1 SD change in the PRS index of expropriation risk is associated with a 1.98

12. AJR present a number of specifications. We show the results of replicating the model they report in Table 4, Column 8, which includes average latitude and continent dummy variables as controls. We have also controlled for colonial experience and legal tradition. The results of those models are stronger and more consistent with each other than what we report in Table 2. Thus, the results summarized in Table 2 are conservative.

13. KEITH result not shown for reasons discussed in footnote 11. For each judicial independence score, Wu-Hausman tests suggest that the institutional measures are not exogenous of the development measure.

increase in the log of GDP per capita. This represents a 1.8 SD change in economic development. The judicial independence measures produce similar results, some slightly stronger and some slightly weaker. For example, a 1 SD change in the PRS law and order score yields a 2.1 increase in the log of GDP. Similar changes in the CIM, the XCONST score, and the GCR judicial independence score yield 1.45, 1.54, and 0.68 changes, respectively. Changing the binary State Department measures of judicial independence from 0 to 1 results in no less than a 3.26 change in the log of GDP. On the whole, the judicial independence measures, the *de facto* measures especially, perform well on this test.

In addition to promoting economic development, scholars have suggested that judicial independence can promote democratic values (e.g., North et al. 2000).¹⁴ La Porta et al. (2004: 447) write: “In principle, judicial independence promotes both economic and political freedom, the former by resisting the state’s attempts to take property, the latter by resisting its attempts to suppress dissent.” Table 3 summarizes the results of our replication of La Porta et al. (2004) finding that *de jure* judicial independence is positively associated with political rights. The dependent variable in each model is the Freedom House Political Rights index (reversed so that higher values reflect higher levels of political rights).¹⁵ All of the estimates in Table 3 are signed similarly to the original measure, and only two measures fail to reach statistical significance at the 0.10 level. Six are significant at the 0.01 level. Interestingly, and as was true in the AJR models, the Feld and Voigt *de facto* measure is not significant, yet the Feld and Voigt *de jure* measure is. The measures also produce similar substantive effects as the original. A 1 SD in the LAP measure is associated with a 0.52 change in the Freedom House Index. The average change associated with a 1 SD change in the continuous judicial independence measures is 0.56. Also, moving the La Porta measure from its minimum to its

14. Our primary interest here is in replicating LAP’s models, and so we adopt their specification. Of course, via a variety of mechanisms, other scholars argue that vigorous democracy breeds judicial independence (e.g., Ginsburg 2003). To consider this possibility, we estimated 2SLS models of the Freedom House Index, using AJR’s settler mortality scores. However, in no model could we reject the assumption of exogeneity.

15. We replicate the models summarized in Table 3, Panel B of LAP. These models include controls for the natural log of GDP per capita, average latitude, and ethnolinguistic fractionalization. The LAP replication files only contain data for the dependent variable and controls for a limited set of states. Obtaining measures of economic development, latitude, and the Freedom House scores were straightforward. We were unable to locate the average ethnolinguistic fractionalization measure LAP used for a more general sample of states. Instead, we use the measure developed by Fearon (2003), which is one of the measures LAP use in their average. Although not ideal, we are encouraged by the fact that the correlation between this measure and LAP’s is 0.87. More importantly, LAP find no influence for average ethnolinguistic fractionalization and that measure is only very weakly correlated with all of the judicial independence scores for the set of states on which LAP conducted their study. Finally, we have replicated LAP’s findings in which they exclude these controls, and the results are stronger than what we report in Table 3.

Table 3. LA Porta Replication

	Freedom House Political Rights Index		
	β (RSE)	p	N
Original study			
LAP	1.64 (0.84)	0.06	68
<i>De facto</i> measures			
CIM	1.91 (1.14)	0.10	158
CIRI	2.23 (0.29)	<0.01	163
HEN	2.31 (0.30)	<0.01	154
H&C	2.17 (0.26)	<0.01	163
F&V-F	0.77 (0.75)	0.31	83
GCR	0.19 (0.10)	0.08	54
PRS	0.04 (0.19)	0.82	119
T&K	1.83 (0.30)	<0.01	165
XCONST	0.75 (0.05)	<0.01	144
<i>De jure</i> measures			
F&V-J	2.14 (0.92)	0.02	99
KEITH	0.12 (0.04)	<0.01	157

Replications OLS models reported in Table 3, Panel B of La Porta et al. (2004). The dependent variable is the Freedom House Political Rights scale in 1996, reversed so that higher numbers reflect higher levels of rights. We take the score of each judicial independence measure in 1996 or the closest available year when 1996 is missing. Control variables, the results for which are not reported, include a standardized score of average latitude, a measure of ethnolinguistic fractionalization, and the natural log of GDP/cap. To match LAP, our measure of development is from 1998 though the results are not sensitive to this choice.

maximum value induces a change of 1.64 units on the Freedom House index. The average change in that index associated with changing the binary measures of judicial independence from 0 to 1 is 2.2.

3.4 Summary

The results suggest some optimism for using the available measures of judicial independence, though the results are stronger for the *de facto* indicators. Content, convergent, and construct validation analyses show that the menu of available measures offers a number of reasonable choices for a variety of research questions. Of course, we should let our theoretical models, within which our concepts operate, guide our measurement choices. Here, we echo Collier and Adcock (1999: 539) who argued that “specific methodological choices are often best understood and justified in light of the theoretical framework, analytic goals, and context of research involved in any particular study.” More often than not, however, there is not a single best choice of measurement. In this case, given subtle conceptual and measurement differences across the available measures, it is advisable to perform robustness analysis using different measures. As we now discuss, however, this approach is not free from challenges.

4. Missing Data, Convergence, and Economic Development

Table 1 underscores a well-understood feature of judicial independence data—measures are missing for much of the latter half of the 20th century. Only the HEN, XCONST, PRS, and CIM measures offer anything close to a reasonable time series. For this reason, we are largely left with time-invariant measures. In so far as theories of judicial independence commonly predict significant temporal change or at least change across political institutional arrangements that vary with time (e.g., the number of veto players), cross-sectional measures are not ideal. This much, of course, is understood.

A more interesting, less well-publicized issue concerns missingness patterns in these data within blocks of years for which measures are available. For example, the Bertelsmann transformation index (BTI) project excludes all OECD states, so that it almost certainly produces a judicial independence measure that is missing systematically by economic development. If this pattern is general, and if scholars do not address it (and we have no evidence that they do), then it may be that all estimates of judicial independence effects are subject to the biases associated with various haphazard strategies of imputation or data deletion (King et al. 2001). The question we address now concerns how bad the missing data problem are in these data.

There are many possible systematic explanations for missingness. We pursued the simplest mechanism that we could imagine, considering whether missingness could be explained by a state's level of economic development, largely because it is more difficult to locate appropriate information about poorer states. We evaluated this possibility in two related ways. First, we created a binary missing data indicator for each variable, and then estimated a logistic regression of that indicator on the natural log of GDP per capita. We also conducted difference of means tests for the development measure, across the subsample for which we have data, and the subsample for which we have missing values.

Table 4 summarizes the results. For each test, the measures in the left column reflect those for which missingness seems to be related to development. The measures on the top of the table are more likely to be missing for less developed states, whereas those on the bottom are more likely to be missing for more developed states. More than half of the measures have missingness patterns that appear dependent on development, and of these measures, they are most likely to be missing for developing states. None of the measures developed from the State Department reports (T&K, CIRI, and H&C) fit either pattern—they are either broadly available or for whatever reason not correlated with development.

On its own, this finding is not all that striking. Missingness is a feature of most data empirical scholars use, and we all have a responsibility to deal with it when it emerges and seems consequential. But it is important to stress that scholars in our field *are not dealing with it*. More importantly, Figure 3 suggests that the problem has implications for the robustness

Table 4. Missingness Across Levels of Development

	Regression		Difference of means	
	Yes	No	Yes	No
Missing for low development?	F&V-F	CIM	CIM	CIRI
	F&V-J	CIRI	F&V-F	H&C
	GCR	H&C	F&V-J	HEN
	LAP	HEN	GCR	KEITH
	PRS	KEITH T&K*	LAP PRS	T&K*
Missing for high development?	XCONST		XCONST	
	BTI		HEN	
			BTI	

Table summarizes a simple missingness analysis. For each indicator, we estimated a logistic regression of a missing value indicator on the natural log of GDP/cap. Measures for which the development coefficient was statistically significant at the 0.05 level (two-tailed test) are placed on the left side of the table. Those for which the coefficient was negative are placed at the top of the table, identifying measures that are less likely to have missing values as development increases. The table also summarizes difference of means tests for the natural log of GDP/cap across state-years where data are missing and where data are present. *Tate and Keith's measure has very wide coverage; however, the team did not code the United States. Removing the United States from the analysis eliminates the pattern of missingness suggested in the table.

analysis suggested by the convergent validity tests above. The figure shows a dot plot of the correlations between the T&K judicial independence measure and the remaining indicators, above (black diamonds) and below (red circles) the sample's median GDP/capita. In every case, correlations are stronger above median development than they are below. The correlation between the Feld and Voigt *de facto* score is actually negative below median GDP per capita! Providing this information for all measures is cumbersome, but the results generalize. Developing countries may not only have lower quality data but also may experience different paths toward developing the rule of law, which can cause greater divergence in expert evaluations (see Weingast 2009).

The clearest implication of the results in Figure 3 is precisely that where scholars want to know most about the causes and consequences of judicial independence among developing states; our measures seem to provide the least consistent information. But there is another concern that is related to earlier findings. Recall that the natural implication of having easy access to distinct measures of judicial independence, which are generally but not perfectly correlated with each other, is to evaluate the extent to which results are robust to alternative indicators. Haggard et al. (2008) strongly recommend such a strategy in light of the lack of a benchmark measure of the concept. While we agree with this suggestion, if we do this without dealing with the missing data problem, we risk being overly confident about our robustness analysis.

Consider two facts. Judicial independence data are missing systematically by level of development for many measures and measures are more

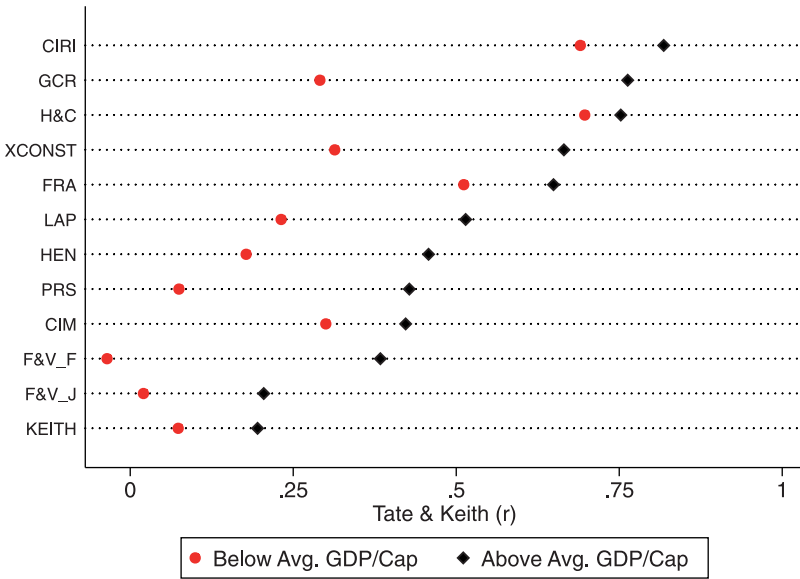


Figure 3. *Convergence Above and Below Median Development.* Dot plots of correlation coefficients measures of judicial independence above median GDP per capita (black diamonds) and below median GDP per capita (red circles).

likely to agree among more developed states. If we ignore the missing data problem and we are using measures that are more likely to be missing for underdeveloped states, we will conduct our robustness checks largely on observations from more developed states, as these are the states for which we have the most data. Yet, these are also the states where our measures are more likely to agree.

To see how this might happen, consider Table 5. The left column of the table shows the results from the LAP replication described above, where missing data are addressed via listwise deletion, i.e., it is not addressed. The right column shows multiple imputation estimates of the same models of political rights.¹⁶ The results at the top of the table are associated with measures that we found to be more likely to be missing for less developed

16. In our analysis, missing values are assumed to be missing at random. We assume further that the variables are characterized by a joint multivariate normal distribution. All imputations were conducted in Stata 11 using the data augmentation algorithm described in Tanner and Wong (1987). Making use of the fact that our measures are correlated positively with each other, the imputation model includes each judicial independence measure used in the LAP replication, in addition to the control variables in the LAP specification. Since we have found that the AJR settler mortality score is negatively correlated with the judicial independence measure, we include it, as well. We create 10 imputations for each missing value. Following the imputation stage and accepted practice, we round each imputed value for our binary variables to 0 or 1. Specifically, we enter a 0 for each imputed score where the imputation is less than 0.5 and 1 if the imputation is at least 0.5 (Schafer 1997).

Table 5. LAP Missing Data Analysis

	Listwise deletion			Multiple imputation		
	β (SE)	p -value	N	β (SE)	p -value	N
F&V-F	0.77 (0.75)	0.31	83	1.58 (0.99)	0.14	207
F&V-J	2.14 (0.92)	0.02	99	3.53 (1.84)	0.09	207
GCR	0.19 (0.10)	0.08	54	0.35 (0.09)	<0.01	207
LAP	1.64 (0.84)	0.06	68	0.73 (0.62)	0.25	207
PRS	0.04 (0.19)	0.82	119	-0.05 (0.21)	0.78	207
CIM	1.91 (1.14)	0.10	158	3.34 (1.30)	0.01	207
CIRI	1.90 (0.38)	<0.01	147	2.63 (0.29)	<0.01	207
HEN	2.31 (0.30)	<0.01	154	2.54 (0.33)	<0.01	207
H&C	2.17 (0.26)	<0.01	163	2.57 (0.29)	<0.01	207
KEITH	0.12 (0.04)	<0.01	157	0.16 (0.04)	<0.01	207
T&K	1.83 (0.30)	<0.01	165	2.27 (0.29)	<0.01	207
XCONST	0.75 (0.05)	<0.01	144	0.80 (0.04)	<0.01	207

Results at top of the table are for measures where there is consistent evidence suggesting that scores are more likely to be missing for underdeveloped states; results at the bottom are for measures where this is not true. The left of the table shows listwise deletion estimates; the right shows multiple imputation estimates; imputation model described in footnote 13.

states across both tests summarized in Table 4.¹⁷ Scanning the bottom half of the table, we can see that the multiple imputation results among the measures that are not more likely to be missing for less developed states are actually quite similar to the listwise deletion estimates. If anything they are stronger no more precisely estimated. This is not true for the results summarized in the upper half of the table where we observe significant changes in magnitude.

Suppose that LAP had evaluated the robustness of their findings to the measures listed in the upper left half in Table 5, but done so via listwise deletion. The authors would have concluded that all of the measures are positively associated with political rights, as measured by the Freedom House Index, and that three of the measures of judicial independence, including their own, were significant at the 0.10 level. Yet, the multiple imputation estimates suggest a different conclusion. Most striking, the LAP estimate itself drops by >50% and it is very far from statistically significant. Although the size of the Feld and Voigt *de jure* estimate increases, so does its standard error. Further, the sign of the PRS estimate is now negative. Although evaluating the robustness of the findings with

17. In principle, if scholars use measures that are more likely to be missing for developed states, it may be that a naïve robustness analysis would suggest less consistent results than one that addresses the missing data problem, precisely because models would have been run on the states for which measures are least likely to agree. Unfortunately, we only have one measure, XCONST, which appears to be missing for wealthier states on both of our tests, making it impossible to consider that possibility here.

respect to alternatives is somewhat subjective, we would suggest that the upper right portion of the table suggests a much more uncertain picture than the upper left.

5. Conclusions

The increasing number of scholarly attempts to define and measure judicial independence speaks to how crucial this variable has become in several theories of economic development, rights protection, and the rule of law. We provided a conceptual map of judicial independence that can serve as a guide not only to the users of existing measures but also to potential creators of new measures. Our map distinguishes *de jure* judicial independence, that is, institutions aimed at insulating judges, and *de facto* independent judicial behavior, and within the latter concept between the notion of judges authoring their own opinions (autonomy) and the extent to which judicial decisions are actually complied with (influence). We collected and conceptually categorized 13 of the most widely used available measures of judicial independence in order to systematically evaluate their content, convergent, and construct validity. Our findings reveal that, despite differences in content, *de facto* measures tend to correlate reasonably well. That said, we also found nonrandom patterns of missing data that undermine traditional robustness checks, as well as no systematic efforts to deal with potential measurement error resulting from attempts to measure the behavior of strategic actors. Our analysis also shows that *de jure* and *de facto* measures of judicial independence do not covary and, more surprising, that *de jure* measures are not correlated among them.

The implications for measuring *de facto* judicial independence would appear relatively clear. In light of its latency, scholars ought to be creative in their efforts to identify the ways in which independent judicial behavior is likely to manifest. Some manifestations may not directly reflect features of the judiciary at all. Instead, they may reflect features of state and society that we would expect to observe in the presence of judicial independence. What this means is that *de facto* judicial independence indicators may lack face validity in certain cases. It may only be in the context of a theoretical argument that the logic of the measurement strategy becomes clear. Further, since there is evidence that extant indicators are providing information about the same concept, not only would it be appropriate to consider the robustness of our results across distinct indicators, but also we should certainly consider combining the indicators in ways designed to uncover better the latent concept. A measurement model could be particularly helpful in this context.

The *de jure* concept presents a related but distinct research challenge. There the concern is not so much that we wish to find a way of cutting through noise to uncover good estimates of the latent concept. Rather, it is not yet clear that we have identified well the rules (or sets of rules) that produce the incentives we hope to measure. The perennial question on

whether and how institutions impact behavior, that is, the relation between *de jure* and *de facto* judicial independence, requires thinking carefully about two sets of issues: the conditions under which the institutions tend to work effectively and the incentives set by specific institutions, such as the appointment, removal, or constitutional review powers of judges. The length of judicial tenure as established in the constitution is a good measure if one wants to study the relationship between the *de jure* and the actual length of tenure. But it is far less clear whether life tenure in the constitution produces “independent judicial behavior,” even if the actual tenure is also long. The latter question requires a conceptual clarification of what amounts to independent judicial behavior; for instance, what we have identified as autonomy or influence and a theoretical model of how a long tenure incentivizes such behavior. Since it is unclear that we yet understand well the properties of particular institutions, our findings suggest that we ought to be cautious in aggregating *de jure* information into additive indexes. This suggests that whereas we may be in a position to aggregate different efforts to measure *de facto* independence, disaggregation is the more prudent course in the case of *de jure* independence.

Appendix A

A.1. PRS “Law and Order” measure

The Law and Order measure is a part of the political risk rating that includes 12 components and 15 subcomponents covering both political and social attributes. The PRS staff collects political information and financial and economic data, converting these into risk points for each individual risk component. The political risk assessments are made on the basis of subjective analysis of the available information. In the “law and order” component, Law and Order are assessed separately; each component can take values from 0 to 3 to a total of six joint points together. The Law subcomponent is an assessment of the strength and impartiality of the legal system, whereas the Order subcomponent is an assessment of popular observance of the law. Thus, a country can enjoy a high rating—3—in terms of its judicial system, but a low rating—1—if it suffers from a very high crime rate or if the law is routinely ignored without effective sanction (e.g., widespread illegal strikes). See: http://www.prsgroup.com/ICRG_Methodology.aspx

A.2. Wittold Henisz’s (2000) measure of “judicial independence”

The existence of an independent judiciary ($J=1$ or $J=0$) is determined through the joint existence of a POLITY score on executive constraints of at least 3 (see definition below) and, where data are available, a PRS score on law and order of at least 4 (see definition above). HEN’s measure on judicial independence is available for the period after 1960. See: <http://www-management.wharton.upenn.edu/henisz/>

Table 6. Correlation Matrix for all Measures

	H&C	CIRI	T&K	F&V-F	PRS	CIM	GCR	BTI	XCONST	HEN	F&V-J	KEITH	LAP
H&C	1.00												
CIRI	0.69 ^a	1.00											
T&K	0.72 ^a	0.75 ^a	1.00										
F&V-F ^b	0.34	0.41	0.36	1.00									
PRS	0.45	0.38	0.44	0.40	1.00								
CIM	0.50	0.49	0.49	0.21	0.58	1.00							
GCR	0.62	0.66	0.71	0.54	—	0.41	1.00						
BTI ^c	0.67	0.67	0.63	0.35	0.21	0.44	0.46	1.00					
XCONST	0.69	0.56	0.59	0.33	0.35	0.49	0.36	0.71	1.00				
HEN	0.46	0.37	0.38	0.25	—	0.35	0.38	0.27	—	1.00			
			De jure measures										
F&V-J ^b	0.40	0.32	0.28	0.20	-0.03	0.20	0.05	0.32	0.24	-0.03	1.00		
KEITH	0.18	0.26	0.14	0.02	0.09	0.26	-0.13	0.37	0.49	0.26	0.29	1.00	
LAP ^b	0.32	0.31	0.34	-0.04	0.21	0.16	0.37	0.44	0.25	0.12	0.23	0.21	1.00

Correlation coefficients for judicial independence measures (all available data). For correlations between measures that have no overlapping years, we used the closest available years. We do not include correlations between HEN and PRS or XCONST since HEN is constructed from these two indicators.

^aKendall's Tau b.

^b2003 value.

^c2006 value.

A.3. Polity measure on “Constraints on the Executive”

As noted in the POLITY II Codebook (Gurr 1990): “Operationally, this variable refers to the extent of institutionalized constraints on the decision-making powers of chief executives, whether individuals or collectivities. Such limitations may be imposed by any “accountability groups.” In Western democracies, these are usually legislatures. Other kinds of accountability groups are the ruling party in a one-party state; councils of nobles or powerful; advisors in monarchies; the military in coup-prone polities; and in many states a strong, independent judiciary. The concern is therefore with the checks and balances between the various parts of the decision-making process. A seven-category scale is used (e.g., on evidence used to code countries, see Gurr 1990). See: <http://garnet.acns.fsu.edu/whmoore/polity/polity.html>

A.4. GCR’s measure of “judicial independence”

The World Economic Forum, an independent international organization, produces a series of annual reports among which are the GCR. This measure of judicial independence is from the GCR (data generated by a survey of Firm Executives operating in a given country) question: Is the judiciary in your country independent from political influences of members of government, citizens, or firms? No heavily $d (= 1)$ or Yes entirely independent $(= 7)$. The question’s wording has varied slightly over the years. We take the values for this measure from the Fraser Institute’s annual *Economic Freedom of the World Report* that uses 42 distinct pieces of data to measure economic freedom in 141 nations. According to the Fraser Institute, judicial independence is one of the key ingredients of a legal system consistent with economic freedom. The data sources of the Index of Economic Freedom include the GCR from which they take their measure of judicial independence. See: <http://www.weforum.org/en/initiatives/gcp/index.htm> and <http://www.freetheworld.com/>

A.5. La Porta et al. (2004) measure “Judicial Checks and Balances”

La Porta and his coauthors measure judicial independence and constitutional review. “Judicial independence is of obvious value for securing property and political rights when the government is itself a litigant, as in the takings of property by the state. But judicial independence is also socially valuable in purely private disputes when one of the litigants is politically connected, and the executive wants the court to favor its ally.” (2003: 3) “Besides seeking to influence judges, the executive and the legislature would also wish to pursue policies and pass laws that benefit themselves, or democratic majorities, or allied interest groups. Constitutional review is intended to limit these powers. By checking laws against a rigid constitution, a court can limit such self-serving efforts.” (2003: 4)

Their measure of “judicial checks and balances” comes from national constitutions, including all 71 countries covered in the Maddex (1995) Encyclopedia of Constitutions, with the exception of transition economies because “their constitutions are rapidly changing.” Judicial independence

is proxied by looking at the tenure of judges in the highest ordinary court, the tenure of judges in the highest administrative court, and to whether courts have “law making” powers and judicial decisions that are constrained by prior judicial decisions. Constitutional review is measured by the degree of rigidity of the constitution and the extent of judicial review powers.

A.6. Feld and Voigt’s (2003) measure of *De Jure* and *De Facto* Judicial Independence

Feld and Voigt measure judicial independence *de jure* and judicial independence *de facto*. They focus on the independence of the court with authority to interpret the constitution.

The authors sent a questionnaire to experts (judges, law professors, lawyers and activists of the NGOs) asking several quite detailed questions regarding the different components of both of their concepts. The components of judicial independence *de jure* are (a) whether the highest court is anchored in the constitution; (b) how difficult is it to amend the constitution; (c) appointment procedures of the judges; (d) their length of tenure; (e) whether there is a fixed retirement age of the judges in the court; (f) removal procedures; (g) whether re-election of judges is possible; (h) protection and adequacy of the salary of the judges; (i) accessibility to the highest court; (j) procedure for allocation of cases in the court; (k) judicial review powers; and (l) transparency of the court.

The components of judicial independence *de facto* are: (a) average length of tenure; (b) deviation of average length of tenure from *de jure* prescription; (c) number of judges removed from the office; (d) frequency of changes in the number of judges in the court; (e) real salary of the judges; (f) real court’s budget; (g) number of constitutional changes in the relevant articles; and (h) compliance by other branches on court rulings.

A.7. Cingranelli and Richards (2008) measure of judicial independence

CIRIs develop their measure of judicial independence from the US State Department yearly country reports. Their concept of judicial independence requires that judges be free from control by the government or military. In particular, CIRIs ask whether judges can be removed from the office, whether there is judicial review, whether judges are free from corruption, and whether case outcomes are not influenced by the government.

The measure is ordinal and contains three categories. A state received a score of 2, fully independent judiciary if its judiciary satisfies the following criteria:

- it has the right to rule on the constitutionality of legislative acts and executive decrees;
- judges at the highest level of the courts have a minimum of seven-year tenure;

- the President or Minister of Justice cannot directly appoint or remove judges and the removal of judges is restricted (e.g., allowed for criminal misconduct);
- actions of the executive and legislative branch can be challenged in the courts;
- all court hearings are public; and
- judgeships are held by professionals.

A state receives a 1 if there are structural limitations on judicial independence (e.g., the ability of the chief executive or minister of justice to appoint and dismiss judges at will, even if they do not actually do so in the particular year being coded or there is limited corruption or intimidation of the judiciary). A state receives a 0 if there are “active and widespread constraints on the judiciary” (e.g., active government interference in cases or judicial dismissals for political reasons. See: http://ciri.binghamton.edu/documentation/ciri_coding_guide.pdf)

A.8. Tate and Keith (2009) measure of judicial independence

Tate and Keith develop their measure of judicial independence from the US State Department yearly country reports. The measure is ordinal and falls into three categories. Tate and Keith (2009) provide a comprehensive conceptual review of judicial independence, but do not state an explicit concept of their own. Regarding their concept, Tate and Keith (2009: 4) write: while the dimensions of judicial independence conceptualized by each of these scholars and others do not fit together perfectly, we do see a common core across them that allows us to identify two somewhat overlapping sets of distinctions. The first is the distinction between (a) institutional (or collective) independence from the other branches or private and public actors and (b) the independence of the individual judge from the same influences.

States receive a 2 if “the judiciary is reported as *generally independent* or is independent in practice with no mention of corruption or outside influence,” 1 if “the judiciary is reported to be somewhat independent in practice with reports of (some) pressure from the executive *at times* or with occasional reports of corruption,” and 0 if “the judiciary is reported as not being independent in practice; is reported to have significant or high levels of executive influence or interference; or is reported to high levels of corruption (17).”

A.9. Howard and Carey (2004) measure of judicial independence

Howard and Carey (2004: 286) define judicial independence as, “The extent to which a court may adjudicate free from institutional controls, incentives, and impediments imposed or intimidated by force, money, or extralegal, corrupt methods by individuals or institutions outside the judiciary, whether within or outside government.” Their measure, which is derived from US State Department country reports, is a three-category

ordinal measure, which indicates whether the high court of a state in a particular year is fully independent, partially independent, or dependent. According to Howard and Carey (2004: 287–88), a state has a fully independent judiciary if the high court functions in practice:

- independently of the executive and legislature, and is
- relatively free from corruption and bribery and provides basic criminal due process protections to criminal defendants.

A state has a partially independent judiciary if its high court either satisfies the first or the second condition or partially satisfies both. A state has a dependent judiciary if its high court satisfies neither condition.

A.10. Apodaca–Keith Scale (2004) of judicial independence

Keith (2002) produces seven ordinal measures based on the United Nations principles of judicial independence. The scores indicate whether a state’s constitution formally guarantees the tenure of high court judges; ensures the finality of judicial decisions; grants judges exclusive authority over their jurisdiction; bans special or military jurisdiction for civilian crimes; financial autonomy; a separation of powers system; and specifically enumerates appointment qualifications. For each component, the Keith measure returns a value of 2 if a constitution explicitly provides a feature of *de jure* independence; 1 if it does so partially or vaguely; and, 0 otherwise. In the case of military jurisdiction, Keith also includes a code of –1 in the event that a constitution explicitly endorses military jurisdiction. Apodaca (2004) creates an additive index of these seven items whose range is from –1 to 14.

A.11. Clague et al. CIM (1999)

The CIM measure created by Clague et al. is “the ratio of non-currency money to the total money supply” (Clague et al. 1999: 188). Conceptually, high values of CIM reflect a society’s trust in judicial institutions that enforce the banking industry’s contractual obligations. The CIM was conceptualized as a measure of legal protections for property rights. In order to use it as a general measure of *de facto* independence, one must assume that, on average, states that possess judicial institutions that protect property rights are likely to have judicial institutions that protect rights generally.

A.12. BTI—Judicial Independence (2008)

The BTI is derived from country expert reports. The aggregate index includes a component for judicial independence. Bertelsmann (2008: 20) write:

An independent judiciary refers first and foremost to how far the courts can interpret and review norms and pursue their own reasoning free from the influence of rulers or powerful groups and individuals. This requires a

differentiated organization of the legal system, including legal education, jurisprudence, regulated appointment of the judiciary, rational proceedings, professionalism, and channels of appeal and court administration.

The BTI judicial independence measure is a scale from 1 (low) to 10 (high). Experts consider whether the judiciary is free both from unconstitutional intervention by other institutions and from corruption. There are mechanisms for judicial review of legislative or executive acts. The judiciary is established as a distinct profession and operates relatively independently, but its functions are partially restricted by factors, such as corruption and insufficient territorial or functional penetration. The judiciary is institutionally differentiated, but its decisions and doctrine are subordinated to political authorities or severely restricted by functional deficits, such as territorial penetration, resources, or severe corruption. The judiciary is not institutionally differentiated or is significantly subordinated to religious or political authorities. See: <http://www.bertelsmann-transformation-index.de>

References

- Acemoglu, Daron, Simon Johnson, and James A. Robinson 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation," 91 *American Economic Review* 1369–401.
- Adcock, Robert, and D. David Collier 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research," 95 *American Political Science Review* 529–46.
- Apodaca, Clair. 2004. "The Rule of Law and Human Rights," 87 *Judicature* 292–9.
- Barro, Robert J. 1997. *Determinants of Economic Growth: A Cross-Country Empirical Study* Cambridge, MA: MIT Press.
- Becker, Theodore. 1970. *Comparative Judicial Politics*. University Press of America.
- Burbank, Steven B., and Barry Friedman. 2002. "Reconsidering Judicial Independence," in Steven B. Burbank, and Barry Friedman, eds., *Judicial Independence at the Crossroads: An Interdisciplinary Approach*. 9–44. New York, NY: Sage Publications Inc.
- Cameron, Charles M. 2002. "Judicial Independence: How Can You Tell It When You See it? And, Who Cares?," in Steven B. Burbank, and Barry Friedman, eds., *Judicial Independence at the Crossroads: An Interdisciplinary Approach*. 134–47. New York, NY: Sage Publications Inc.
- Carothers, Thomas. 2006. *Promoting the Rule of Law Abroad: in Search of Knowledge*. Carnegie Endowment for Intl Peace.
- Carrubba, Clifford J. 2005. "Courts and Compliance in International Regulatory Regimes," 67 *Journal of Politics* 669–89.
- Cingranelli, David L., and David L. Richards 2008. "The Cingranelli-Richards (CIRI) Human Rights Dataset," <http://www.humanrightsdata.org> (accessed August 21, 2012).
- Clague, Christopher, Philip Keefer, Stephen Knack, and Mancur Olson 1999. "Contract-Intensive Money: Contract Enforcement, Property Rights, and Economic Performance," 4 *Journal of Economic Growth* 185–211.
- Collier, David, and David Adcock 1999. "Democracy and Dichotomies. A Pragmatic Approach to Choices about Concepts," 2 *Annual Review of Political Science* 537–65.
- Cross, Frank B. 1999. "The Relevance of Law in Human Rights Protection," 19 *International Review of Law and Economics* 87–98.
- Epp, Charles R. 1998. *The Rights Revolution: Lawyers, Activists, and Supreme Courts in Comparative Perspective* Chicago, IL: University of Chicago Press.

- Fearon, James D. 2003. "Ethnic and Cultural Diversity by Country," 8 *Journal of Economic Growth* 195–222.
- Feld, Lars P., and Stefan Voigt 2003. "Economic Growth and Judicial Independence: Cross-country Evidence Using a New Set of Indicators," 19 *European Journal of Political Economy* 497–527.
- Ferejohn, John, and Charles Shipan 1990. "Congressional Influence on Bureaucracy," 6 *Journal of Law, Economics, & Organization* 1–20.
- Frye, Timothy. 2004. "Credible Commitment and Property Rights: Evidence from Russia," 98 *American Political Science Review* 453–66.
- Ginsburg, Tom. 2003. *Judicial Review in New Democracies: Constitutional Courts in Asian Cases* New York, NY: Cambridge University Press.
- Gurr, Ted R. 1990. *Polity II: Political Structures and Regime Change, 1800–1989* Boulder, CO: Center for Comparative Politics.
- Haggard, Stephan, Andrew MacIntyre, and Lydia Tiede 2008. "The Rule of Law and Economic Development," 11 *Annual Review of Political Science* 205–34.
- Hamilton, Alexander, James Madison, and John Jay. 2009. *The Federalist Papers* New Haven, CT: Yale University Press.
- Hammergren, Lynn. 2006. *Envisioning Reform*. Penn State University Press.
- Hathaway, Oona A. 2007. "Why Do Countries Commit to Human Rights Treaties?," 51 *Journal of Conflict Resolution* 588–621.
- Helmke, Gretchen. 2005. *Courts Under Constraints* Cambridge, MA: Cambridge University Press.
- Helmke, Gretchen, and Jeffrey K. Staton. 2011. "The Puzzling Judicial Politics of Latin America. A Theory of Litigation, Judicial Decisions, and Interbranch Conflict," in Gretchen Helmke, and Julio Rios-Figueroa, eds., *Courts in Latin America* New York, NY: Cambridge University Press.
- Henisz, Witold J. 2000. "The Institutional Environment for Economic Growth," 12 *Economics and Politics* 1–31.
- Herron, Eric S., and Kirk A. Randazzo 2003. "The Relationship between Independence and Judicial Review in Post-Communist Courts," 65 *The Journal of Politics* 422–38.
- Howard, Robert M., and Henry F. Carey 2004. "Is an Independent Judiciary Necessary for Democracy?," 87 *Judicature* 284–90.
- Kastellec, Jonathan P., and Jeffrey R. Lax 2008. "Case Selection and the Study of Judicial Politics," 5 *Journal of Empirical Legal Studies* 407–46.
- Keele, Denise, Robert W. Malsheimer, Donald W. Floyd, and Lianjun Zhang 2009. "An Analysis of Ideological Effects in Published Versus Unpublished Judicial Opinions," 6 *Journal of Empirical Legal Studies* 206–39.
- Keith, Linda Camp. 2002a. "Constitutional Provisions for Individual Human Rights: Are They More Than Mere Window Dressing," 55 *Political Research Quarterly* 111–43.
- . 2002b. "Judicial Independence and Human Rights Protection Around the World," 85 *Judicature* 195–200.
- King, Gary, James Honaker, Anee Joseph, and Kenneth Scheve 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation," 95 *American Political Science Review* 49–69.
- Kornhauser, Lewis A. 2002. "Is Judicial Independence a Useful Concept?," in Steven B. Burbank, and Barry Friedman, eds., *Judicial Independence at the Crossroads: An Interdisciplinary Approach*. 45–55. New York, NY: Sage Publications Inc.
- La Porta, Rafael, Florencio López de Silanes, Cristian Pop-Eleches, and Andrei Shleifer 2004. "Judicial Checks and Balances," 112 *Journal of Political Economy* 445–70.
- Martin, Andrew D., Clifford J. Carrubba, Matthew Gabel, Gretchen Helmke, and Jeffrey K. Staton 2006. "Collaborative Research: A Cross-National Study of Judicial Institutionalization and Influence,". National Science Foundation; Law and Social Science Program.

- North, Douglass, and Barry Weingast 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in 17th Century England," 49 *Journal of Economic History* 803–32.
- North, Douglass, William Summerhill, and Barry Weingast. 2000. "Order, Disorder, and Economic Change: Latin America vs. North America," in Bruce Bueno de Mesquita, ed., *Governing for Prosperity* New Haven, CT: Yale University Press.
- Powell, Emilia J., and Jeffrey K. Staton 2009. "Domestic Judicial Institutions and Human Rights Treaty Violation," 53 *International Studies Quarterly* 149–74.
- Pozas-Loyo, Andrea, and Julio Rios-Figueroa. 2006. "When and Why 'Law' and 'Reality' Coincide? De Jure and De Facto Judicial Independence in Chile and Mexico," in Alejandra Rios-Cazares, and David Shirk, eds., *Evaluating Accountability and Transparency in Mexico*. University of San Diego/Trans-Border Institute.
- Raz, Joseph. 1997. "The Rule of Law and its Virtue," 93 *The Law Quarterly Review* 195.
- Rios-Figueroa, Julio. 2011. "Institutions for Constitutional Justice in Latin America," in Gretchen Helmke, and Julio Rios-Figueroa, eds., *Courts in Latin America* New York, NY: Cambridge University Press.
- Rosenberg, Gerald N. 1991. *The Hollow Hope: Can Courts Bring about Social Change?* Chicago, IL: University of Chicago.
- Russell, Peter H., and O'Brien, David M. *Judicial Independence in the Age of Democracy: Critical Perspectives from Around the World*. University of Virginia Press.
- Schafer, J.L. 1997. *Analysis of Incomplete Multivariate Data*. Chapman & Hall/CRC.
- Simmons, Beth A. 2009. *Mobilizing for Human Rights: International Law in Domestic Politics* New York, NY: Cambridge University Press.
- Stasavage, David. 2002. "Private Investment and Political Institutions," 14 *Economics and Politics* 41–63.
- Staton, Jeffrey K. 2010. *Judicial Power and Strategic Communication in Mexico* New York, NY: Cambridge University Press.
- Tanner, Martin A., and Wing Hung Wong 1987. "The Calculation of Posterior Distributions by Data Augmentation (With Discussion)," 82 *Journal of the American Statistical Association* 528–50.
- Tate, C. Neal, and Linda Camp Keith 2009. "Conceptualizing and Operationalizing Judicial Independence Globally".
- Tate, C. Neal, Stacia L. Haynie, Reginald Sheehan, and Donald Songer 2006. "Collaborative Research: Extending a Multi-Country Database and Program of Research," National Science Foundation: Law and Social Science Program.
- Treier, Shawn, and Simon Jackman 2008. "Democracy as a Latent Variable," 52 *American Journal of Political Science* 201–17.
- Vanberg, Georg. 2005. *The Politics of Constitutional Review in Germany* New York, NY: Cambridge University Press.
- Weingast, Barry. 1997. "The Political Foundations of Democracy and the Rule of Law," 91 *American Political Science Review* 245–63.
- . 2009. "Why Developing Countries Prove So Resistant to the Rule of Law," in Robert L. Nelson, ed., *Global Perspectives on the Rule of Law*. Routledge.