Available online at www.sciencedirect.com
ScienceDirect
Journal of Comparative Economics 35 (2007) 711–728
Journal of COMPARATIVE ECONOMICS
www.elsevier.com/locate/jce

Measuring legal systems

Howard Rosenthal [a], Erik Voeten [b,*]

[a] *Wilf Family Department of Politics, New York University, 19 West 4th Street, #305, New York, NY 10012-1119, USA*
[b] *Edmund A. Walsh School of Foreign Service and Government Department, Georgetown University, 520B InterCultural Center, 37th & O Streets, NW Washington, DC 20057, USA*

Received 4 January 2007; revised 7 August 2007

Available online 15 August 2007

Rosenthal, Howard, and Voeten, Erik—Measuring legal systems

Factor-analytic models can substantially improve the measurement of comparative legal systems and thereby our understanding of how legal systems influence economic outcomes. These methods yield better estimates of latent constructs, allow us to evaluate whether institutional features are representative of a theoretical construct and whether allegedly distinct theoretical constructs can be separated empirically. We illustrate these points through a re-analysis of a 2003 study by Djankov, La Porta, Lopez-de-Silanes and Shleifer, using a factor-analytic method that combines continuous and categorical indicators. Our results strengthen these authors' findings with respect to how legal formalism relates to legal origin and the quality of the legal system. Yet, the results also show that many of the original index items are not significantly positively related to formalism. The results thus shed light on what institutional features should be prioritized for reform if we seek to make legal systems less formalistic. Moreover, we question the evidence that the formalism model better predicts the quality of the legal system than does the alternative "incentives" model. We argue, instead, that formalism and incentives may both relate to the tendency of a legal system to use bureaucratic rule-making. Our approach can readily be applied to the analysis of legal concepts other than formalism. *Journal of Comparative Economics* **35** (4) (2007) 711–728. Wilf Family Department of Politics, New York University, 19 West 4th Street, #305, New York, NY 10012-1119, USA; Edmund A. Walsh School of Foreign Service and Government Department, Georgetown University, 520B InterCultural Center, 37th & O Streets, NW Washington, DC 20057, USA.

JEL classification: K10; K40; K41; K42; O10

* Corresponding author.
E-mail addresses: hr31@nyu.edu (H. Rosenthal), ev42@georgetown.edu (E. Voeten).

0147-5967/$ – see front matter
doi:10.1016/j.jce.2007.08.001

Keywords: Law and economics; Legal origin; Measurement; Development; Courts; Factor analysis; Item-response; Bayesian

1. Introduction

"Law and Finance," a seminal article by La Porta et al. (1998) advanced the legal origin theory. The theory claimed that the cross-national variation in corporate governance regulation can, to a large extent, be traced back to the historical origin of the legal system of each country and that the type of origin exerts a powerful influence on economic outcomes. That paper focused on investor protection and the breadth of the equity market. Subsequently, the authors extended their comparative analyses to other economic institutions, including regulations for starting a business (Djankov et al., 2002), protections of corporate shareholders and creditors (La Porta et al., 1998), securities laws (La Porta et al., 2006), labor regulations (Botero et al., 2004), and dispute resolution mechanisms (Djankov et al., 2003).

These articles have been extraordinarily influential, yielding not only large numbers of academic citations[1] but also exerting considerable policy influence. The World Bank financed much of the research and relies heavily on the index measures for institutions developed in the legal origin articles.[2] The legal origin theory has, nonetheless, been subject to extensive debate and criticism (Acemoglu and Johnson, 2005; Gourevitch and Shinn, 2005; Musacchio, 2006; Rajan and Zingales, 2003; Pistor and Xu, 2005). This paper contributes to this larger debate by focusing on the issue of measuring variation in legal institutions across countries. Despite the centrality of measurement to the study of the economic effects of legal institutions, measurement issues have received relatively little attention in the literature. Generally, scholars collect indicators of characteristics of a country's regulatory or legal system and then combine these into an aggregate index that supposedly measures the theoretical concept of interest. Subsequently, these indices are related to upstream variables, such as legal origin, and downstream variables, such as performance of the legal system or economic outcomes. We argue that substantively and theoretically important information can be uncovered by scaling the data with models similar to factor analysis instead of using indices that are constructed ad hoc. The models yield better estimates of latent constructs, allow us to evaluate whether individual institutional features are indeed representative of a theoretical construct and whether allegedly distinct theoretical constructs can be distinguished empirically.

We illustrate these points with a re-analysis of the results in the article "Courts" by Djankov et al. (2003). This paper argues both that countries with different legal origins display varying levels of "legal formalism" in their dispute resolution mechanisms and that high levels of formalism have negative economic consequences. We apply a measurement model to the Djankov et al. (2003) data that explicitly models a mixture of categorical and continuous indicators (Quinn, 2004). Our findings show that the latent variable estimate performs better than the ad-hoc Djankov et al. (2003) index of formalism in the sense that it correlates more strongly with legal origin and the performance variables. On the other hand, we show that several of the indicators included in the formalism index do not fit the concept of formalism and thus, that there

[1] For example, the "Law and Finance" article already has 534 citations to the published version in the *Social Science Citation Index* (accessed May 3, 2006) and is largely responsible for making its four authors the top four cited authors in economics and business between 1995–2005 (*Science Watch* November/December 2005).

[2] Especially in its *Doing Business* project, see: http://www.doingbusiness.org.

is no evidence that these aspects of a countries should be reformed if we wished to make its system less formalistic (as the World Bank recommends). In this sense, our results give guidance to policy-makers as to which aspects of a legal system's reform should be prioritized in an effort to make it less formalistic. In addition, we conclude that Djankov et al. (2003) have inappropriately tested formalism against an alternative theory of "incentives." We show that the items used to measure "incentives" do not hang together as the authors expected, with a ban on contingency fees having an effect opposite to the authors' expectations. Indeed, incentives and formalism may both express a more basic differentiation of countries with respect to the extent of rule-making. This plausibly reflects the trade-offs nations face when opting for centralized or decentralized contract enforcement (see Djankov et al., 2003). More generally, we conclude that scaling can contribute to both more refined policy conclusions and the further refinement of theories about the economic effects of institutions.

2. Measurement of legal systems: the "Courts" project

In "Courts," Djankov et al. (2003) observe that while economic theory strongly implies that levels of investment and trade depend on institutions that efficiently resolve property and contractual disputes, economic theory is much less conclusive about the qualities that make some dispute resolution mechanisms more efficient than others. Moreover, there is little empirical information about the characteristics and efficiency of actual courts, especially in comparative perspective.

Djankov et al. (2003) propose three theories that may explain variation in the efficiency of courts. First, the "development" theory suggests that efficient court systems result largely from demands that arise when a society develops economically and when its populace becomes more educated (Demsetz, 1967; North, 1981). Second, the "incentive" theory stipulates that court systems work more efficiently when judges, lawyers, and litigants have strong incentives to avoid delay in the resolution of disputes, for instance because there are mandatory deadlines (Busciaglia and Diakiolas, 1999; Messick, 1999). Third, and more novel, the "procedural formalism" theory suggests that court systems are more efficient as they more closely resemble an ideal-typical model in which disputes can be resolved on fairness grounds without heavy reliance on written law, professional judges and lawyers, and without restrictions on the presentation of arguments and evidence (Shapiro, 1981). Djankov et al. (2003) argue that levels of procedural formalism are primarily determined by historical circumstances, with French civil law nations having especially high levels of formalism in contrast to the low levels in British common law nations.

For the "Courts" data, Djankov et al. (2003) cooperated with the international association of law firms *Lex Mundi* to produce 38 indicators that detail the procedures that litigants face in two fairly straightforward matters of dispute resolution: the eviction of a tenant for non-payment of rent (*evict*) and the collection of a bounced check (*check*). *Evict* and *check* can be viewed as basic settings that may reveal how nations perform in a broader set of dispute resolution situations. The indicators were collected for 109 countries. Thirty-one of these indicators are assumed to measure some aspect of "procedural formalism," while the remaining 7 indicators are used to capture the rival "incentive" theory. Table 1 gives an overview of the indicators.[3] Djankov et al. (2003) provide much more detailed information about the indicators and the motivation for their inclusion.

[3] All data is available from: http://www.andrei-shleifer.com/data.html.

Table 1
Overview of the indicators in the courts data set

	Indicators of Procedural Formalism
jud_court	Equals one if a court of general jurisdiction would be chosen or assigned to hear the case under normal circumstances, zero if a court of limited jurisdiction
jud_profe	Equals one if the judge, or the members of the court or tribunal, could be considered as professional
jud_repre	Equals one if the law requires the intervention of a licensed attorney
jud_index	***Professionals vs. Laymen:*** Index from above three items
wrt_filin	Equals one if the complaint is normally submitted in written form to the court, and zero if it can be presented orally
wrt_servi	Equals one if the defendant's first official notice of the complaint is most likely received in writing, and zero otherwise
wrt_oppos	Equals one if under normal circumstances the defendant's answer to the complaint should be submitted in writing, and zero if it may be presented orally to court
wrt_evide	Equals one if evidence is mostly submitted to the court in written form
wrt_finar	Equals one if final arguments on the case are normally submitted in writing
wrt_judge	Equals one if the judge issues the final decision in the case in written form, and zero if orally
wrt_notif	Equals one if normally the parties receive their first notice of the final decision in written form
wrt_enfmt	Equals one if the enforcement procedure is mostly carried out through the written court orders or written acts by the enforcement authority
wrt_index	***Written vs. Oral:*** Index from previous 8 items
leg_compl	Equals one if the complaint is required, by law or court regulation, to include references to the applicable laws, legal reasoning, or formalities that would normally require legal training
leg_legal	Equals one if the judgment must expressly state the legal justification (articles of the law or case-law) for the decision
leg_julaw	Equals one when judgment must be on law only, and zero when judgment may be based on equity grounds
leg_index	***Legal Justification:*** Index from previous 3 items
evi_evi	Equals one if, by law, the judge cannot freely request or take evidence that has not been requested, offered, or introduced by the parties
evi_rej	Equals one if, by law, the judge cannot refuse to collect or admit evidence requested by the parties, even if she deems it irrelevant to the case
evi_ofc	Equals one if statements of fact that were not directly known or perceived by the witness, but only heard from a third person, may not be admitted as evidence
evi_preq	Equals one if, by law, the judge must pre-qualify the questions before they are asked of the witnesses
evi_int	Equals one if parties and witnesses can only be orally interrogated by the judge
evi_orig	Equals one if only original documents and "authentic" or "certified" copies are admissible documentary evidence
evi_auth	Equals one if the authenticity and probative value of documentary evidence is specifically defined by the law
evi_rec	Equals one if, by law, there must be a written or magnetic record of all evidence introduced at trial,
evi_index	***Statutory Regulation of Evidence:*** Index from previous 8 items
rev_enf	Equals one if the enforcement of judgment is automatically suspended until resolution of the appeal when a request for appeal is granted. Equals zero if the suspension of the enforcement of judgment is not automatic, or if the judgment cannot be appealed at all
rev_capp	Equals one if issues of both law and fact (evidence) can be reviewed by the appellate court
rev_iapp	Equals one if interlocutory appeals are allowed
rev_index	***Control of Superior Review:*** Index from previous 3 items
Sta_concl	Equals one if the law requires plaintiff to attempt a pre-trial conciliation or mediation before filing the lawsuit
sta_sepro	Equals one if the law requires the complaint to be served to the defendant through the intervention of a judicial officer
sta_notju	Equals one if the law requires the judgment to be notified to the defendant through the intervention of a judicial officer
sta_index	Engagement Formalities: Index from previous 3 items

Table 1 (*continued*)

stp_sfise	The total minimum number of independent procedural actions required to complete filing, admission, attachment, and service
stp_strju	The total minimum number of independent procedural actions required to complete opposition to the complaint, hearing or trial, evidence, final arguments, and judgment
stp_sexju	The total minimum number of independent procedural actions required to complete notification and enforcement of judgment
stp_indexn	***Independent Procedural Actions:*** normalized index from previous 3 items
	Indicators of Incentives Theory
mand_judge	Equals one if the judge is required by law to admit or reject the lawsuit within a certain period of time
mand_evid	Equals one if the period in which the parties may collect or present evidence is fixed by law to a certain number of days after service or number of days before hearing
mand_def	Equals one if the defendant is required by law to file the opposition within certain time limit, either in terms of number of days from service or number of days before the hearing.
mand_jugmnt	Equals one if the judge is required by law to enter judgment within a specified period of time after the conclusion of the hearing or the final pleadings
mand_notif	Equals one if the court is required by law to notify the parties within a specified period of time after judgment is entered
mand_index	***Mandatory Time Limits:*** Index from previous 5 items
quota_proh1	Equals one if quota litis or contingent fee agreements are prohibited by law in all cases
loser_pay	Equals one if the loser is required to pay all the costs of the dispute

Djankov et al. (2003) proceed in two steps to create estimates of the levels of "procedural formalism" and "incentives" from the 2×31 procedural formalism indicators in Table 1. First, as outlined in Table 1, they create seven sub-indices by computing the normalized sum of related individual indicators. So, for example, the *Professionals vs. Laymen* index uses three binary indicators: whether a court of general or limited jurisdiction is generally used, whether judges can be considered professional, and whether the law requires the participation of a licensed attorney. The sum of these indicators is divided by 3 to obtain an index in the 0–1 range. For the index that relies on continuous variables (*Independent Procedural Actions*), the normalization is achieved based on the minimum and maximum number of procedural actions in the sample. The seven sub-indices are then summed into a single indicator for procedural formalism.

A similar process is repeated to create an index for the presence of mandatory time limits. This index, together with binary indicators for the existence of "loser pays" rules and the prohibition of "quota litis" or contingency fee agreements, are taken to be the indicators for the incentives theory. There is, unlike formalism, no overall index of incentives.

3. Method

Djankov et al. (2003) argue that "The exact method of the construction of the formalism index is not crucial, since the various sub-indices generally point in the same direction" (p. 476). Yet, the methodology for measuring formalism relies on several important but untested assumptions. First, it assumes that each indicator is equally informative about the concept that underlies the sub-index and that each sub-index is equally informative about the umbrella concept of procedural formalism. Second, and related, it assumes that these indicators measure a single dimensional concept. In addition, it is implied that the indicators for the incentive theory comprise a distinct dimension of variation across legal systems. Third, even though individual indicators of

complex concepts are generally measured with error, the procedure ignores measurement error in the observations.

These assumptions could be defended if we had particularly strong a priori conceptions about precisely what indicators should be included in scales of procedural formalism and judicial incentives. Yet, as Djankov et al. (2003) acknowledge, this is a very novel area of research. Neither "procedural formalism" nor "incentives" were previously operationalized and measured across countries. There is no strong theory that specifies precisely what aspects of a legal system contribute to its procedural formalism or its ability to create performance incentives. As such, one should be cautious in picking indicators and it would seem prudent to assess the validity and consequences of the assumptions underlying the measurement approach of Djankov et al. (2003).

Factor analysis models the relationship between a large set of manifest indicators and a small number of latent constructs. Let $i = 1, \ldots, N$ denote countries and $j = 1, \ldots, J$ denote individual characteristics of the legal system. Each observation x_{ij} is the value for country i on aspect j of its legal system. If, as Djankov et al. (2003) argue, legal formalism were the defining characteristic underlying these individual qualities, then we have J imperfect indicators of a single unobserved variable. More generally, the J indicators arise probabilistically from a set of K unobserved (or latent) fundamental characteristics of legal systems, where K is much smaller than J.

A complication in the analysis of the "Courts" data, and many similar datasets of comparative institutions, is that it contains both dichotomous and continuous indicators. If we had only continuous data, we could use standard factor analysis. If we had only dichotomous indicators, we could use item-response theory (IRT) to model the relationship between the observed indicators and our unobserved concept of interest. IRT models are not widely used in economics but they have a long history in the educational testing literature and, more recently, in the political science literature where they form the basis for analyses of committee voting (e.g. Jackman, 2000; Londregan, 1999; Martin and Quinn, 2002; Poole, 2005; Poole and Rosenthal, 1991, 1997, 2007; Voeten, 2000).

The normal factor model and the IRT model with probit link are special cases of a more general model, which can be labeled mixed factor analysis (MFA). (For a more formal exposition, see Quinn, 2004). In normal theory factor analysis, the $N \times J$ matrix of observed indicators $\mathbf{X}$ is assumed to follow a multivariate normal distribution. This assumption is obviously untenable with a mixture of continuous and dichotomous data. Instead, the discrete variables are expressed in terms of a latent continuous variable (the random utility probit model). We first map the observed matrix of indicators $\mathbf{X}$ to an $N \times J$ matrix of latent variables $\mathbf{X}^*$, depending on their level of measurement[4]:

$$x_{ij} = \begin{cases} x_{ij}^* & \text{if variable } j \text{ is continuous,} \\ 1 & \text{if } x_{ij}^* > 0, 0 \text{ otherwise, if variable } j \text{ is dichotomous.} \end{cases} \tag{1}$$

We now assume that $\mathbf{X}^*$ follows a multivariate normal distribution. We then use the usual factor analytic model to describe the J vector of individual country scores $\mathbf{x}_i^*$ as a function of a $J \times (K+1)$ matrix of factor loadings $\boldsymbol{\Lambda}$, and a $K+1$ vector of factor scores $\boldsymbol{\varphi}_i$, with the first element set to 1 for all i:

$$\mathbf{x}_i^* = \boldsymbol{\Lambda}\boldsymbol{\varphi}_i + \boldsymbol{\varepsilon}_i, \quad \text{with } \boldsymbol{\varepsilon}_i \sim N(0, \boldsymbol{\Psi}). \tag{2}$$

[4] Quinn (2004) develops MFA more generally, including ordered multichotomies as well as dichotomies. The "Courts" data contains only continuous and dichotomous variable. We consequently use simplified notation in Eq. (1).

We assume that $\boldsymbol{\Psi}$ is diagonal.[5] The diagonal entries of $\boldsymbol{\Psi}$ that represent dichotomous variables are assumed to equal 1. This serves to identify the model (similar to assuming that $\boldsymbol{\varepsilon}_i \sim N(0, 1)$ in regular probit models). For convenience of interpretation, we standardize all continuous variables to have mean zero and standard deviation 1. Consequently, elements of the first column of $\boldsymbol{\Lambda}$ for the continuous variables are constrained to 0. The elements of the remaining columns can be interpreted as factor loadings. For dichotomous indicators, the elements of the first column of $\boldsymbol{\Lambda}$ can be interpreted as difficulty parameters from IRT models. The remaining elements are discrimination parameters, which are analogous to factor loadings.

We follow Quinn in assuming independent standard normal prior distributions for factor scores $\boldsymbol{\Phi}$ and independent diffuse normal priors for the parameters in $\boldsymbol{\Lambda}$.[6] The diagonal elements of $\boldsymbol{\Psi}$ that correspond to continuous items are given independent inverse gamma priors.

We estimate the model using Bayesian methods as implemented by the *R* package *MCMCpack* (Martin and Quinn, 2006). For our data, where only continuous and dichotomous indicators are present, *MCMCpack* relies on a Gibbs sampling algorithm to fit the model. We identify the polarity of the results by constraining one item to have a positive loading on each dimension (this has no substantive consequences). After discarding the initial 100,000 iterations as burn-in, we run 1,000,000 iterations and store every 100th scan in our posterior sample. We estimate one-dimensional models for

(1) the 31 formalism indicators for the check case,
(2) the 31 indicators used for the evict case, and
(3) all 62 indicators for procedural formalism. In additional, we estimated a separate model for all incentives indicators.

We also estimated a two-dimensional model that includes all indicators that Djankov et al. (2003) used to represent the incentives and the formalism theories.[7] We identified the two-dimensional model by restricting an item (whether the law requires the intervention of a licensed attorney) to have zero loading on the second dimension for both the evict and the check cases. We chose this item because of its centrality to the formalism theory (and thus should load primarily on the first

[5] It is theoretically possible to relax this assumption and model the correlation between the measurement error of related items (e.g. the items for the different subscales). There is, however, very little work on the properties of such models, especially how much data one would need to make estimates of the diagonal elements of $\boldsymbol{\Psi}$ reasonable.

[6] These become truncated normal distributions if we wish to constrain an indicator parameter. The inverse of the variance on the diffuse normal priors is 0.001.

[7] The MCMCPack code for the two-dimensional model is: post2<-MCMCmixfactanal(~JUDXCOUR + JUDXPROF + JUDXREPR + WRTXOPPO + WRTXEVID + WRTXFINA + WRTXJUDG + WRTXNOTI + LEGXCOMP + LEGXLEGA + LEGXJULA + EVIXEVI + EVIXREJ + EVIXOFC + EVIXPREQ + EVIXINT + EVIXORIG + EVIXAUTH + EVIXREC + REVXENF + REVXCAPP + REVXIAPP + STAXCONC + STAXSEPR + STAXNOTJ + MANDXJUD + MANDXEVI + MANDXDEF + MANDXJUG + MANDXNOT + QUOTAXPR + LOSERXPA + JUDXCOXA + JUDXPRXA + JUDXREXA + WRTXFIXA + WRTXOPXA + WRTXEVXA + WRTXFIXB + WRTXJUXA + WRTXNOXA + LEGXCOXA + LEGXLEXA + LEGXJUXA + EVIXEVXA + EVIXREXA + EVIXOFXA + EVIXPRXA + EVIXINXA + EVIXORXA + EVIXAUXA + EVIXREXB + REVXENXA + REVXCAXA + REVXIAXA + STAXCOXA + STAXSEXA + STAXNOXA + MANDXJXA + MANDXEIV + MANDXDXA + MANDXJXB + MANDXNXA + QUOTAXXA + LOSERXXA + ZSTP_ SFI + ZSTP_STR + ZSTP_SEX + ZSTP_S_A + ZSTP_S_B + ZSTP_S_ C, factors = 2, Data = Courtsdata, lambda.constraints = list(WRTNOXA = list(2, "+"), JUDXREPR = list(3, 0), JUDXREXA = list(3, 0), REVXENF = list(3, "+")), burnin = 100000, mcmc = 1000000, thin = 100, verbose = 10000, L0 = 0.25, store.lambda = TRUE, store.scores = TRUE, tune = 3.0.

dimension). Indeed, variation on this item was nearly perfectly explained in a one-dimensional model.[8]

The two-dimensional model is central to our attempt to evaluate the measurement assumptions underlying the "Courts" paper. We now turn to that task.

4. Interpretation of items

An important advantage of scaling over a priori aggregate indicators is that the analyst can study not just the scores for respondents (countries) but also the items that go into the construction of indices. For example, item-response models are used in the educational testing literature to examine how successful individual questions on the SAT, GRE, or other tests are in discriminating between high-quality and poor-quality students. Similarly, we can use the discrimination parameters and factor loadings from an MFA analysis to examine how successfully the various indicators that enter into the "procedural formalism" index discriminate between countries with high levels and low levels of formalism. In effect, Djankov et al. (2003) assume that discrimination parameters are 1 for each item in the formalism index. Moreover, we can evaluate whether the "incentives" indicators truly measure a distinct concept.

Figure 1 presents the 95 percent credible intervals[9] for the discrimination parameters (for the dichotomous items) and the factor loadings (for the continuous items) for a two-dimensional MFA analysis that includes both incentives and formalism indicators. The labels correspond to the names of the items in Table 1. There is only one label for each pair of items from the check and evict cases. The item discrimination parameters of the same items on different cases track each other closely, thus indicating that we do not have distinct *check* and *evict* dimensions but rather that the scores reflect more fundamental characteristics of legal systems.[10] Given that the incentive indicators comprise only about one-fifth of the total number of indicators, we would expect the resulting scale to reflect procedural formalism with the incentives indicators, perhaps, constituting a second dimension. The figure includes only 71 of a possible 76 discrimination parameters as 5 items lacked variation in the sample so item parameters could not be estimated.[11]

First dimension MFA estimates indeed correlate strongly with the Djankov et al. (2003) index for procedural formalism on both the check (Pearson $R = 0.90$) and evict ($R = 0.86$) cases. The second dimension country estimates are not significantly correlated with the formalism scores. Yet, it is also clear that the second dimension does not constitute an incentive dimension: none of the incentives indicators loads significantly on the second dimension (i.e. all credible intervals include zero).

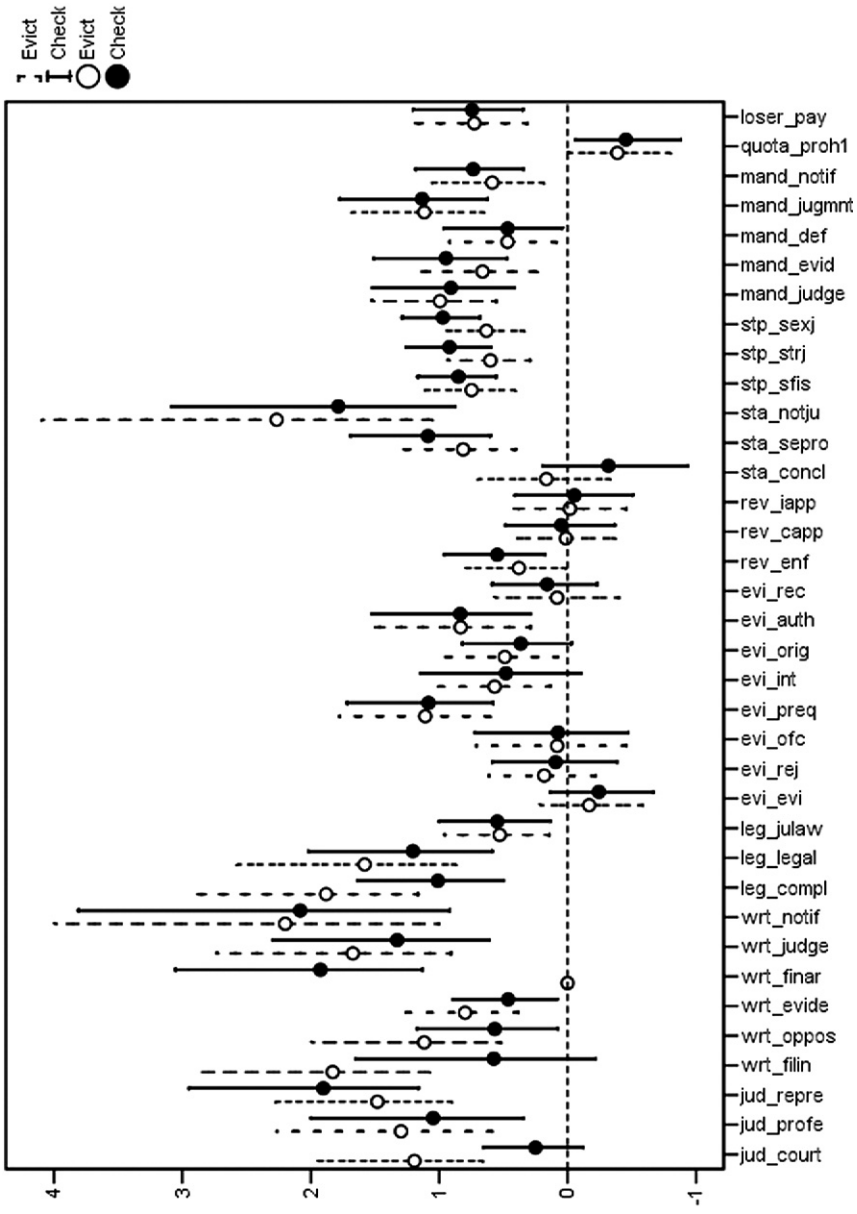
There is considerable variation in the item parameters from the formalism scales on the first dimension. For seven indicators, the point estimates are below zero, implying that the indicators are related to formalism in a manner opposite to that expected by Djankov et al. (2003). The 95

[8] One of the weaknesses of MFA is the absence of a good standard for fit and thus, for evaluating how many dimensions are appropriate. Since most indicators in our data are binary, one way to evaluate the fit is classification errors. Ranking countries along a single dimension classifies 84.1% of all values on binary indicators correctly. Adding a second dimension brings this percentage up to 88.2% and a third dimensional solution yields 90.9% correctly classified observations. The excellent classification on the first dimension and the weak additions by the second and third dimension echoes our finding that the second dimension relates poorly to legal origin and evaluations of performance of the legal system.

[9] The credible interval is the Bayesian equivalent of the confidence interval.

[10] Alternatively, check and evict cases may be highly similar because, in each country the same law firm was asked to evaluate both cases.

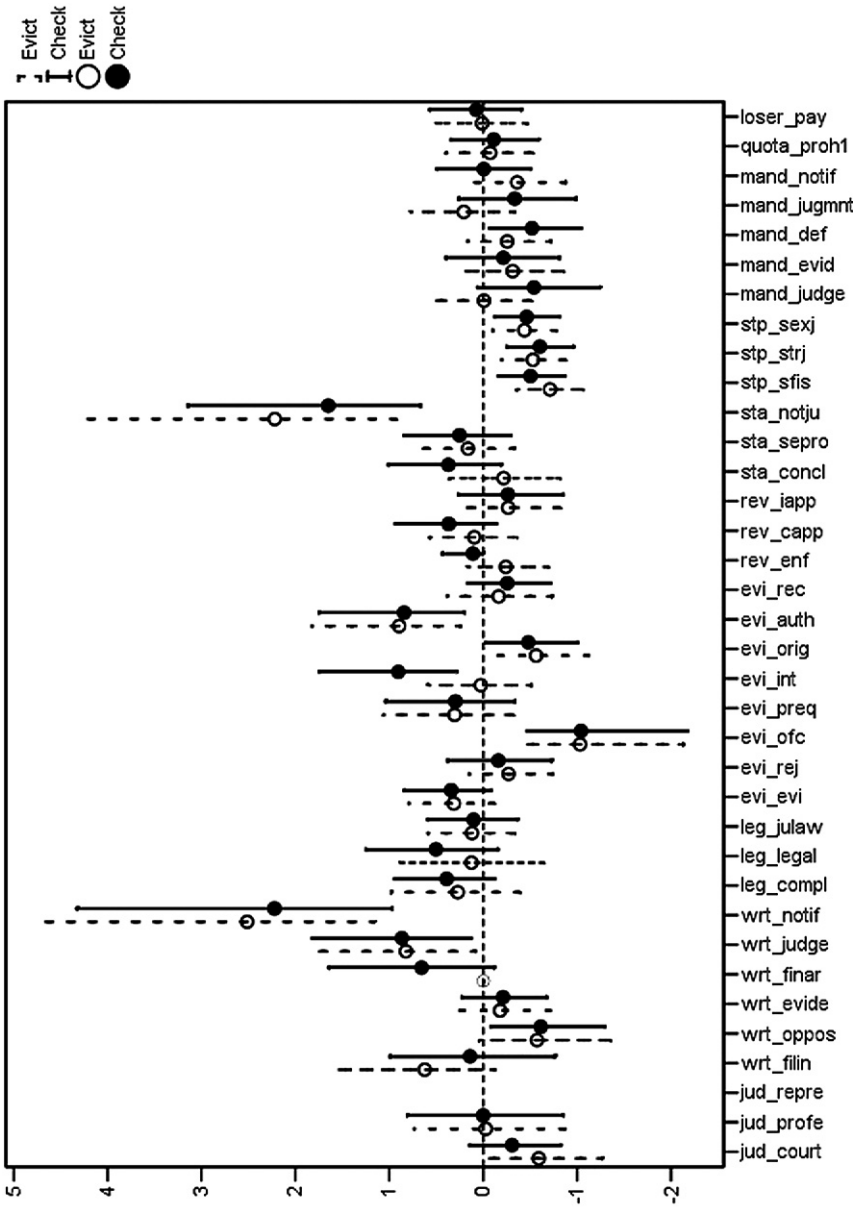
[11] These are wrt_servi and wrt_enfmt for both cases as well as wrt_finar for the evict case.



(a) First dimension MFA solution.

Note to Fig. 1(a). Two points correspond to each label of the horizontal axis. The first point is for "evict," the second for "check." The items at the right end of the axis are "incentives" items, beginning with "mand_judge." The item wrt_finar has no variation in the sample for the evict case and was thus omitted from the analysis.

Fig. 1. Discrimination Parameters and Factor Loadings Indicators (95 percent credible intervals. Names and ordering as in Table 1).



(b) Second dimension MFA solution.

Note to Fig. 1b. The discrimination parameters for the item jud_repre were fixed at 0 for identification purposes.

Fig. 1. (*continued*)

percent credible interval for 18 of the 57 formalism item discrimination parameters includes 0, indicating that variation on these indicators is not significantly associated with the latent construct. Of the indicators that were insignificant for both the check and evict cases, five were from the *evidence* sub-index[12] and two from the three-item *review* index.[13] Substantively, this implies that countries that regulate evidence and review in the manner measured by these indicators are neither more nor less "formal" in the overall regulation of contract enforcement than countries that lack these rules. Consequentially, the analysis provides no reason to advocate reform of these aspects of court systems.

All indicators for incentives are significantly associated with variation among countries on the first dimension. This suggests that the presence of mandatory time limits in the law reflects a taste for formal rules rather than a system designed to provide incentives for quick resolution to participants. A plausible argument is that the enforcement environment shapes the choice of legal institutions to a considerable degree (Glaeser and Shleifer, 2002). Mandatory time limits are superfluous in countries where contracts can effectively be enforced at a decentralized level. Instead, countries with professional lawyers and judges with poor performance incentives could plausibly benefit from mandatory time limits. The prohibition of "quota litis" or contingency fees loads *negatively* on the latent construct (for both the check and the evict cases). This result should not be surprising. Contingency fees provide an incentive for litigation, not just of spurious lawsuits but also of suits where the plaintiff has a valid claim. "Quota litis" may improve the efficiency of a legal system. The result does, however, again indicate that a taste for rules (namely one that prohibits particular fees) determines the first dimension. Our findings lead us to express some doubt as to whether Djankov et al. (2003) accurately capture incentive theory with the indicators that they include. The presence of formal rules to create positive incentives is likely at least partially a reflection of an overall poor incentive structure.

We suspect that the second dimension is mostly related to levels of economic development. Countries with high scores stand out in particular because they require written notifications of judgments (wrt_notif positive sign), require a judicial officer to notify the defendant of a judgment (sta_notju), require the judge to reach a written judgment, and have specific definitions of documentary evidence in the law (evi_auth). These are all requirements that presume high bureaucratic capacity. Countries with low scores, on the other hand, also create hurdles for litigants, but these are mostly ones that do not require much capacity. These countries do not allow defendants to present answers orally in courts (wrt_oppos has negative signs), have large numbers of independent steps (stp_ vars have negative signs), and prohibit hearsay as evidence (evi_ofc). Indeed we find that second dimension MFA country-scores are significantly correlated with the natural log of per capita GDP (Pearson $R = 0.39$), whereas the first dimension scores are unrelated to levels of development.

5. Formalism, legal origin, and economic performance

An additional advantage of the MFA analysis is that it produces better estimates of our concept of interest. Figure 2 presents the 90 percent credible intervals of the posterior distribution for the levels of procedural formalism as estimated by the mixed factor analysis. The estimates are based on a combination of the check and evict cases. The countries with the highest levels of formalism

[12] "Out-of-court statements are inadmissible," "Judge can not reject irrelevant evidence," "Judge can not introduce evidence," "Mandatory recording of evidence," and "Mandatory pre-trial conciliation."

[13] "Comprehensive review in appeal" and "Interlocutory appeals are allowed."

Fig. 2. Estimates of procedural formalism (90 percent credible intervals, high scores indicate high levels of formalism).

are primarily lesser developed Latin American countries. The countries with the lowest levels of procedural formalism are lesser developed former English dependencies. The zero point on the scale indicates roughly average levels of formalism.[14]

The length of the credible intervals indicates uncertainty about the point estimates of levels of formalism. The sizes of the credible intervals are reasonably similar across countries. There

[14] The mean of the prior distribution was set at zero.

are three main reasons why the size of the credible intervals may differ. First, some countries may have more missing data than others, although this is not a serious problem in the current case. Second, point estimates tend to be less precise near the ends of a scale, as there is little information to separate countries with extremely high (or low) levels of formalism from each other.[15] Third, in some countries the indicators line up in less predictable ways than in others. For example, Greece and Bulgaria have near-identical point-estimates for their levels of procedural formalism, yet the standard deviation of the posterior is higher for Greece (0.24 vs. 0.20). This suggests that Greece's legal system exhibits some characteristics that make it extremely formalistic but also some indicators that point in the opposite direction, whereas Bulgaria's system exhibits a moderately high level of formalism across the board. Overall though, there are relatively few examples along these lines in this dataset.

One could (and probably should) take measurement error into account when performing secondary analyses on levels of formalism (see Lewis and Linzer, 2005). Yet, in what follows we focus on a comparison with Djankov et al. (2003) and thus rely only on our point estimates, taken as the means of the posterior distributions. Overall, the bivariate correlations between the MFA dimension and the Djankov et al. index are high: The Pearson R is 0.92 with the combined Djankov et al. index of the check and evict cases. We also ran MFA on the check and evict cases separately. Here too, the bivariate correlations are high: 0.94 for the check case and 0.92 for the evict case.

5.1. Legal origin and procedural formalism

Table 2 reports a one-way analysis of variance of the relationship between the various point-estimates of procedural formalism and legal origin. The table reveals a striking pattern: legal origin accounts much better for variation in the MFA formalism estimates than in the Djankov et al. index. Legal origin explains 57 percent of the variation in the MFA scaled measure of formalism for the check case and 51 percent for the evict case but only 41 percent for the Djankov et al. index for either case. When we combine the two cases, legal origin accounts for 45 percent of the variation in the Djankov et al. measure and 58 percent of the variance in the MFA estimate. Thus, it appears that using MFA strengthens Djankov et al.'s (2003) conclusion that procedural formalism has its origins in historical legacies. The last three columns of the table look at the influence of legal origins on MFA estimates of the incentives indicators and all indicators. On the incentives scale, "good" incentives indicate positive scores.[16] The table shows that incentives and the second dimension of the combined incentive-formalism MFA are only moderately related to legal origins but that 57 percent of the variance in the first dimension estimate on all indicators is accounted for by legal origins.

5.2. Procedural formalism and the quality of legal systems

Djankov et al. (2003) also present results from regression analyses that show that high levels of procedural formalism correlate strongly and negatively with the performance of legal systems. Moreover, they show that formalism is more strongly associated with outcomes than indicators for the alternative "incentives" and "development" theories. As measures for the latter theory,

[15] The analog with legislative voting is that extremist legislators are generally difficult to pin down precisely.

[16] Note that the absence of a prohibition on contingent fee arrangement is a "good" incentive in the estimate.

Table 2
Legal origins and measures of formalism, MFA and Djankov et al. (2003) estimates

Legal origin		Items used								
		Formalism						Incentives*	All	
		Djankov et al. (2003) check	MFA check	Djankov et al. (2003) evict	MFA evict	Djankov et al. (2003) check + evict	MFA check + evict	MFA	MFA Dim 1	MFA Dim 2
British	Mean	2.76	−0.62	3.02	−0.62	2.89	−0.58	−0.43	−0.60	−0.25
$N = 42$	SD	0.83	0.58	0.77	0.54	0.69	0.43	0.71	0.44	0.49
Socialist	Mean	3.93	0.08	3.83	−0.09	3.88	0.00	0.44	0.09	−0.18
$N = 16$	SD	0.53	0.32	0.52	0.41	0.50	0.32	0.67	0.32	0.55
French	Mean	4.29	0.70	4.38	0.75	4.33	0.68	0.41	0.60	0.18
$N = 40$	SD	1.01	0.71	0.82	0.60	0.89	0.62	0.86	0.58	0.66
German	Mean	3.15	−0.19	3.57	−0.02	3.36	−0.10	0.07	−0.05	0.43
$N = 16$	SD	0.44	0.38	0.33	0.33	0.34	0.29	0.71	0.20	0.17
Scandinavian	Mean	3.15	−0.42	3.32	−0.51	3.24	−0.42	−0.29	−0.39	0.31
$N = 6$	SD	0.59	0.32	0.47	0.42	0.36	0.09	0.80	0.13	0.27
F statistic		18.2[a]	26.6[a]	18.2[a]	34.8[a]	21.1[a]	35.3[a]	7.6[a]	34.1[a]	5.0[a]
Eta squared		0.41	0.51	0.41	0.57	0.45	0.58	0.23	0.57	0.16

[a] $p < 0.01$.
* Djankov et al. (2003) provided only items for incentives and did not construct an overall index.

Djankov et al. (2003) include the natural log of GNP per capita, latitude, and average years of schooling. In addition, they include a measure of ethnic fractionalization to control for the finding that institutional performance generally declines with increased fractionalization (Alesina et al., 2003).

Panel A in Table 3 reports results of regression analyses on the log of the total estimated duration of the *check* and *evict* cases, as in Djankov et al.'s (2003) Table 8.[17] The independent variables include the Djankov et al. formalism and incentives indices for the *check* and *evict* cases and the formalism and incentive scales (produced by MFA). Both the MFA measures and the Djankov et al. (2003) indices give similar results: formalism rather than indicators of development or incentives account for the variation in the duration of the *check* and *evict* cases. Thus, this important finding of Djankov et al. (2003) appears robust to the choice of measurement model.

Djankov et al. (2003) also analyzed the effect of procedural formalism on a broader range of measures of the quality of legal systems. First, they use small firm assessments of the fairness, honesty, affordability, consistency, and confidence in legal systems (from the *World Business Environment Survey*). Second, they use indicators developed by political risk consultancies (based on business surveys) of the perceived enforceability of contracts, corruption, and overall integrity of the legal system (more detail in Djankov et al., 2003). We combined the *check* and *evict* cases to compute a single scale for procedural formalism (or estimate in the case of MFA). We included log GNP per capita, latitude, ethnic heterogeneity, and a constant in all regressions.

[17] The analyses are not an exact replication of Djankov et al. (2003) as they were performed on an updated dataset sent to us by Florencio Lopez-de-Silanes, which corrects for some minor mistakes in the original education variable.

Table 3
The quality of legal systems: comparing Djankov et al. (2003) and MFA
Panel A: Duration of the check and evict cases ($N = 93$)

	Eviction of a Tenant				Check Collection			
	Djankov et al. (2003)		MFA		Djankov et al. (2003)		MFA	
	β	S.E.	β	S.E.	β	S.E.	β	S.E.
(Constant)	4.198[a]	0.747	5.779[a]	0.627	3.271[a]	0.719	4.464	0.601
Development								
Log GNP per capita	−0.008	0.084	−0.039	0.085	0.085	0.081	0.056	0.080
Latitude	0.487	0.570	0.692	0.565	0.678	0.547	0.848	0.540
Ethnic heterogeneity	−0.421	4.198	−0.420	0.389	0.671	0.377	0.668	0.371
Schooling	−0.034	−0.008	−0.037	0.048	−0.041	0.047	−0.035	0.047
Formalism								
Formalism Djankov et al. (2003)	0.369[a]	0.487	–	–	0.301[a]	0.091	–	–
Formalism MFA	–	–	0.404[a]	0.113	–	–	0.384[a]	0.109
Incentives								
Mandatory limits	−0.302	−0.421	–	–	−0.164	0.354	–	–
Quota litis	0.271	−0.034	–	–	0.030	0.174	–	–
Loser pay	0.030	0.369	–	–	0.080	0.165	–	–
Incentives MFA	–	–	−0.174	0.108			−0.007	0.102
R^2_{adj}	0.117		0.136		0.121		0.115	
S.E. Estimate	0.765		0.732		0.738		0.766	

Panel B: Perceptions about quality of legal systems (The regressions also included log GNP, latitude, ethnic heterogeneity, and a constant)

Dependent variables:	Djankov et al. (2003) formalism		MFA formalism		MFA all indicators (DIM1)	
	B (S.E.)	R^2_{adj} (S.E. estimate)	B (S.E.)	R^2_{adj} (S.E. estimate)	B (S.E.)	R^2_{adj} (S.E. estimate)
Fair and impartial ($N = 64$)	−0.521[a]	0.344	−0.747[a]	0.414	−0.816[a]	0.459
	0.088	0.649	0.110	0.613	0.110	0.589
Honest and uncorrupt ($N = 64$)	−0.501[a]	0.363	−0.737[a]	0.440	−0.802[a]	0.474
	0.093	0.683	0.115	0.641	0.116	0.621
Legal system affordable ($N = 64$)	−0.148[b]	0.170	−0.246[a]	0.208	−0.245[a]	0.199
	0.069	0.507	0.089	0.495	0.093	0.498
Legal system consistent ($N = 64$)	−0.345[a]	0.230	−0.530[a]	0.313	−0.583[a]	0.348
	0.082	0.601	0.102	0.568	0.103	0.553
Confidence in legal system ($N = 64$)	−0.178[b]	0.199	−0.259[a]	0.217	−0.297[a]	0.237
	0.074	0.541	0.096	0.535	0.099	0.528
Enforceability of contracts ($N = 50$)	−0.553[a]	0.830	−0.725[a]	0.828	−0.793[a]	0.831
	0.108	0.664	0.143	0.667	0.153	0.661
Corruption ($N = 50$)	−0.549[a]	0.659	−0.902[a]	0.692	−0.903[a]	0.682
	0.151	1.307	0.188	1.243	0.202	1.262
Law and order ($N = 88$)	−0.309[c]	0.529	−0.447[c]	0.533	−0.581[c]	0.543
	0.181	1.570	0.236	1.564	0.247	1.546

[a] $p < 0.01$.
[b] $p < 0.05$.
[c] $p < 0.1$.

Our findings (Panel B) confirm that procedural formalism is a strong and negative correlate of almost all indicators of the quality of legal systems.[18] On 7 of the 8 quality measures, the MFA estimate of procedural formalism yields a better fit of the model than the Djankov et al. (2003) method. Thus, again, using MFA only seems to fortify the Djankov et al. (2003) conclusion that procedural formalism is a powerful determinant of the efficiency and quality of legal systems.

There is, however, some reason for pause. The results in the third column indicate that a model including an MFA scale computed from *all* indicators, so including those intended to measure the incentives theory, fits the data even better for most outcomes. Moreover, and most notably, if we include the MFA estimates for formalism and incentives together in a model to explain the level of law and order, it is incentives rather than formalism that significantly and *negatively* predict the integrity of the legal system.[19] Even though incentives are not significant correlates of on any of the other quality variables, this finding is important. The "Law and Order" variable (based on the Inter Country Risk Guide) is measured across many more countries ($N = 88$) than the other outcome indicators and is widely used in the literature. It may well be that rules that give incentives to lawyers are only necessary in countries that already experience problems with lawyer performance. Either way, these results warrant more attention to the Djankov et al. (2003) assumptions about how individual items fit in the scales and what the substantive interpretation of these scales is. (The second dimension MFA estimates had no significant correlation with any of the output variables in Table 3.)

6. Conclusion

We advance two general claims. First, the use of a statistical measurement model improves the validity and reliability of measurements of complex concepts. Measurement models allow us to assess the validity of assumptions that enter into the scale construction and to characterize measurement error. Our results show that MFA estimates also tend to have greater external validity than the indices constructed by Djankov et al. (2003): they are more strongly related to legal origins and they are better predictors of outcomes than are the Djankov et al. (2003) scales constructed from the same data. Thus, using a statistical measurement model strengthens some results of Djankov et al. (2003).[20]

Second, a major advantage of measurement models over ad-hoc index construction is that it offers researchers a vehicle for analyzing the various components of an index. This often generates substantively important knowledge. In the same way that researchers in educational testing evaluate whether proposed test-items successfully separate smart students from their less able counterparts, researchers of comparative institutions should examine the extent to which individual aspects of legal or political systems contribute to the more diffuse theoretical concepts that they are interested in. Such evaluations are particularly important when the theoretical motivation for including items into particular indices is novel or weak, as is often the case in the study of comparative institutions. Here, we have shown that a large number of items identified by

[18] We omit education from the set of covariates as it does not approach significance in any analysis and because including the variable results in the loss of several cases. Analyses including education yield virtually identical results.

[19] Coefficient for formalism: 0.022 (0.250), for incentives: −0.802 (0.206). R^2_{adj} is 0.600, S.E. of the estimate is 1.4468.

[20] We observe, on the other hand, that formalism and development indicators combine to explain only about one-half to two-thirds of the cross-national variation in the outcome measures. This suggests that more work needs to be done on a variety of matters that may affect performance, including whether disputes are involved through courts or administrative law procedures and whether political power is abused.

Djankov et al. (2003) do not contribute to the procedural formalism scale. Hence, the finding that formalism affects performance does not imply support for reforming those institutional features, something that is obscured by the ad-hoc index construction. We further find that all indicators that Djankov et al. collected to measure the rival "incentives" theory contribute substantially to variation on the formalism scale. One of these indicators, quota litis, has a significant effect that is opposite to the one hypothesized by Djankov et al. The information contained in the incentives indicators and the unanticipated effect of one of the indicators both call into question whether Djankov et al. (2003) have appropriately tested alternative theories of legal institutions.

Our analysis also has substantive implications for the study of comparative economics. We suggest that formal rules that appear to create performance incentives may in fact be a reflection of a country's poor enforcement environment rather than an indication of efficient institutions. For example, formal rules designed to speed up court proceedings may not be necessary in legal systems where enforcement can effectively be decentralized. To further investigate this suggestion and many other topics, it should be worthwhile to apply the techniques developed in this paper to the large set of cross-national variables that have been assembled by not only the "legal origins" scholars but also by many other researchers in the social sciences.

Acknowledgments

We thank Florencio Lopez-de-Silanes and Andrei Shleifer for generous assistance in using their data. Earlier versions of this article were presented in 2006 at Stanford University and at the 2005–2006 "Political Economy of Financial Markets" lunch meetings at the Russell Sage Foundation. We thank participants at both seminars, Simeon Djankov, Peter Gourevitch, Andrei Shleifer, and Holger Spamann for valuable comments.

References

Acemoglu, Daron, Johnson, Simon, 2005. Unbundling institutions. Journal of Political Economy 113 (5), 949–995.

Alesina, Alberto, Devleeschauwer, Arnaud, Easterly, William, Kurlat, Sergio, Wacziarg, Romain, 2003. Fractionalization. Journal of Economic Growth 8 (2), 155–194.

Botero, Juan, Djankov, Simeon, La Porta, Rafael, Lopez-de-Silanes, Florencio, 2004. The regulation of labor. Quarterly Journal of Economics 119 (4), 1339–1382.

Busciaglia, Edgardo, Dakiolas, Maria, 1999. Comparative International Study of Courts Performance. The World Bank Legal Department, Washington, DC.

Demsetz, Harold, 1967. Toward a theory of property rights. American Economic Review 57 (2), 347–359.

Djankov, Simeon, La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, 2002. The regulation of entry. Quarterly Journal of Economics 117 (1), 1–37.

Djankov, Simeon, La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, 2003. The New Comparative Economics. Journal of Comparative Economics 31 (4), 595–619.

Glaeser, Edward, Shleifer, Andrei, 2002. Legal origins. Quarterly Journal of Economics 117 (4), 1193–1229.

Gourevitch, Peter, Shinn, James, 2005. Political Power and Corporate Control: The New Politics of Corporate Governance. Princeton Univ. Press, Princeton.

Jackman, Simon, 2000. Multidimensional analysis of roll call data via Bayesian simulation: Identification, estimation, inference, and model checking. Political Analysis 9 (3), 227–241.

La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, 2006. What works in securities laws? Journal of Finance 61 (1), 1–32.

La Porta, Rafael, Lopez-de-Silanes, Florencio, Shleifer, Andrei, Vishny, Robert W., 1998. Law and finance. Journal of Political Economy 106 (6), 1113–1155.

Lewis, Jeffrey B., Linzer, Drew A., 2005. Estimating regression models in which the dependent variable is based on estimates. Political Analysis 13 (4), 345–364.

Londregan, John, 1999. Estimating legislators' preferred points. Political Analysis 8 (1), 35–56.

Martin, Andrei D., Quinn, Kevin M., 2002. Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999. Political Analysis 10 (2), 134–153.
Martin, Andrew D., Quinn, Kevin M., 2006. MCMCPack 0.7.2 (May 29, 2006). http://mcmcpack.wustl.edu.
Messick, Richard, 1999. Judicial reform and economic development: A survey of the issues. World Bank Research Observer 14 (1), 117–136.
Musacchio, Aldo, 2006. Can civil law countries get good institutions? Creditor rights and bond markets in Brazil 1850–2003. Harvard Business School Paper 6, 40.
North, Douglass, 1981. Structure and Change in Economic History. Norton and Company, New York.
Pistor, Katharina, Xu, Chengang, 2005. Governing stock markets in transition economies: Lessons from China. American Law and Economics Review 7 (1), 184–210.
Poole, Keith T., 2005. Spatial Analysis of Parliamentary Voting. Cambridge Univ. Press, New York.
Poole, Keith T., Rosenthal, Howard, 1991. Patterns of congressional voting. American Journal of Political Science 35 (1), 228–278.
Poole, Keith T., Rosenthal, Howard, 1997. Congress: A Political-Economic History of Roll-Call Voting. Oxford Univ. Press, Oxford.
Poole, Keith T., Rosenthal, Howard, 2007. Ideology in Congress. Transaction Publishers. New Brunswick, NJ.
Quinn, Kevin, 2004. Bayesian factor analysis for mixed ordinal and continuous responses. Political Analysis 12 (4), 338–353.
Rajan, Raghuram G., Zingales, Luigi, 2003. The great reversals: The politics of financial development in the twentieth century. Journal of Financial Economics 69 (1), 5–50.
Shapiro, Martin, 1981. Courts. The Univ. of Chicago Press, Chicago.
Voeten, Erik, 2000. Clashes in the assembly. International Organization 54 (2), 185–215.