

How (Not) to Measure Institutions

Stefan Voigt, MACIE (Philipps University Marburg); CESifo and ICER, Torino*

Abstract:

The statement “institutions matter” has become commonplace. A precondition for it to be supported by empirical evidence, is, however, that institutions are measurable. Glaeser et al. (2004) attacks many studies claiming to prove the relevance of institutions for economic development as being based on flawed measures of institutions, or not even on institutions at all. This paper shows that their criticism deserves to be taken seriously, but that it is somewhat overblown. Some of the difficulties in measuring institutions are described and some ways of measuring them are proposed.

Key words: Institutions, Institutions vs. Policies, Measurement, Formal vs. Informal Institutions.

JEL classification: B41, C81/82, H11, K00, O17, O43, O57.

* Prof. Dr. Stefan Voigt, MACIE (Marburg Center for Institutional Economics), Barfüßertor 2, D-35032 Marburg, Germany. Telephone: +49-6421-282 3080, Telefax: +49-6421-282 8974, e-mail: voigt@wiwi.uni-marburg.de. The author thanks the participants of the 2nd workshop “Measuring Law & Institutions” held in Paris on Dec. 15 and 16, 2007, the World Bank Seminar on “the Empirics of Governance,” and, in particular, Anne van Aaken, Matthias Dauner, Nora El Bialy, Phil Keefer, Marcus Kurtz, Sang Min Park, Katharina Pistor, Janina Satzer, Michael Seebauer, Katharina Stepping, and Kim Eun Young. The paper was written while the author was a Senior fellow at the Institute for Advanced Study in Greifswald, Germany.

How (Not) to Measure Institutions

1 Introduction

That “institutions matter” is a self-evident truth for some, a mantra for new institutional economists, and a robust empirical insight for others.¹ Over the last decade, dozens of empirical papers purporting to present evidence in support of the claim have been published. A more precise version of the phrase might be that “institutions matter crucially for economic development.” This claim has always been attacked by researchers who stress the dominance of geography (e.g., Jeffrey Sachs; see McArthur and Sachs 2001). Recently, the claim has been attacked from another angle, arguing that many—if not most—empirical studies purporting to show the crucial relevance of institutions are based on flawed, if not entirely false, indicators for institutions (e.g. Glaeser et al. 2004).

Saying that “institutions matter” implies that, due to the existence of institutions, actors behave differently than they would in the absence of institutions or in the presence of different institutions. For the statement to be meaningful, two preconditions must be satisfied. First, it must be realized that the universe is comprised of more than just “institutions”; otherwise, the statement is trivial. At times, it seems as though the term “institution” is all encompassing. In the literature, newspapers, supermarkets, and even phone booths have been described as institutions. Often, no explicit distinction is made between institutions and organizations (like firms, churches, governments, etc.). The second precondition arises once a conceptual distinction is made between institutions and noninstitutions: it must be possible to empirically ascertain institutions; otherwise, it is impossible to show their relevance empirically and saying “institutions matter” cannot be proven in any substantiated way.

This paper presents a number of proposals on how to measure institutions empirically. It is thus not another paper trying to “prove” that institutions do or do not matter; rather, it is a discussion of how to measure institutions in the first place. Only if institutions can be measured with a minimum degree of confidence are empirical statements such as “institutions matter for y” credible. There are many ways of delineating institutions;

¹ The extent to which the claim that “institutions matter” has become commonplace is remarkable. Only two decades ago, some textbooks proudly claimed that economic effects were truly independent of any institutional background and for a long time, growth theory ignored outright the possibility that institutions might have an effect on growth.

however, a minimum degree of agreement on what exactly an institution is must be reached before trying to measure them.

The main points of this contribution are: (1) measures of institutions should refer to specific institutions because aggregate measures such as “the rule of law” are too broad and fuzzy to contain meaningful information, (2) objective measures are generally preferable over subjective measures, (3) one should always aim at measuring the institution as formally specified in legislation (*de jure*) and as factually implemented (*de facto*), and finally (4) the ability to measure institutions does not imply the ability to create and modify institutions at will. Institutional optimism—or even institutional naiveté—will lead to disappointment and might even result in throwing out the baby (the New Institutional Economics) with the bathwater.

The next section of the paper presents and critically evaluates the attack that Glaeser et al. (2004) have launched against empirical studies that include institutional measures as explanatory variables. Section 3 proposes a definition for the term “institution.” Section 4 provides some bits and pieces of institutional theory. Section 5 sets forth several pragmatic proposals for measuring institutions. Section 6 contains a number of examples of flawed attempts at measurement. Section 7 concludes.

2 Have We Been Measuring Policies All These Years?

Glaeser et al. (2004) is an attack on the New Institutional Economics (NIE) containing the reproach that much of the empirical work purporting to measure the economic effects of institutions has not been measuring institutions at all, but rather policies. The authors present and evaluate two competing views on economic growth. In the first, “democracy and other checks on government” serve as mechanisms to secure property rights, which spurs investment and, eventually, income and growth—the institutional view.² According to the competing view, increased levels of human capital lead to more benign politics, less violence, and more political stability, which in turn lead to more secure property rights. In that view, better institutions are not a prerequisite to economic growth, but its consequence. The authors end up endorsing the second view but are careful enough not to break entirely with the first one (“The results of this paper do not show that ‘institutions do not matter.’ That proposition is flatly contradicted by a great

² Many scholars would, however, stress the importance of various aspects of the rule of law rather than the relevance of democracy (e.g., Barro 2000). The debate on whether democracy leads to growth or vice versa has been going on ever since Lipset (1959), but has remained largely inconclusive (Sunde 2006 is a recent survey).

deal of available empirical evidence ... Rather, our results suggest that the current measurement strategies have conceptual flaws, and that researchers would do better focusing on actual laws, rules, and compliance procedures that could be manipulated by a policy maker to assess what works.”).³

What, precisely, are the conceptual flaws in the measurement of institutions that Glaeser et al. identify? Drawing on a standard definition of institutions, they stress two chief characteristics of institutions: (1) that they constrain behavior and (2) that they are permanent or stable. Some of the frequently used measures (they cite the International Country Risk Guide, the Governance Indicators of the World Bank (Kaufmann et al. 2003), and the Polity IV measures) would neither measure policy constraints nor would they be stable; they would rather measure outcomes, i.e., policy choices.⁴ To make matters worse, the subjectivity of these measures makes it very likely that improved scores are not due to the institution being improved, but simply based on an increase in income. But if their ascertainment is influenced by income levels, they are not an adequate measure for explaining changes in income levels.

The critique of measuring institutions is well taken. Yet, some of the reasoning appears no less flawed than the measures being criticized. If the indicators used to proxy for institutions are inadequate, then these proxies are, as Glaeser et al. argue, inadequate to support the hypothesis that institutions are a prerequisite for economic growth. But if the indicators are not good proxies for institutions, they are equally poor at supporting the hypothesis that good institutions are the consequence, rather than the prerequisite, of economic development.

³ I confine the discussion to the measurement issue but cannot quite resist pointing out that the level of human capital is, itself, a consequence of institutions. Acemoglu et al. (2005) show that Glaeser et al.’s (2004) endorsement of the second view does not fit the facts. Acemoglu et al. (2007) demonstrate that the evidence in favor of the so-called modernization hypothesis formulated by Lipset (1959) completely vanishes if country fixed effects proxying for historical influences are taken into account.

⁴ It is useful to keep in mind that most institutions make a number of theoretically possible behavioral options more costly. However, this does not imply that most institutions would reduce the action space to just one possible option. In other words, there are choices within constraints. Glaeser et al. (2004) seem to assume otherwise: “These measures do not code dictators who choose to respect property rights any differently than democratically elected leaders who have no choice but to respect them.” It seems, however, reasonable to assume that even democratic governments have some discretion in the degree to which they honor private property rights.

How do Glaeser et al. propose to measure institutions properly? Unfortunately, their paper contains some general observations, but no concrete proposals. According to them, constitutional rules are likely to constrain behavior and be permanent. Yet, they note (ibid., 276) that “it is possible that these constitutional measures are noisy, and it is certain that ‘rules on the books’ are very different from what actually takes place in a country. But this is precisely the point: the institutional outcomes that scholars have used as measures of constraints have very little to do with the constitutional constraints, raising doubts about the effectiveness of changing political rules.”

Here, Glaeser et al. conflate a number of propositions. They first make the point that *de jure* and *de facto* are often worlds apart. We could not agree more. The next sentence supposedly uses constitutional rules as a benchmark and then notes that those measures of institutions that Glaeser et al. criticize are often only very loosely correlated with constitutional rules. But if *de jure* and *de facto* are often worlds apart, one can also argue that measuring *de jure* constitutional rules will not teach us anything about the factual constraints of the real world. And if constitutional rules are useless as a benchmark, why should we measure them? At the end of the day, their entire argument seems to be motivated by the possibility that changing these constraints will not necessarily lead to the intended changes in outcomes. This conflates positive economics with the “art of economics” (à la John Neville Keynes 1955). To test whether *de jure* and *de facto* deviate, we first need to be able to measure both *de jure* and *de facto*. Second, given that we have measured both and we find that they do deviate, we would want to know under what conditions *de jure* institutions are factually implemented and under what conditions they are not. All this is part of positive economics, which tries to explain certain phenomena. This undertaking would be justified even if it was impossible to change the rules in order to change the outcomes.

The constitutional variables that Glaeser et al. suggest as more appropriate benchmarks for long-term constraints are: two dealing with electoral systems (“plurality” and “proportional representation”) and two dealing with judicial constraints on government, namely, judicial independence (JI) and constitutional review. The first two measures are motivated by the work of Persson and Tabellini (2003) and the last two are taken from La Porta, Lopez de Silanes, Pop Eleches, and Shleifer (2004). *Judicial independence* is the average of three components: (1) the tenure of highest ordinary court judges, (2) the tenure of administrative court judges, and (3) a dummy coded 1 if judicial decisions are a binding source of law. *Constitutional review* is the sum of (1) the rigidity of the constitution and (2) the extent of judicial review (none, limited, full). These measures are praised as “objective”—and, indeed, they are. Glaeser et al. find that “the measures of judicial checks and balances—judicial independence and constitutional review—are uncorrelated with per capita income, and only JI is weakly correlated with outcome

indices.” The authors thus state two conclusions. First, the two objective measures for constitutional constraints are only weakly, if at all, correlated with the institutional measures that Glaeser et al. criticize, which they appear to believe is sufficient evidence for the inferiority of the institutional measures they criticize. But they do not stop there. They argue, second, that neither of the preferred objective measures is correlated with income, appearing to believe that this is a sufficient foundation for the claim that institutions do not cause growth.

These arguments are unconvincing for a number of reasons:

- (1) I do not know of any theory claiming that any of the four constitutional measures is sufficient for observing economic growth.
- (2) The two judicial measures rely exclusively on *de jure* information and hence may not accurately reflect the real world.⁵
- (3) The two electoral measures are averages for the period 1975 to 2000, whereas the two judicial measures are intended to reflect the situation in 1995. Their dependent variable, economic growth, is, however, the average for the period 1960 to 2000. It would be truly amazing were they to find that a constitutional rule claimed to be valid in 1995 caused growth that started three and half decades before!
- (4) They do not sufficiently distinguish between procedural and substantive aspects of institutions, apparently subscribing to the idea that a high number of constraints on the executive (a procedural aspect) translates into secure property rights (a substantive aspect). After noting that a commonly used measure for the level of democracy (Polity IV) attributes the worst scores to dictatorships such as those in Cuba and North Korea and also to Pinochet’s Chile, but that communist countries, including China and the USSR, fare better, they state (Glaeser et al. 2004, 277): “It is difficult to see how property is more secure in Mao’s China than in Pinochet’s Chile.” But why should a high number of procedural constraints quasi-automatically translate into secure property rights? During Mao’s tenure, securing private property rights was considered antithetical to the whole communist ideal. In other words: mechanically counting

⁵ Their coding is crude and distinguishes only between tenure that is less than six years (coded 0), tenure longer than six years but less than life (coded 1), and lifelong tenure (2). Whether “lifelong” means until death or until a fixed retirement age is not mentioned. Witold Henisz (2000) has constructed a variable that captures the *de facto* tenure of supreme court judges with very high precision. The correlation between his variable and the tenure variable constructed by Glaeser et al. is only 0.366.

constraints on the executive is not a good proxy for the security of private property if the underlying ideology of the executive's country is completely ignored.

(5) Finally, Glaeser et al. (2004, 274f.) note that “[i]f the experience of poor countries in the last 50 years is a guide, politically constrained government may not be a viable strategy for them to secure property rights.” The authors seem to believe that this insight is incompatible with the NIE, but that is not necessarily the case. Many NIE scholars, time and again, emphasize that the existence of formal—or external—institutions should not be fundamentally at odds with the existence of informal—or internal—institutions. If a society's internal institutions do not support the factual implementation of tight (constitutional) constraints on its government, wonderful formal institutions will, most likely, have few beneficial effects.

This section has both summarized and criticized Glaeser et al. (2004). Two points are worth emphasis: institutional measures should explicitly take the factual enforcement of the respective institution into account and the measures should be as objective as possible. Unfortunately, the four constitutional constraints relied upon by Glaeser et al. ignore the enforcement aspect. Measurement of institutions is key to an empirical test of whether “institutions matter.” But in order to measure institutions, we first need a definition of what they are. The next section proposes such a definition and discusses a number of possible implications.

3 Defining Institutions: A Proposal

The NIE is a young field and still lacks a commonly agreed upon definition of “institutions.” The majority of scholars defines institutions as “the rules of the game.” North (1990, 3) defines them as “the humanly devised constraints that shape interaction. In consequence, they structure incentives in human exchange, whether political, social, or economic.” His definition of institutions comprises implicit constraints, formal rules, and enforcement mechanisms. According to North, any formal rule is at least partially backed, supplemented, or contradicted by a number of implicit rules that can take the form of taboos, customs, traditions, codes of conduct, routines, conventions, and so forth (1990; 6, 43, 83).

We propose a definition that explicitly takes into account the difference between informal and formal rules, on the one hand, and between rules and enforcement, on the other. It is inspired by Ostrom (1986, 5), according to whom, rules

refer to prescriptions commonly known and used by a set of participants to order repetitive, interdependent relationships. Prescriptions refer to which actions (or states of the world) are required, prohibited, or permitted. Rules

are the result of implicit or explicit efforts by a set of individuals to achieve order and predictability within defined situations ...

Two things are of particular note and deserving of emphasis in this definition: (1) “commonly known” implies that purely private rules do not qualify as rules and (2) rules are the result of human action, but not necessarily the outcome of deliberate human design.⁶

Institutions can then be defined as commonly known rules used to structure recurrent interaction situations that are endowed with a sanctioning mechanism. North (1990) distinguishes between formal and informal institutions, using the rule component as the criterion. I prefer to distinguish between internal and external institutions, where the classification depends on who sanctions rule-breakers: if rule-breaking is sanctioned by the state, the institution is an “external” one; if rule-breaking is sanctioned by members of society, the institution is “internal.” Within the internal institutions category, a more fine-grained taxonomy could again focus on who does the sanctioning: unorganized actors (due, e.g., to norms of reciprocity) or organizations (churches, chambers of commerce, private arbitration courts)?⁷ Internal institutions thus include mores, traditions, norms, and so forth.

In their critique of frequently used measures of institutions, Glaeser et al. focus on the particular subset of “political institutions” without ever explicitly defining what these are. They do, however, explicitly refer to “democracy and other checks on government.” Other scholars also make a distinction between political and economic institutions. In Acemoglu et al. (2005a) economic institutions “determine the incentives of and the constraints on economic actors” Similarly, political institutions “determine the constraints on and the incentives of the key actors, but this time in the political sphere.” According to Acemoglu et al., political institutions allocate *de jure* political power. Political institutions determine economic institutions and the authors thus think of these institutions as hierarchically structured.

Empirically, there are a number of differences between political and economic institutions, one of the more important being that economic institutions are generally very flexible regarding the structure of an interaction.⁸ Political institutions are more likely to specify a procedure: for example, elections are held every so many years, votes

⁶ Hayek attributes this statement to the Scottish moral philosopher Adam Ferguson (1767), who, however, attributes it to French Cardinal de Retz.

⁷ Voigt and Kiwit (1998) contains a proposal for a more fine-grained taxonomy of internal institutions.

⁸ A car can be bought using cash or a bank account; the transaction can be connected to a credit contract and so forth.

are transformed into delegates in a specific way, and so forth. Another difference between the two types of institutions is that in the case of economic institutions, sanctions for failure to comply with expected behavior (e.g., a contract) are more or less precisely specified *ex ante*. Sanctions for deviations from political institution expectations are rarely so explicit.⁹

Whereas North emphasizes the difference between formal and informal rules, the distinction between economic and political institutions uses the kind of interaction as a classification criterion, and we emphasize the difference between internal and external sanctioning of rule-breakers. At times, it is not easy to precisely differentiate between the political and the economic sphere (e.g., how to classify institutions constraining state owned enterprises?), the strategy is closely related to a standard classification often used by legal scholars: economic institutions structure interactions in which all involved parties act as private actors. In case of conflict, the state may act as a neutral arbiter (judges). These are interactions based on private law. However, interactions between private parties can also be nonvoluntary, for example, car accidents or theft. In these cases, the state usually acts on behalf of the victim. This type of interaction is not based on but, rather, adjudicated according to, criminal law (which is part of public law). Finally, there are interactions between government representatives and private actors, including taxation, regulation and the safety of citizens. These interactions are based on public law, as are interactions between representatives of the state. Interpreted like this, economic institutions are largely congruent with private law and political institutions with public law.

4 Measurement Follows Theory: Assumptions and Implications

The previous section clarified our notion of institutions. Now, we take the next step and discuss how to measure them. Measurement is always based on a number of assumptions concerning potential effects of the concept to be measured (here, institutions). Our proposals on how to measure institutions will be more convincing if the underlying assumptions are set out explicitly, which we now proceed to do.

Institutions are supposed to constrain actors. A constraint implies that there are situations in which an actor who is subject to an institution prefers not to abide by the

⁹ For example, who is going to sanction a government for refusing to hold elections on the date specified in the constitution?

rule.¹⁰ If the rule reduces the number of allowed actions by an actor, then he or she might prefer not to be constrained by the rule because one of the nonallowed actions promises higher expected utility than the most attractive allowed action. In the absence of *any* sanction, then, the agent is expected not to conform to the rule. Enter sanctions and assume that an actor will never be sanctioned if he or she complies with the rules (we exclude judicial and other errors for the moment). The expected utility from breaking the rule needs to be greater than the product of the probability of being sanctioned times the utility loss of the sanction (the fine, the prison term, etc.). Assuming that the expected utility of rule-conforming behavior is higher than that of rule-breaking behavior, we expect that the agent will conform to the rule.

To know to what degree institutions constrain behavior, it is insufficient to merely assign expected utilities to the situations “comply with rule/not be sanctioned” and “not comply with rule/be sanctioned.” We also want to know the expected utility of any rule/sanction combination. Ideally, this would allow identification of the “compliance elasticity,” defined as the percentage increase in compliance after a 1 percent increase in sanctions.

Unfortunately, these expected utilities cannot be ascertained directly because the utilities of either complying with the rule or breaking the rule are not observable. Another problem is that most real-world actors will not be able to calculate the relevant expected utilities with any degree of certainty. Suppose a powerful and directly elected president considers canceling the next election. The margin of error in predicting how various groups will react—in other words, both the probability that they will act and the damage this will cause him—is extremely high. A high degree of uncertainty implies that actors will make “wrong” decisions every now and then. A (Nash) equilibrium is defined by the absence of incentives for any actor involved to unilaterally change his or her behavior. Uncertainty can induce actors to deviate unilaterally once but if their expectations from doing so are not met, they will likely soon revert to the equilibrium. Therefore, accurately measuring *de facto* institutions would seem to require observing behavior over a number of periods so as not to be misled by one-time deviations from equilibrium behavior.

To this point, we have assumed that constraints are exogenously given. However, this is obviously not true if economic institutions are determined by political institutions. Political institutions are not exogenous either. In many countries, formal constitutional

¹⁰ For the sake of completeness, we add that institutions whose purpose is to solve (pure) coordination games do not need sanctions because noncoordination and the consequent reduction in payoff is sanction enough.

change requires supermajorities, but even constitutional constraints can be changed. If different rules promise higher benefits, people will lobby for the change. In short, the content of an institution is not permanent, at least not in the long run.

Now that we have set out our definitions and provided a simple presentation of the relevant expected utility calculus, we now present a number of assumptions and implications. The attempt to measure institutions needs to be driven by an underlying theory. These theoretical considerations will thus have an influence on how we propose to make institutions measurable.

Assumption 1: The effects of institutions are due both to their substantial content and their factual implementation. Institutions bring order to an otherwise chaotic world. They allow actors to form reasonable expectations about the future, which in turn enables actors to develop a longer time horizon, make long-term investments, engage in the division of labor, and so forth. This increased level of predictability can be achieved in two interdependent ways: (1) the substantial content of a rule (e.g., degree of protection of intellectual property) and/or (2) the degree of the rule's factual implementation. For example, suppose the substantial content provides only slight protection of intellectual property, but the rule is meticulously enforced. Such an environment leads to predictability although the substantial content provides for little protection only. One can think of the two dimensions (the strength of the rule and the degree of its enforcement) as being in a substitutive relationship and some "iso predictability curve" as the result of their interaction.

Implication 1: Both *de jure* and *de facto* institutions need to be measured; otherwise, it is impossible to separate the effect of the substantive content of a rule from the effect of enforcing a rule. Neglecting to measure *de jure* institutions implies that all of them are completely identical to each other, which is obviously not the case.

Corollary 1: Many creators of indicators seem to assume simple linear relationships between an institution and some outcome. If such were indeed the case, then setting policy would be simple: just maximize or minimize (but never optimize) the institution's content. Sadly, the real world is not that simple. For example, it is very likely that up to a certain level, the strength of a rule might have positive returns, but once past that level, the returns may be marginal or even negative.

Assumption 2: The constraining effect of institutions largely depends on their factual implementation and enforcement. To ascertain whether institutions have a significant influence on any outcome variables, it is thus necessary to take their factual implementation and enforcement explicitly into account.

Institutions that are intended to formally constrain behavior substantively but whose rules are only weakly or erratically enforced are expected to constrain behavior to only a limited degree. Factual enforcement depends on the behavior of the enforcers.

Implication 2: Measures aiming at including the factual enforcement of institutions need to reflect the behavior of the enforcers. These include, most obviously, judges, prosecutors, the police, and prison staff, but the press, lobby groups, and even the public at large often act as very effective enforcers. Once again, it turns out that measuring institutions is a messy affair. Glaeser et al. (2004) might say that this would imply a mix between institutional constraints on the one hand, and factual behavior on the other. And we would reply that behavior is indeed key to effect enforcement. Thus, if one is interested in factually enforced institutions, one must take behavior explicitly into account.

Assumption 3: Factual behavior is likely determined by more than a single institution. Internal institutions can reinforce, but also weaken, the constraining effect of external institutions.

Implication 3: Institutions serve to structure specific recurring interaction situations. To understand the behavior of actors involved in a specific interaction situation, one should attempt to identify all potentially relevant institutions, i.e., both formal and informal rules as well as both internal and external sanctions. To predict likely effects of institutions, it is insufficient to focus on the analysis of single institutions. In many situations, more than one institution is likely to affect the observed behavior. In such a case, it would be premature to attribute the effect (the observed behavior) exclusively to an external institution. One should also not ignore the possibility that external and internal institutions are in conflict, i.e., following an external institution will mean breaking an internal one or vice versa.¹¹

Note that when taking into consideration a number of possibly relevant institutions, one must be careful not to dump them all together in a sort of “mixed bag” approach. It is important to try to measure each involved institution by itself so as to enable a more precise attribution of effects. This careful attention to detail will allow distinguishing

¹¹ Voigt (2004) deals with the possibility and possible consequences of the coexistence of different property rights regimes.

those institutions that truly drive the effects from those that are only marginally relevant.¹²

Assumption 4: The factual enforcement of political institutions is often extremely precarious. Think of a hierarchy of institutions. Noncompliance with economic institutions can be sanctioned via political institutions. Noncompliance by members of the administration with administrative law can be challenged via administrative courts. But who enforces formal constitutional constraints against the government? Checks and balances are an attempt to reduce the expected utility of noncompliance. If legislators pass a new law that contravenes the constitution, constitutional review via a constitutional court can lead to annulment of the law. But what if the government simply ignores the court's decision?

Assumption 5: The factual enforcement of all institutions—and political institutions in particular—is a function of informal or internal institutions. In many, many cases sanctions for noncompliance by informal means is at least and sometimes far more effective than sanctioning by formal institutions. For example, if a merchant reneges on a contract, others are going to be less than eager to enter into contracts with him. He is thus punished by his potential partners. This sanctioning by actors other than the representatives of the state promises to be potentially most important with regard to political institutions: precisely because formal enforcement is highly precarious (see Assumption 4), political institutions will often only be factually enforced if there is a credible threat of being sanctioned informally for noncompliance.

Suppose a government considers not complying with an annulment decision of a supreme court. If there is no outcry in the press, no opposition by organized interest groups, no protest by the people at large, then the government might very well expect to be better off by breaking some formal constitutional constraint. If, however, the press, interest groups, and the public do react to this government noncompliance, the expected utility of complying with the rule might be higher than that of breaking it. In other words, the factual constraining effect of some formal institution might depend on the presence of (complementary) internal institutions. To understand why a constitutional constraint is factually complied with in some countries but not in others, it is necessary to take informal institutions explicitly into account. If one is interested in understanding why constitutional constraints bind politicians in some cases but not in others, it is not

¹² Some concepts, such as judicial independence, may not be directly observable. In such cases, measuring a number of single institutions can be complemented by identifying a latent and, hence, nonobservable variable made up of the various observable institutions.

sufficient to look at a particular rule in isolation—the institutional environment needs to be taken into account explicitly.

Assumption 6: Internal institutions are largely exempt from intentional modification.

Internal institutions are enforced without reliance on the state. Exclusion from a relevant group is a sanction that has worked for millennia. There are many such institutions embodied in the traditions, mores, and norms of societies. Domestic revolutionaries and foreign colonizers alike have often been surprised by the strength of these institutions. Since they do not rely on the power of the state for their enforcement, the state often has little influence on their substantive content.

Implication 4: Suppose the factual implementation of external institutions crucially depends on a number of internal institutions (Assumption 5). Further suppose that internal institutions are largely exempt from intentional modification (Assumption 6). If the factual enforcement of external institutions does indeed depend on internal institutions, then external institutions should not be entirely at odds with internal institutions. The capacity to create external institutions that have a high chance of being factually implemented could thus be seriously constrained by the relevant internal institutions. The identification of an external institution that causes desired results is a necessary but hardly sufficient condition for the establishment of such an institution. Only if extant internal institutions are such that one can expect them to support the enforcement of the external institutions will establishment of the external institution be a success.¹³

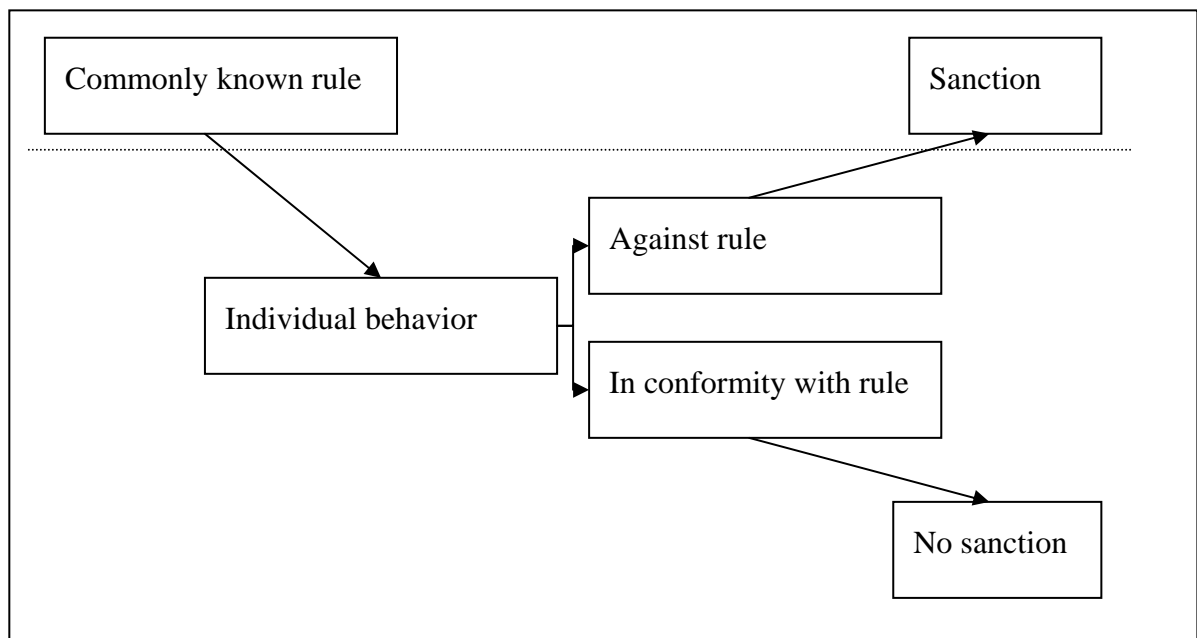
In this section, we not only proposed a taxonomy of institutions, but also conjectured as to whether and to what degree formal constraints—such as constitutional rules setting up checks and balances—will be factually enforced. The results might seem discouraging with regard to the task of measuring institutions—is such a thing even possible? In the next section, we brighten up this gloomy outlook by showing that there are, indeed, pragmatic ways of measuring institutions.

¹³ There are, of course, a number of additional considerations in the decision to introduce a new external institution, including that the benefits will outweigh the costs. In this context, Williamson's (1996, 195) delineation of efficiency is relevant: "An outcome for which no feasible superior alternative can be described and implemented with net gains is presumed to be efficient."

5 A Pragmatic Approach to Measuring Institutions

Before proposing a pragmatic approach to measuring institutions, let us review some basic preconditions to the task. First, the institutions need to be “perceptible” or “recognizable.” Generally, in regard to institutions formally passed as legislation this should not be a problem; however, in some countries legislation is published only after long delay, if at all. In countries without a written constitution, it is at times hard to ascertain what the constitutional constraints precisely are. Also, the precise content of many informal or internal institutions is never published, making it very difficult for outsiders to measure them. Ostrom (1996, 208) notes: “These rules may be almost invisible to outsiders, especially when they are well accepted by participants who do not even see them as noteworthy.”

Figure 1: The Interplay Between Institutions and Behavior



The figure above illustrates how to clarify the relationship between institutions and factual behavior. It has the following implications.

- (1) An actor subject to an institution needs to know the content of the underlying rule and understand its implications, i.e., the actor needs to be able to judge whether a behavior under consideration complies (or not) with the rule.
- (2) An actor needs to be able to anticipate more or less correctly the possible sanctions in case of noncompliance with the rule.
- (3) The potential “sanctioner” must be able to determine whether an individual has complied with the rule or not.

- (4) Finally, an external observer (such as an academic) must be able to judge all of the above!

In addition, the assumption that individual behavior that conforms to a rule will never be sanctioned may be too optimistic. Judges who wrongly convict somebody who has complied with a rule are only one possibility. In many cases, actors who want to rely on institutions are sanctioned by regime representatives simply for having dared to use an institution. This could be the case with regard to freedom of opinion, for example.

We now propose a pragmatic approach to measuring institutions. The central message is that it is essential to measure factually implemented institutions and their measurement is a lot less messy than might have been expected considering the previous discussion.

We assume all actors have unobservable preferences; however, their behavior *is* observable. In addition, an external observer can evaluate whether their behavior is in conformity with a valid rule or not. It might even be possible to introduce a scale informing us about the “distance” between the rule and the factually observed behavior. It is further assumed that the larger the distance between the behavior to be expected according to the rule and the factually observed behavior, the less binding is the formal constraint. One needs, however, to be careful not to make the reverse statement. If behavior that is in compliance with the relevant rule is observed, one does not know whether compliance is due to the threat of a sanction, or simply reflects the preferences of the actor.

Let us now make a number of pragmatic proposals on how to measure institutions. They are in line with the assumptions and implications developed in the last section, but presented in a different order.

- (1) To estimate differences between behavior expected according to some institution and factually observed behavior, we first need to select the institution in whose effects we are interested. Before starting to measure institutions, a clear and concise conception of the institution is essential. This sounds self-evident but, apparently, it is not. How else can one explain that measures for “democracy” or “the rule of law” are interpreted as measures of institutions? Neither “democracy” nor the “rule of law” are single institutions but are made up of dozens, even hundreds, of institutions. If one is interested in ascertaining the effects of specific institutions, one needs to measure these as a first step. If one believes that the effects are brought about by a whole system of institutions, one can aggregate all the single institutions into a more encompassing indicator later

on,¹⁴ but to find out what really drives the results, measures of single institutions are essential.

Starting from a clear and concise theoretical delineation can often be interpreted as involving a value judgment. It has frequently been argued, for example, that “human rights” is a concept firmly rooted in Western civilization. If we are, indeed, interested in estimating the effects of various human rights, we need to delineate them as a first step. Some might respond that this is a sort of Western imperialism, but the only way to know whether human rights have the hypothesized effects is to delineate them as clearly as possible.

(2) After having delineated an institution as precisely as possible, the next step consists in predicting the behavior that would be observable were actors to comply with the institution. If, for example, judges are appointed for 12 years, we would expect average tenure to be 12 years, with exceptions only for judges who voluntarily leave their positions early or die in office. It was pointed out above that formal constraints can be modified. For every single point in time, it is possible to identify behavior that accords with the formal rule. Spelling out *de jure* explicitly in the first place is necessary because otherwise we implicitly assume that all countries have identical rules. But some countries might not even try to secure private property rights in certain areas, such as real estate, in the first place. If this is clearly stated in the law and is also implemented as such, such a country should score high on predictability, as discussed above in Section 3.

(3) Next, factually observed behavior needs to be measured. With regard to economic institutions, this will often appear almost impossible as the behavior of thousands or even millions of actors would need to be taken into account. Political institutions pose less of a problem. Usually, there is only one head of government and only so many ministers and thus the number of potentially relevant actors is relatively limited. In some instances, empirical complexity can be reduced by choosing an appropriate sample. One example is our own attempt to measure *de facto* judicial tenure (Feld and Voigt 2003). Many countries have thousands of judges, and it would be optimal to calculate factual average tenure based on all of them. This figure would then need to be corrected by considering voluntary early retirements, deaths in office, and so forth. If one is interested in a large cross-country sample, such a task could necessitate tracing the careers of hundreds of thousands of judges. To make the task less tedious, we focused

¹⁴ To synthesize measures of single institutions into more encompassing indicators, aggregation or weighting rules are needed. Often, specific theoretical arguments on how to weigh specific institutions are lacking. However, this is a follow-up problem and need not concern us here.

only on judges sitting in the highest court of a country.¹⁵ This simplification can be justified because the judiciary is structured hierarchically and if there are problems with factual tenure at the top of the hierarchy, it can have an influence on the entire legal development of a country.

Whether political institutions are factually implemented cannot be answered based on a single point in time but must be looked at across some longer period. Suppose the constitution guarantees judges that their incomes cannot be reduced. Whether this institution is factually implemented will depend on whether it is complied with for several years or even decades. Likewise with tenure: whether factual tenure corresponds with *de jure* tenure can be decided only after many years. To measure the factual implementation of institutions over very long periods also seems justifiable because predictability is not an overnight phenomenon. Rather, predictability is conjectured to be a function of the number of periods over which a *de jure* institution has been factually enforced.

Deciding on the “optimal” period for measuring factual implementation of institutions involves various considerations:

- (a) For measurement of some institutions, there is a natural minimum period. For example, if we want to ascertain if the factual tenure of supreme court justices accords with their formally ensured tenure, we need to take at least the formal period into account (say 9 or 12 years).
- (b) The time period chosen can also have an impact on the number of available observations. Although it might be interesting to look at the implementation record of some institution for the last 100 years, if accurate observations cannot be made over the entire period, it might be more productive to look at a shorter period with better data.
- (c) If we want to use institutions as independent variables, we should make sure that they have been in place long enough to make an effect on our dependent variable plausible. Some institutions might need to be in place for a number of years before we can expect them to have any effects.
- (4) If one is interested in ascertaining the effect of institution “x” on variable “y,” it is crucial that measurement of x is not influenced by y. Again, this sounds self-evident, but many institutional measures disregard this basic rule. Many of the currently available measures are constructed on the basis of survey responses. Those surveyed can be local

¹⁵ Note that this presupposes that it is possible to determine the “highest court” without ambiguity, which might be difficult in a number of countries.

businesspeople, foreign investors, or others. Suppose the conjecture to be tested is that “secure property rights” are conducive to growth and income. When answering a question on the security of property rights in country a, the answers are very likely to be influenced by recent growth rates of that country or the country’s income level. If that is the case, the researcher is very likely to find a “significant” impact of x on y simply because the “measurement” of x is already done by taking y into account.

How to avoid this pitfall? By relying on objective data—instead of subjective evaluations—as much as possible.¹⁶ Subjective evaluations are tainted by the theories, ideologies, prejudices, and so forth of the respondent. If one is interested in the “security of property rights,” it would seem more productive to describe a specific situation in which the respondent’s security is at stake and then inquire into how many days it would take and how much money would be involved to obtain one’s rights. This is the procedure used by Djankov et al. (2003) in their Lex Mundi project. The disadvantage is, of course, that general inferences about the “security of property rights” are not possible since it is likely that the security of property rights is not identical over all kinds of property.¹⁷

Collecting “truly” objective data is no mean feat. In many countries, the number of times that, say, any judge or prosecutor has been retired against his or her will is information not readily available. The availability of accurate information could be influenced by the degree to which freedom of the press is factually existent.

(5) “Objectivity” in measurement implies that anybody repeating the identical measurement exercise should end up with exactly the same results. This is, however,

¹⁶ It is often assumed that there is no way to construct objective measures of corruption. However, there have been a number of attempts: for example, Golden and Picci (2005) propose comparing the quality of the physical infrastructure with the money that went into it. Controlling for differences in the natural environment, higher prices for identical quality imply higher levels of corruption. We are, however, not aware of any objective measure of corruption that would appropriate on a cross-country basis.

Subjective indicators do have their merits. After all, investment decisions are made by individuals whose subjective evaluations are crucial. Additionally, subjective indicators can implicitly control for a number of potentially relevant factors that might be difficult to control for using objective controls.

¹⁷ Djankov et al. (2003) describe two paradigmatic situations (cashing in of a bounced check and getting rid of a nonpaying tenant) and then ask local lawyers how long it would take to have these cases settled. These measures are, in other words, only “hypothetical de facto” measures in the sense that it is the lawyers’ beliefs about how long it would take, not an objective measure of how long it really does take to have these cases settled.

only possible if the criteria, the coding rules, the various components of a measure, and so forth are all disclosed; in other words, if the construction of the measure is transparent. Unfortunately, some of the most frequently used current measures are completely nontransparent. The Freedom House indicators are a case in point.

(6) To measure the “distance” between behavior expected according to the letter of the law and factually observed behavior, some measuring rod is needed. The appropriate measuring rod depends on the issue at stake: if it is essential that not a single deviation from the formal rule has occurred (e.g., an election entirely cancelled), then a simple dummy variable could suffice. Quite often, however, deviation from a *de jure* institution is a matter of degree. In such cases, the number of times that an institution has not been enforced over a decade or half a century can be measured. The various coding choices will also be affected by the variance of the number of deviations. This is also the case with regard to the issue of whether the measuring rod should have a linear or a logarithmic form.

(7) Sometimes, we are interested in the effects of a notion that is broader than just one single institution, such as judicial independence or procedural formalism. In these cases, the notion one is interested in is often not directly observable. To make such latent variables observable, one can resort to factor analysis, in which a number of variables are synthesized into a (lower) number of factors, or principal components. Synthesizing different variables that are interrelated among each other by different levels of correlation into one indicator follows a theory-based algorithm. Reliance on simple arithmetic means between the variables making up the indicator is not necessary. Drawing on factor analysis implies an important theoretical conjecture, namely, that the correlations between the directly measurable variables can be causally ascribed to latent concepts. Factor analysis thus condenses the information contained in the original variables into latent factors by analyzing the common variation of the variables. The values of the factors in the single countries (the factor values) are presented as deviations from the mean, which is normalized to 0. Factor analysis allows us to keep our theoretical concepts clearly separate. Within the factors, it is not one single variable that drives the results but a mix of variables. Between the groups, factor analysis has the advantage of zero correlation between the factors. The relationship between the original variables and the factors (both in terms of strength as well as direction) is represented by so-called factor loadings, which can, in turn, be interpreted as correlations.

Rosenthal and Voeten (2007) use factor analysis to identify the principal components of procedural formalism. In our own work, we use factor analysis to tease out the various dimensions hiding behind different indicators of federalism and fiscal decentralization (Blume and Voigt 2008).

Table 2, which is a simple 2x2 matrix, summarizes the most important point of this section: the most serious challenge in making institutions measurable lies in the dark cell. Although this will demand most resources, the cell shaded gray should also be taken into account as the cells cannot be compared without data for each of them.

Table 2: Desirable Institutional Measures

		Way to measure:	
		subjective	objective
Type of institution:	<i>De jure</i>		
	<i>De facto</i>		

6 How Not to Measure Institutions

This section contains a critical discussion of a few recently proposed attempts to measure institutions.

- (1) Input measures. To compare, say, judicial systems, on the basis of input measures implicitly assumes that the inputs are transformed into outputs relying on the same production function everywhere. Countries that have achieved higher levels of efficiency will not be discriminated from countries having a highly inefficient judiciary. Attempting to make any inferences about quality of the judicial system based on, for example, the “budget allocated to the judicial system per inhabitant in 2004” seems risky at best. Input measures can provide useful information on a number of topics, but without information on *de facto* implementation, they are difficult to use. An example of such measures are those provided by the so-called CEPEJ project of the European Council.
- (2) Declared output if incentives to declare are not equally distributed. The U.N. Office on Drugs and Crime regularly publishes a *Survey on Crime Trends and Operations of the Criminal Justice System*. These surveys contain a great many highly interesting and relevant data, but they should not be used as a proxy for the functioning of any institution. To give an example: part of the *Survey* documents “crimes recorded in criminal (police) statistics, by type of crime including attempts to commit crimes.” The number of crimes is then normalized by 100,000 inhabitants, which appears to give the data an aura of objectivity. Before putting too much trust into them, however, a number of questions should be dealt with: it might be necessary not only to control for population but also for the object of the

crime. In Romania, for example, only 5.5 attempts to steal cars per 100,000 were recorded. In Switzerland, this number is 899.47, which would imply that the likelihood of having one's car stolen in Switzerland is 160 times higher than in Romania. Obviously, one should also correct for the number of cars. But this is not our main point: If the police are perceived as highly unlikely to ever catch a criminal and, moreover, they are even perceived as being corrupt, then the likelihood of even notifying the police is likely to decrease. If many people perceive a low institutional quality, they might have low incentives to declare or reveal x ; a small denominator could then bias institutional quality and let it appear high precisely because it is perceived as being low!

- (3) Highly aggregate measures. The Worldwide Governance Indicators initiated by the World Bank (e.g., Kaufmann et al. 2003) have been criticized extensively and specifically.¹⁸ They purport to measure such broad concepts as “voice and accountability,” “government effectiveness,” or “rule of law.” The main critique is that the indicators are not based on a thoroughly systematized concept but that the (implicit) definition of the various concepts is based on the available surveys. These change over time, which makes comparison over various years impossible.

7 Conclusions and Outlook

In this paper, we argue that measures of institutions should be precise, objective, and take into account *de jure* as well as *de facto* elements. We hypothesize that the factual enforcement of formal institutions is likely to be heavily influenced by a number of informal institutions. When trying to estimate the (economic) effects of institutions, this possibility should be reflected by incorporating a number of covariates proxying for these informal institutions; otherwise, the danger of omitted variable bias looms large. We also point out that measuring institutions combined with econometric findings showing their significance for explaining variation in dependent variables is absolutely no basis for assuming that it is possible to modify institutions at will. If their factual enforcement is, as hypothesized, indeed dependent on informal institutions, then these could be hard constraints preventing the factual enforcement of “better” or “more modern” institutions.

But we will only know whether this is empirically correct after having estimated appropriate models. To do so, data are needed. As discussed, many of the currently

¹⁸ Among the critics are Arndt and Oman (2006), Knack (2006), Thomas (2006), and Kurtz and Schrank (2007a, 2007b). Kaufmann et al. (2007) is a reply to critics.

available institutional measures are not sufficient to refute the hypothesis that institutions do (not) matter. Major data collection exercises lie ahead.

Below is a list of some areas for which better data could increase our knowledge.

- (1) One of the first fields in which objective measures were introduced was central bank independence, the question being, of course, whether higher degrees of independence were causing lower inflation levels. Most of the indicators were, however, *de jure*. To my knowledge, the only proxy for the factual independence of central bank governors that is frequently used is their turnover rate. This is, of course, an important aspect of *de facto* independence, but there are others: How many times have the legal bases on which the bank operates changed over some given period? Have the qualification requirements for central bank governors (given that there are any) always been followed? How has the salary of the governors developed, or the budget of the entire bank? And so forth.
- (2) The central bank can be viewed as an (independent) regulatory agency entrusted with the task of providing a stable money supply. Many of the reasons for making central banks independent also apply to other regulatory agencies, implying that the independence of these other agencies, as well as their accountability, could be measured using almost the same criteria as those used to measure the independence of central banks.
- (3) Many actions by the judiciary can be interpreted as solving credible commitment problems involving actors wanting to make promises, be they private actors or government representatives. This means that frequently there are time-inconsistency problems involved, and the independence of the courts can be analyzed using criteria very similar to those used for the analysis of central bank independence. My own indicators for *de jure* and *de facto* judicial independence are examples. However, it would be extremely helpful to have an organization such as the World Bank take up these indicators and extend them to more countries, double check the answers, and so forth.
- (4) Depending on the resources available for data gathering, one might want to consider whether it is possible to measure some of the central institutions making up the rule of law. These could include its generality, its abstractness, and its certainty, along with more concrete provisions, such as the prohibition of retroactive legislation, the prohibition of expropriation without just compensation, *habeas corpus*, the protection of confidence, the principle of proportionality, and so forth.

Since collecting such data will require both money and expertise, large international organizations such as the World Bank are the best candidates for both the collection and dissemination of such data.

References

- Acemoglu, D., S. Johnson, J. Robinson (2005a); Institutions as the Fundamental Cause of Long-Run Growth, in: Ph. Aghion and St. Durlauf (eds.): *Handbook of Economic Growth*, 1A, North Holland: Elsevier, chapter 6, pp. 385-472.
- Acemoglu, D., S. Johnson, J. Robinson, P. Yared (2005); From Education to Democracy? *American Economic Review* 95(2):44–49.
- Acemoglu, D., S. Johnson, J. Robinson, P. Yared (2007); Reevaluating the Modernization Hypothesis, MIT Department of Econ WPS, Working Paper 07-23.
- Arndt, Chr. and Ch. Oman (2006). *Uses and Abuses of Governance Indicators*, Paris: OECD.
- Barro, R. J. (2000); Rule of Law, Democracy, and Economic Performance, in: *2000 Index of Economic Freedom*, Washington, DC: Heritage Foundation, 31–49.
- Blume, L. and S. Voigt (2008); Federalism and Decentralization—A Critical Survey of Frequently Used Indicators, mimeo, Philipps University Marburg.
- Djankov, S., R. La Porta, F. Lopez-de-Silanes, A. Shleifer (2003); Courts: The Lex Mundi Project, *Quarterly Journal of Economics*, 118(2):453–517.
- Feld, L. and S. Voigt (2003); Economic Growth and Judicial Independence: Cross Country Evidence Using a New Set of Indicators, *European Journal of Political Economy* 19(3):497–527.
- Glaeser, E., R. La Porta, F. Lopez-de-Silanes, A. Shleifer (2004); Do Institutions Cause Growth? *Journal of Economic Growth* 9(3):271–303.
- Golden, M. and L. Picci (2005); Proposal for a New Measure of Corruption, Illustrated with Italian Data, *Economics & Politics* 17(3):37–75.
- Henisz, W. (2000); The Institutional Environment for Economic Growth, *Economics and Politics* 12:1–31.
- Kaufmann, D., A. Kraay, M. Mastruzzi (2003); Governance Matters III: Governance Indicators for 1996–2002. World Bank Policy Research Department Working Paper.
- Kaufmann, D., A. Kraay, M. Mastruzzi (2007); The Worldwide Governance Indicators Project: Answering the Critics, The World Bank February.
- Keynes, John Neville (1955); *The Scope and Method of Political Economy*, New York.
- Knack, St. (2006); Measuring Corruption in Eastern Europe and Central Asia: A Critique of the Cross-Country Indicators, World Bank Policy Research Working Paper 3968.
- Kurtz, M. and A. Schrank (2007a); Growth and Governance: Models, Measures, and Mechanisms, *Journal of Politics* 69(2):538–54.
- Kurtz, M. and A. Schrank (2007b); Conceptualizing and Measuring Institutions: A View from Political Science, mimeo prepared for 2nd workshop on Measuring Law & Institutions held in Paris Dec. 14 and 15, 2007.
- Lipset, S. M. (1959); Some Social Requisites of Democracy: Economic Development and Political Legitimacy, *American Political Science Review* 53(1):69–105.
- McArthur, J. and J. Sachs (2001); Institutions and Geography—Comment on Acemoglu, Johnson, and Robinson, NBER Working Paper 8114.

- North, D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
- Ostrom, E. (1986); An Agenda for the Study of Institutions, *Public Choice* 48:3–25.
- Ostrom, E. (1996); Incentives, Rules of the Game, and Development, in: M. Bruno (ed.): *Annual World Bank Conference on Development Economics*, Washington, DC: World Bank, 207–34.
- Persson, T. and G. Tabellini (2003). *The Economic Effects of Constitutions*. Cambridge, MA: The MIT Press.
- Rosenthal, H. and E. Voeten (2007); Measuring Legal Systems, *Journal of Comparative Economics* 35:711–28.
- Sunde, U. (2006); Wirtschaftliche Entwicklung und Demokratie – Ist Demokratie ein Wohlstandsmotor oder ein Wohlstandsprodukt?; *Perspektiven der Wirtschaftspolitik* 7(4):471–99.
- Thomas, M. (2006); What Do the Worldwide Governance Indicators Measure? mimeo, Johns Hopkins University.
- Voigt, S. (2004); On the Coexistence of Different Property Rights Systems—And Their Consequences for Economic Growth and Development, in: E. Colombatto (ed.): *Edward Elgar Companion on Property Rights*, Cheltenham: Edward Elgar, 154–76.
- Voigt, S. and D. Kiwit (1998); The Role and Evolution of Beliefs, Habits, Moral Norms, and Institutions, in: Herbert Giersch (Hrsg.): *Merits and Limits of Markets*, Berlin et al.: Springer, 83–108.
- U.N. Office on Drugs and Crime (200*); *Eighth UN Survey of Crime Trends and Operations of Criminal Justice Systems, Covering the Period 2001–2002*.
- Williamson, O. (1996); The Politics and Economics of Redistribution and Efficiency, in: *The Mechanisms of Governance*, Oxford: Oxford University Press.