

How (Not) to measure institutions

STEFAN VOIGT*

Director, Institute of Law & Economics, University of Hamburg, Rothenbaumchaussee 36, 20148 Hamburg, Germany

Abstract: The statement ‘institutions matter’ has become commonplace. A precondition for it to be supported by empirical evidence is, however, that institutions are measurable. Some of the difficulties in measuring institutions are described and some ways of measuring them are proposed.

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 (Aron, 2000 is an early overview on the connection between institutions and growth). 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). More 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 non-institutions: it must be possible to empirically ascertain

*Email: stefan.voigt@uni-hamburg.de

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.

institutions; otherwise, it is impossible to show their relevance empirically and saying ‘institutions matter’ cannot be falsified 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.

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.

This paper deals with a straightforward, yet very basic question, namely how to measure institutions. A convincing answer is necessary before empirical insights can claim any credibility. But the paper is also confined to this question. It is not an attempt to sketch the possible future of the New Institutional Economics. Neither is it an attempt to discuss adequate econometric techniques or even empirical approaches more generally. But the question dealt with here needs to be answered before discussions concerning adequate empirical techniques even make sense.

Academics have sought to demonstrate the relevance of institutions relying on a host of approaches: from case studies via natural experiments and comparative institutional analysis to cross-country regressions. No matter what approach one relies upon, if one wants to show that institutions matter – or that they do not – one needs a reliable way to identify and measure them. It is this issue I am concerned with here.

The next section summarizes a prominent attack that has been 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 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. This paper has been very influential, at least if one takes

the number of citations as an indicator. Rather than offering a brief overview of the critique concerning the empirics of the NIE, we discuss this paper in a little more detail.² 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.³ 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, writing ‘The results of this paper do not show that “institutions do not matter”. That proposition is flatly contradicted by a great 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’.

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

2 There are hundreds of papers criticizing at least some aspects of the NIE. Bardhan (2005), for example, doesn’t doubt that institutions matter but suspects that too much focus has been laid on individual property rights and – inversely – not sufficient focus on democratic participation rights on the one hand and institutions that address coordination failures on the other. Woodruff (2006) is also interested in identifying the institutions that are fundamental for growth. He notes that one problem is that many of the currently available measures are highly correlated and critically evaluates some frequently used indicators such as those offered by Political Risk Services, Polity IV, and the World Bank’s Doing Business project. Acemoglu and Johnson (2005) is an attempt to ‘unbundle’ institutions and to distinguish between property rights on the one hand and contracting institutions on the other. Woodruff (2006) interprets their result as indicating that enforcement (he calls it informal institutions) is more important than the substantial content of the rule (which he calls formal institution). Keefer (2004) critically discusses ‘political institutions’ as coded by Persson and Tabellini (2003) who refer to them as form of government, i.e., the distinction between presidential and parliamentary systems. Keefer (2004) criticizes that Persson and Tabellini make assumptions and emphasize institutional characteristics which do not reflect the theory developed by others. They would, for example, not take sufficiently into account the considerable variation in the vote of confidence in parliamentary systems. For the (parliamentary) survival of governments, these can, however, be crucial. Dixit (2007) is interested in the policy implications of the NIE and summarizes his disappointment in four points, namely that (1) development might depend on historical/geographical conditions, (2) recipes for development would be contradictory, (3) recommendations would frequently be motivated ideologically, and (4) recommendations would often be much too abstract.

3 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).

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.⁵ 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.

How do Glaeser *et al.* propose to measure institutions properly? Unfortunately, their paper contains some general observations, but no concrete proposals. This section has summarized Glaeser *et al.* (2004). Two points are worth emphasizing: institutional measures should explicitly take the factual enforcement of the respective institution into account and the measures should be as objective as possible. 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 define institutions as ‘the rules of the game’.⁶ North (1990, 3) defines them as ‘the humanly devised constraints that

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.

⁵ Glaeser *et al.* (2004) are not the only ones stressing difficulties in the measurement of institutions. In her contribution to the Handbook of Institutional Economics, Shirley (2005) identifies some of the issues involved in measuring institutions, e.g., that the institutional variable is typically a broad aggregate and that measurement is complicated when laws do not reflect practice.

⁶ There have been many different proposals on how to define institutions. Others (e.g., Schotter 1981) have proposed to think of them as the *outcomes* of a game or as the *actors* involved in a game (e.g. Aoki 1998). More recently, Greif (2006) has proposed a very broad definition deliberately attempting

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).

I 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.⁷

Institutions can then be defined as commonly known rules used to structure recurrent interaction situations that are endowed with a sanctioning mechanism.

to encompass many previous definitions many of which are mentioned and briefly evaluated by him (ibid., 39). He proposes to define institutions as '*a system of rules, beliefs, norms, and organizations that together generate a regularity of (social) behavior*' (italics in original) and points to some advantages of his proposal. It would, e.g., circumvent the fault line between an agency perspective on the one hand and a structural perspective on the other. This is an advantage but I think it also holds for the definition I am about to propose. For the measurement of institutions, the issue of whether the agency or the structural approach deserve more merit is only marginally relevant.

Greif claims another advantage of his definition that it does not have to take sides on the issue of intentionality versus evolutionary emergence (ibid., 40ff.). These are, indeed, important advantages. But I prefer not to use Greif's definition because it also comes with some disadvantages.

His definition does, indeed, encompass both the rule of the game and the outcome of the game approach. If there is no regularity in behavior, no institution exists. Economists are interested in estimating the importance of institutions for economic results such as growth and income. If the definition of an institution is already based on the outcome, then a cause-effect relationship is difficult to establish. The conflation of at least four elements – namely (1) rules, (2) beliefs, (3) norms, and (4) organizations – into a 'system' seems to make exact measurement almost impossible. As an economist, I would, moreover, like to know the 'production function' that transfers these ingredients into behavioral regularities. Greif is aware of the impossibility of observing beliefs directly and proposes to 'deductively restrict them' (ibid., 124) by drawing on game theory.

Measuring institutions is not *l'art pour l'art*. It is the precondition for empirically assessing their relevance. An approach that interprets them as entire systems that lead to regularities in behavior is, therefore, of little help because it does not help us in making them measurable.

⁷ Hayek attributes this statement to the Scottish moral philosopher Adam Ferguson (1767), who, however, attributes it to French Cardinal de Retz.

Table 1. Types of internal and external institutions

Rule	Kind of enforcement	Type of institution	Example
1. Convention	Self-enforcement	Type-1 internal	Grammatical rules of language
2. Ethical rule	Self-commitment	Type-2 internal	Ten commandments, categorical imperative
3. Custom	Informal societal control	Type-3 internal	Norms of etiquette
4. Formal private Rule	Organized private enforcement	Type-4 internal	Rules created by merchants (arbitration courts e.g.)
5. State law	Organized state enforcement	External	Private and criminal law

North (1990) distinguishes between formal and informal institutions, using the rule component as the criterion. Since North bases his distinction on the rule component, we will refer to the distinction as formal *versus* informal rules. In addition to distinguishing the kind of rule which is part of an institution, one can also distinguish who is to do the sanctioning in case a rule is broken.⁸ If it is the state that sanctions rule-breaking, the enforcement is external to society and I propose to call the institution ‘external’; if rule-breaking is sanctioned by members of society, I propose to call the institution ‘internal’. One can also think of this as ‘public’ versus ‘private’ sanctioning⁹. Table 1 contains a proposal for a more fine-grained taxonomy of internal institutions that focuses on who does the sanctioning.¹⁰

In pure coordination games, all participants are better off if they coordinate their behavior. There is no conflictual element, so no participant has a preference for any particular Nash-equilibrium in case there is more than one. Once a particular equilibrium, a convention, has emerged nobody is able to make herself better off by deviating unilaterally from it. It is thus self-enforcing. Some authors (e.g. Sugden, 1986) have extended the convention-concept to games that do involve a certain amount of conflict and are thus mixed-motive games. Since unilateral defection does still not make any player better off, they remain self-enforcing. Conventions are called type-1-institutions here.

Ethical rules are a second kind of institution. Individuals might have internalized the strategy to cooperate as ‘the right thing to do’. Due to the

⁸ It is worth noting that the distinction does not always hold water: often, rules emerge spontaneously and become ever more formalized over time. It is unclear how formalized a rule needs to be to qualify as a formal rule. Does it need to be written down somewhere? Does it need to pass some formal legislative procedure? And so forth.

⁹ This taxonomy builds on the old distinction between ‘state’ and ‘society’ which may be defunct now. But the state still claims the monopoly to use force (Weber). I thank a reviewer for suggesting the distinction between ‘public’ and ‘private’ sanctioning.

¹⁰ Note that most interaction situations are likely to involve more than a single institution: I might follow a state law not only because I am afraid of the sanction but also because I believe the state law should be like it is and I have internalized a corresponding norm. Possible consequences of this are taken up next.

internalization, participants want to comply with the ethical rules even if they conflict with narrowly defined self-interest. Internally enforced institutions are called type-2-institutions here.

Whereas the first two ways of enforcement are rooted within the structure of the game or the actor himself, there are other ways of enforcement which rely on other actors. Enforcement via societal control is one of them. An unspecified variety of persons survey the compliance by way of spontaneous control. This is the third type of enforcement, one possible example being to sanction non-compliance by informing others about this behavior in order to diminish the reputation of the person who did not comply. Customs that are enforced via non-organized societal control are called type-3-institutions here.

We call the fourth type of internal institutions to be discussed now private rules. Enforcement that makes use of other actors, i.e., third-party-enforcement, can also be based on some kind of organization. Organized private enforcement may, for example, rely on private courts of arbitration that monitor the compliance with private rules. The enforcement of rules by private organizations is called type-4-institution here.

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.* (2005) 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.

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? Nevertheless, a number of differences between political and economic institutions appear noteworthy. Economic institutions are generally very flexible regarding the structure of an interaction. For example, the civil code of most countries allows a large variety of ways in which property can be transferred from one owner to another: a car can be bought using cash or a bank account; the transaction can be connected to a credit contract and so forth. Political institutions are more likely to tell the addressees what to do: for example, to hold elections every so many years, how to transform votes into parliamentary seats, and so forth. In a sense, economic institutions are enabling (they enable private actors to achieve their goals by using institutions as tools) whereas political institutions are constraining (they tell politicians what to do, hence reducing their individual leeway). Another difference between the two types of institutions is that in the case of economic institutions, sanctions for

Table 2. Delineating institutions according to 3 different criteria

Rules:	Informal		Formal	
	Internal	External	Internal	External
Sanctions:				
Political institutions				
Economic institutions				

failure to comply with expected behavior (e.g., a contract) are more precisely specified than sanctions for renegeing upon political institutions. A thief who has stolen a car and is caught might have a pretty accurate expectation of the sanction he will face whereas the sanction government will face for not holding a constitutionally guaranteed election is highly uncertain.

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 I emphasize the difference between internal and external sanctioning of rule-breakers. It seems even possible to combine these three criteria. Table 2 is an attempt to depict this graphically.

As a last step, I propose to explicitly distinguish between *de jure* and *de facto* institutions: studying formal institutions will often not teach us a lot about the real world. Many rules formally passed by parliaments are never enforced. If we are interested in ascertaining how institutions really matter, we should not be content with measuring formal rules but also their factual enforcement. This promises to be messy: to learn anything about factual enforcement, we need to study the behavior of the police, judges, and so on.

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 rule.¹¹ 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 non-allowed actions promises higher expected utility than the most attractive allowed action. In the absence of *any* sanction, then, the agent is expected not

¹¹ For the sake of completeness, we add that institutions whose purpose is to solve (pure) coordination games do not need sanctions because non-coordination and the consequent reduction in payoff is sanction enough.

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% 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 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 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 the substantial content of their rule component 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 is the consequence of two interdependent components: (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 or enforcement (terms which are used interchangeably here). 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.¹²

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 everywhere, 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. Completely in line with standard economics, it would seem to make sense to assume decreasing marginal returns. Beyond that, it cannot even be excluded that some institutions will only have positive effects up to a certain level, but once past that level, the returns may be even negative.

Assumption 2: The constraining effect of institutions largely depends on their factual implementation. To ascertain whether institutions have a significant influence on any outcome variables, it is thus necessary to take their factual implementation 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 aimed at including the factual enforcement of institutions need to reflect the behavior of the enforcers. These include, most obviously, the police, prosecutors, judges, and prison staff, but the press, lobby groups, and even the public at large often also act as enforcers. 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

12 The notion of 'iso predictability' might have policy implications: given that an identical level of predictability can be attained by different combinations between substantial content and factual enforcement, it would be rational to choose that combination that causes the lowest costs.

behavior is indeed key to effective 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. Four relationships between internal and external institutions can be conceived of: (1) It can be *neutral*, i.e., the institutions regulate different areas of human interaction. (2) It can be *complementary*, i.e., the institutions constrain human behavior in an identical or similar fashion and rule-breaking behavior is sanctioned by private individuals as well as representatives of the state. (3) It can be *substitutive*, i.e., the institutions constrain human behavior in a similar fashion but rule-breaking behavior is sanctioned *either* by private individuals *or* by representatives of the state. (4) It can be *conflicting*, i.e., the institutions constrain human behavior in different ways. Abiding by an internal institution would then be equivalent to breaking an external one and *vice versa*.

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. Not taking internal institutions explicitly into account would lead to omitted variable bias.

Examples where external institutions and internal ones are in conflict abound: Think of de Soto's (1990) case study of three informal sectors within the Peruvian economy: informal housing, informal trade, and informal transportation. De Soto's central conjecture is that the size of the informal sector is a function of the compatibility of external and internal institutions, or, in his own words (1990, 12): 'We can say that informal activities burgeon when the legal system imposes rules which exceed the socially accepted legal framework – does not honor the expectation, choices, and preferences of those whom it does not admit within its framework – and when the state does not have sufficient coercive authority'.

Ellickson (1986, 1991, 1994) studies how disputes between ranchers and farmers are settled in Shasta County, California. This county was chosen because in some parts of it ranchers are strictly liable for cattle trespassing, while in others they are not. Ellickson was able to show that, no matter what legal rules prevailed, the way neighbors resolved their conflict remained unchanged, that is, institutions run by the state did not have any effect on the kind of conflict-resolution chosen. This study tells us that, under certain circumstances (repeat interactions), internal institutions still trump state-run ones, even in such highly developed societies as California.

Berkowitz *et al.* (2002) deal with legal transplants, i.e., with attempts to transplant legislation from one country to another. They argue that the way the law was initially transplanted and received is an important determinant of its subsequent factual enforcement. According to their findings, countries that have adapted transplanted legal orders to local conditions and (or) had a population that was already familiar with basic legal principles of the transplanted law have more effective legality than countries that received foreign law without any similar pre-dispositions.¹³

A correct assessment of the effects that institutions have on observed outcomes would thus need to take all potentially relevant institutions into account. According to the examples just offered, these are not confined to external institutions but should often include internal institutions.

Note that when taking into consideration a substantial 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 those institutions that truly drive the effects from those that are only marginally relevant.¹⁴ To give an example: Some organizations offer indicators for ‘the rule of law’. Although academics still quarrel how best to define the concept, a consensus on one thing should exist, namely that the rule of law consists of many different dimensions such as (1) the separation of powers, (2) judicial review, (3) judicial independence, (4) judicial accountability, (5) prosecutorial independence, (6) fair trial, and (7) basic human rights. This list is admittedly subjective and other academics will come up with other lists. But what is clear is that the rule of law comprises many dimensions. For the list just offered, I have shown elsewhere that most partial correlations are amazingly low, implying that valuable information is lost by aggregating them into a one-dimensional indicator (Voigt 2012).

¹³ A very early example for a failed attempt to transplant public law was given by the former Lord Chancellor of Henry VI, Sir John Fortescue who fled to France and then described the differences between England and the Continent in ‘A Learned Commentation of the Politique Laws of England’ (first published in the 15th century). The superiority of the English system concerning wealth, happiness, and the entire rule-system was so evident that it was hard to understand why the whole world did not simply try to emulate the British law system. His answer: The institution of trial by jury - which he evaluates positively - depended on a specific economic and social structure that was present only in England (Macfarlane, 1978, 179ff.).

MacArthur’s Japanese Constitution is often cited as an example for a successful transplant of public law. Yet, Inoue (1991) shows that the Japanese language version of the constitution is more compatible with Japanese internal institutions than the English language version and how the misunderstandings between American and Japanese participants to the translation of the American draft facilitated the acceptance of the new Japanese Constitution.

¹⁴ 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, non-observable variable made up of the various observable institutions.

Assumption 4: The factual enforcement of political institutions is often extremely precarious. Think of a hierarchy of institutions. Non-compliance with economic institutions can be sanctioned via political institutions. Non-compliance 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 non-compliance. 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? Recognition of the assumption that factual enforcement of political institutions is extremely precarious directly leads to the next assumption.

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 non-compliance by informal means is at least as 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.¹⁵ Sanctioning by actors other than the representatives of the state promises to be potentially most relevant with regard to political institutions: precisely because state enforcement is highly precarious (see Assumption 4), political institutions will often only be factually enforced if there is a credible threat of being sanctioned for non-compliance by other than state representatives.

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 non-compliance, 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 sufficient to look at a particular rule in isolation – the institutional environment needs to be taken into account explicitly.¹⁶

¹⁵ This type of sanctioning has been studied extensively with regard to exchange between merchants. Greif (2006) recounts the story of the Maghribi traders and Bernstein (e.g., 1992) shows that this kind of coordination is still very present in today's world. Her paper refers to Jewish diamond traders in New York.

¹⁶ De jure institutions can thus create focal points à la Schelling (1960). Their creation, even if not taken very seriously by their creators, can at times have far-reaching effects: It is argued (e.g., Thomas 1999) that the Conference for Security and Cooperation in Europe created such focal points for the populations

Assumption 6: Internal institutions are resistant to 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.

Again, examples abound: The Hindu tradition of burning widows on the funeral pyre of their dead husbands was officially abolished in British India in 1829. Yet, instances of ‘suttee’ – as this tradition is also called – occur until today. Also, in India, although the dowry is officially prohibited, advertisements in the classified sections of many newspapers take little effort in paraphrasing that a substantial dowry is expected. Many states the world over tried to ban the consumption of alcohol by outlawing its manufacture, transportation, or sale. The prohibition of trading some goods and services has often had very similar effects; just think of prostitution or drug trade. Massell (1968) is a very detailed and illuminating description of how the Soviets tried to make the populations in their Central Asian Republics give up the Sharia – and how their efforts were an almost complete disaster.

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 resistant to 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.¹⁷

In this section, I presented a number of assumptions 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

of Central and Eastern Europe. Demonstrators referred to the documents later on. A similar argument can be found in Sen (2009) who argues that at the beginning, human rights declarations might be nothing more than ethical statements that contain, however, the invitation to pass legislation in conformity with them. Given that this legislation is enforced, purely ethical proclamations might lead to changes in the factually enforced institutions. I thank Reiner Eichenberger for pointing out Sen (2009) to me.

17 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’.

with regard to the task of measuring institutions – is such a thing even possible? In the next section, I brighten up this gloomy outlook by showing that there are, indeed, pragmatic ways of measuring institutions.

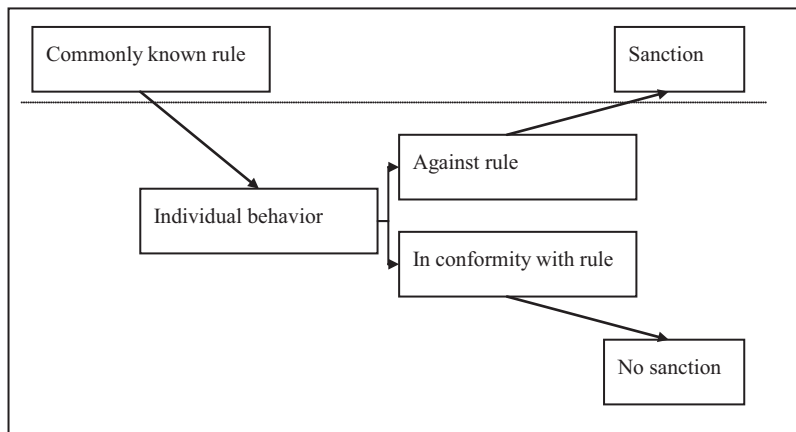
5. A Pragmatic approach to measuring institutions

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 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 non-compliance 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.

Figure 1. The interplay between institutions and behavior.



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. Daring to take basic human rights seriously, such as to freely assemble, is a case in point.

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.

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.¹⁸ 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,¹⁹ 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.

18 Cheibub *et al.* (2010) develop strong arguments in favor of a dichotomous variable for measuring political regimes, distinguishing between democracies and dictatorships. Now political regimes are definitely not identical with institutions, but many of their arguments also apply to our measurement issue. They plead in favor of a minimalist definition confined to procedural aspects (‘are there contested elections for office?’) and to abstain from a multitude of additional aspects. This can be interpreted as the attempt to minimize the number of institutions to be measured which needs to be based on a clear theoretical formulation on what constitutes the core of political regimes. Inclusion of additional aspects (more institutions) would only lead to additional problems such as introducing subjective evaluations and the necessity of a theoretically founded aggregation rule. Both of their concerns are picked up later in this section.

19 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. But see Nardo *et al.* (2005) for a useful handbook.

If we are, indeed, interested in estimating the effects of various human rights, we need to delineate them as a first step. The only way to know whether human rights have the hypothesized effects is to delineate them as clearly as possible. The result itself is entirely positive science.²⁰

(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 tenure to be 12 years, with exceptions only for judges who voluntarily leave their positions early or die in office. 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 nevertheless, 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 only on judges sitting in the highest court of a country.²¹ 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

20 There are many basic human rights (Law and Versteeg (2011) identify up to 60 different such constitutionally guaranteed rights) but the number of countries in which their effects can be observed is severely limited. Also, a high degree of multicollinearity in the factual implementation of some of these rights is likely. Factor analysis can help to identify a limited number of principal components to mitigate these problems. This is what Blume and Voigt (2007) do to ascertain the economic effects of human rights.

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

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:

- (1) 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).
- (2) 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.
- (3) 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.²²

(4) If one is interested in ascertaining the effect of institution ‘*x*’ on variable ‘*y*’, it is crucial that measurement of *x* is not tainted 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 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.²³ Subjective evaluations are tainted by the

22 All institutions are endogenous. It is therefore crucial not to neglect problems due to endogeneity. But again, this is not the place to discuss pros and cons of various techniques such as instrumental variables, matching etc. Here, the only question is how to measure institutions.

23 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. Olken (2009) is

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.²⁴ 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. For someone from Western Europe, the attempt to proxy for the security of property rights by estimating the costs of cashing in on a bounced check seems bizarre, as checks are not in common use there anymore.

The Djankov *et al.* project does not rely on ‘real’ objective data but on hypothetical data only. It is not information on real cases on which the dataset is based but beliefs of lawyers on how long it would take were such a case to arise in their country. Analyzing a standardized situation (as, e.g., the difficulty of cashing in on a bounced check) has the advantage of making comparisons across countries possible. Yet, it comes at an important cost, namely that the hypothetical might not be representative of the real situation in individual countries.²⁵

Collecting ‘truly’ objective – and not only hypothetically objective – data thus remains a desideratum but 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. It might, hence, be sensible to control for it in estimated models.

(5) ‘Objectivity’ in measurement implies that anybody repeating the identical measurement exercise should end up with exactly the same results. This is, however, 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 not completely transparent. ‘Freedom in the World’ by the New York-based NGO Freedom House (2008) is a composite indicator depicting both political rights and civil liberties. The broad range of issues for which these

an ingenious attempt to confront corruption perceptions with corruption reality regarding costs of road construction in Indonesian villages. Razafindrakoto and Roubaud (2010) as well as Donchev and Ujhelyi (2008) show that actual corruption experiences as reported in household surveys are weak predictors of corruption perceptions by ‘experts’.

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.

24 Djankov *et al.* (2003) describe two paradigmatic situations (cashing in of a bounced check and getting rid of a non-paying tenant) and then ask local lawyers how long it would take to have these cases settled.

25 Thanks to an anonymous referee for insisting on this.

indicators have been used is somewhat disturbing: some use them as a proxy for democracy, others for the security of property rights, still others for the rule of law, but it is certainly not the creators who should be held liable for the (mis-)use of their product. Although Freedom House recently began to make the scores of three (four) sub-categories available for the political rights (civil liberties) indicator, this is still a far cry from publishing single codings. For example, the third sub-category of civil liberties is ‘rule of law’ which is itself based on many different aspects the codings of which are still not published.²⁶

(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.

The distance between expected and factually observed behavior will be ascertained on the basis of many observations, i.e., on the basis of some average. Given that the primary function of institutions is to enhance predictability, we are not only interested in the average of factually observed behavior itself but also in the observed variation: if it is low, then predictability can, nevertheless, be high. If variation is, however, high then forming expectations that have a high probability of materializing is correspondingly difficult. *De facto* measures of institutions should, hence, also report the corresponding degree of variation.²⁷

(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

²⁶ Freedom House’s webpage lists four sub-sub-categories with 24 different aspects. Although many of the indicator-producing organizations have somehow responded to the demand for more transparency, many indicators remain opaque. This also holds for the very widely used Worldwide Governance Indicators published by Kaufmann *et al.* (2009). In their critique on governance indicators, Oman and Arndt (2010) point out that their greatest pitfall is lacking transparency in their production.

²⁷ Shleifer and Vishny (1993) argue that it is not necessarily corruption as such that it is bad but the uncertainty in its implementation: is it sufficient to bribe one government official – or will an unpredictable number of officials come up with additional demands for a bribe?

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 2011).

Many of the statements on how to measure institutions have been made with external institutions in mind. So how about measuring internal institutions? *Prima facie*, this seems to be the harder challenge.²⁸ But this is not necessarily true: given that institutions are not only internal but also informal, i.e., nowhere codified, the *de jure de facto* divergence collapses and one problem regarding the measurement of external/formal institutions vanishes. One strong argument in favor of unbundling entire sets of external institutions into single institutions was made with a view to policy implications: if governments want to change institutions to foster economic growth, then they need to know exactly what institutions to modify. But if internal institutions are much less amenable to deliberate change, then this reason in favor of unbundling is no longer valid.²⁹

Over the last couple of years, many datasets purporting to measure institutions have become available. Their number is so large that some overviews of the available datasets have been published (for example by the Interamerican Development Bank (without year) or UNDP (2009)). There is even an organization whose primary task it is to collect, re-organize, and disseminate such datasets (The Quality of Governance Institute at the University of Gothenburg in Sweden). It is thus impossible to evaluate the quality of many such measures here, let alone all of them. For the better known indicators, this is, in addition, superfluous as many very sharp critiques already exist. Take the Worldwide Governance Indicators initiated by the World Bank (e.g., Kaufmann *et al.*, 2003)

²⁸ As one referee of this journal suggested.

²⁹ One reason to aim at the unbundling of internal institutions is exactly that we cannot be sure whether our assumptions regarding their longevity are correct. To know we first have to measure them (separately).

as an example for highly aggregate measures that have been criticized extensively and specifically.³⁰ 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.

In a companion paper (Voigt 2012), I try to be more constructive and apply the reasoning presented here to propose a measure for the rule of law. One reason for choosing the rule of law as an example was that it is part of the very prominent Worldwide Governance Indicators.

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

30 Among the critics are Arndt and Oman (2006), Knack (2006), Langbein and Knack (2010), Thomas (2010), and Kurtz and Schrank (2007a, 2007b). Kaufmann *et al.* (2007) is a reply to critics.

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 stable money supply. The most frequently cited reason for establishing independent central banks is to mitigate the difficulty of politicians seeking to be re-elected to credibly commit to stable monetary policy. The difficulty to credibly commit to certain policies also arises in other policy areas such as competition or environment. Therefore, the reasoning underlying central bank independence can also be applied to 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.

Rodrik *et al.* (2004) argue not only that institutions are clearly more relevant for explaining development than both geography and trade but also that institutions ought to be conceptualized as ‘the cumulative outcome of past policy actions’ (*ibid.*, 156). They propose to think of policy as a flow variable and institutions as a stock variable and thus turn one established way to think about the relationship between institutions and policies on its head. Such a view does not only have important implications for adequately estimating the effects of institutions (it is inappropriate to regress outcomes on institutions and policies simultaneously as measures of institutional quality already contain relevant information on the effects of policies). Such a view also points at the necessity to research into the determinants of institutional change. Needless to say: to ascertain its determinants empirically, institutions need to be measured first.

References

- Acemoglu, D. and S. Johnson (2005), ‘Unbundling Institutions’, *Journal of Political Economy*, 113(5): 949–995.

- Acemoglu, D., S. Johnson, and J. Robinson (2005), 'Institutions as the Fundamental Cause of Long-Run Growth', in P. Aghion and S. Durlauf (eds.), *Handbook of Economic Growth*, vol. 1A, North Holland: Elsevier, chapter 6, pp. 385–472.
- Arndt, Chr. and Ch. Oman (2006), *Uses and Abuses of Governance Indicators*, Paris: OECD.
- Aron, J. (2000), 'Growth and Institutions: A Review of the Evidence', *World Bank Research Observer*, 15(1): 99–135.
- Bardhan, P. (2005), 'Institutions Matter, But Which Ones?', *Economics of Transition*, 13(3): 499–532.
- Barro, R. J. (2000), 'Rule of Law, Democracy, and Economic Performance', in *2000 Index of Economic Freedom*, Washington, DC: Heritage Foundation, pp. 31–49.
- Berkowitz, D., K. Pistor, and J.-F. Richard (2002), 'Economic Development, Legality, and the Transplant Effect', *European Economic Review*, 47(1): 165–195.
- Bernstein, L. (1992), 'Opting Out of the Legal System: Extralegal Contractual Relations in the Diamond Industry', *Journal of Legal Studies*, 21(1): 115–157.
- Blume, Lorenz and Stefan Voigt (2007), 'The Economic Effects of Human Rights', *Kyklos*, 60(4): 509–538.
- Blume, Lorenz and Stefan Voigt (2011), 'Federalism and decentralization – a critical survey of frequently used indicators', *Constitutional Political Economy*, 22(3): 238–264.
- Cheibub, J., J. Gandhi, and J. Vreeland (2010), 'Democracy and Dictatorship Revisited', *Public Choice*, 143: 67–101.
- Dixit, A. (2007), 'Evaluating Recipes for Development Success', *World Bank Research Observer*, 22(2): 131–157.
- Djankov, S., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer (2003), 'Courts: The Lex Mundi Project', *Quarterly Journal of Economics*, 118(2): 453–517.
- Donchev, D. and G. Ujhelyi (2008), What Do Corruption Indices Measure?, Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1124066.
- Ellickson, R. (1986), 'Of Coase and Cattle: Dispute Resolution Among Neighbors in Shasta County', *Stanford Law Review*, 38(3): 623–687.
- Ellickson, R. (1991), *Order Without Law*, Cambridge, MA: Harvard University Press.
- Ellickson, R. (1994), 'The Aim of Order Without Law', *Journal of Institutional and Theoretical Economics*, 150(1): 97–100.
- 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.
- Freedom House (2008), Freedom in the World 2008. Available at: <http://www.freedomhouse.org/report/freedom-world/freedom-world-2008>.
- Glaeser, E., R. La Porta, F. Lopez-de-Silanes, and 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.
- Greif, A. (2006), *Institutions and the Path to the Modern Economy – Lessons from Medieval Trade*, Cambridge, MA: Cambridge University Press.
- Inoue, K. (1991), *MacArthur's Japanese Constitution – A Linguistic and Cultural Study of Its Making*, Chicago: The University of Chicago Press.
- Interamerican Development Bank (without year). DataGov- Governance Indicator Database, Available at: <http://www.iadb.org/datagob/>
- Kaufmann, D., A. Kraay, and M. Mastruzzi (2003), Governance Matters III - Governance Indicators for 1996–2002. World Bank Policy Research Working Paper No. 3106.

- Kaufmann, D., A. Kraay, and M. Mastruzzi (2007), The Worldwide Governance Indicators Project: Answering the Critics, The World Bank February.
- Kaufmann, D., A. Kraay and M. Mastruzzi (2009), Governance Matters VIII: Governance Indicators for 1996–2008. World Bank Policy Research Department Working Paper No. 4978.
- Keefer, P. (2004), ‘What Does Political Economy Tell Us About Economic Development – And Vice Versa?’, *Annual Review of Political Science*, 7: 247–272.
- 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): 538–554.
- 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.
- Langbein, L. and S. Knack (2010), ‘The Worldwide Governance Indicators: Six, One, or None?’, *Journal of Development Studies*, 46(2): 350–370.
- Law, D. and M. Versteeg (2011), The Evolution and Ideology of Global Constitutionalism. *California Law Review*, 99(5): 1163–1257.
- Lipset, S. M. (1959), ‘Some Social Requisites of Democracy: Economic Development and Political Legitimacy,’ *American Political Science Review*, 53(1): 69–105.
- Macfarlane, A. (1978), *The Origins of English Individualism*, Oxford: Blackwell.
- Massell, G. (1968), ‘Law as an Instrument of Revolutionary Change in a Traditional Milieu: The Case of Soviet Central Asia’, *Law and Society Review*, 2: 179–214.
- McArthur, J. and J. Sachs (2001), Institutions and Geography –Comment on Acemoglu, Johnson, and Robinson, NBER Working Paper 8114.
- Nardo, M., M. Saisana, A. Saltelli, and S. Tarantola (2005), *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Paris: OECD.
- North, D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge, MA: Cambridge University Press.
- Olken, B. (2009), ‘Corruption Perceptions vs. Corruption Reality,’ *Journal of Public Economics*, 93(7–8): 950–964.
- Oman, C. and C. Arndt (2010), Measuring Governance, Policy Brief No. 39, OECD Development Centre, Paris.
- 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, pp. 207–234.
- Persson, T. and G. Tabellini (2003), *The Economic Effects of Constitutions*, Cambridge, MA: The MIT Press.
- Razafindrakoto, M. and F. Roubaud (2010), ‘Are International Databases on Corruption Reliable? A Comparison of Expert Opinion Surveys and Household Surveys in Sub-Saharan Africa,’ *World Development*, 38(8): 1057–1069.
- Rodrik, D., A. Subramanian, and F. Trebbi (2004), ‘Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development’, *Journal of Economic Growth*, 9(2): 131–165.
- Rosenthal, H. and E. Voeten (2007), ‘Measuring Legal Systems’, *Journal of Comparative Economics*, 35: 711–728.
- Schelling, Th. (1960), *The Strategy of Conflict*, Cambridge, MA: Harvard University Press.

- Schotter, A. (1981), *The Economic Theory of Social Institutions*, Cambridge, MA: Cambridge University Press.
- Sen, A. (2009), *The Idea of Justice*, London: Penguin Books.
- Shirley, M. (2005), Institutions and Development. in C. Ménard and M. Shirley (eds.), *Handbook of the New Institutional Economics*, Heidelberg: Springer, pp. 611–638.
- Shleifer, A. and R. Vishny (1993), ‘Corruption’, *Quarterly Journal of Economics*, 108(3): 599–617.
- Soto, H. de (1990), *The Other Path – The Invisible Revolution in the Third World*, New York: Harper & Row.
- Sugden, R. (1986), *The Economics of Rights, Cooperation and Welfare*, New York: Blackwell.
- Sunde, U. (2006); ‘Wirtschaftliche Entwicklung und Demokratie – Ist Demokratie ein Wohlstandsmotor oder ein Wohlstandsprodukt?’, *Perspektiven der Wirtschaftspolitik*, 7(4): 471–499.
- Thomas, D. C. (1999), ‘The Helsinki Accords and Political Change in Eastern Europe’, in T. Risse, S. Ropp, and K. Sikkink (eds.), *The Power of Human Rights: International Norms and Domestic Change*, Cambridge, MA: Cambridge University Press.
- Thomas, M. (2010), ‘What Do the Worldwide Governance Indicators Measure?’, *European Journal of Development Research*, 22: 31–54.
- UNDP (2009). Governance Indicators – A User’s Guide, Oslo Governance Centre. 2nd edition.
- Voigt, S. (2012), ‘How to Measure the Rule of Law’, *Kyklos*, 65(2): 262–284.
- Williamson, O. (1996), ‘The Politics and Economics of Redistribution and Efficiency’, in *The Mechanisms of Governance*, Oxford: Oxford University Press.
- Woodruff, C. (2006). ‘Measuring Institutions’, in S. Rose-Ackerman (ed.), *International Handbook on the Economics of Corruption*, Cheltenham, UK: Edward Elgar, pp. 105–124.