Abstract:

The statement “institutions matter” has become commonplace. For the statement to be supported by empirical evidence, institutions need to be measurable. Recently, Glaeser et al. (2004) attacked many studies claiming to prove the relevance of institutions for economic development as 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 is overstretched at the same time. Some of the difficulties in measuring institutions are described and some ways how to measure institutions are proposed.

Key words: Institutions, Institutions vs. Policies, Measurement, Formal vs. informal institutions.

JEL classification: B41, C81/82, H11, K00, O17, O43, O57.
How (Not) to Measure Institutions

1 Introduction

The statement "institutions matter" is a self-evidenced truth for some, the mantra of the New Institutional Economists for others, and a robust empirical insight for yet others. Over the last decade, dozens of empirical papers purporting to present evidence in favor of the claim have been published. A more complete version of the statement could read "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 were based on flawed – if not entirely false – indicators for institutions.

The statement "institutions matter" implies that due to the existence of institutions, actors behave differently than they would behave in the absence of institutions – or under different institutions. In order for the statement to be meaningful, two preconditions need to be satisfied: (1) The universe does not only exist of institutions, otherwise the statement becomes trivial. At times, one gets the impression that the term "institutions" is almost 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 and the like). After having separated institutions from non-institutions conceptually, another precondition needs to be satisfied: (2) Institutions can be empirically ascertained, otherwise it is impossible to show their relevance empirically – and the statement "institutions matter" could not be an empirical statement.

This paper makes a number of proposals on how to measure institutions empirically. It is, hence, not another paper trying to ‘prove’ that institutions matter - or that they do not. It is, rather, discussing how to measure institutions in the first place. Only if institutions can be measured with a minimum degree of confidence are empirical

---

1 The spread with which the claim that “institutions matter” has become commonplace is worth to be explicitly mentioned: only two decades ago, some textbooks proudly claimed that insights from economics were true independently of any institutional background. For long, growth theory outright ignored the possibility that institutions might have an effect on growth.
statements such as “institutions matter for y” possible. There are many ways of
delineating institutions. A minimum degree of consent on what institutions are needs to
be achieved before trying to measure them.

Our main points in 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 disappointments and might even entail the danger to throw out the
baby (the New Institutional Economics) with the bathwater.

The next section of the paper presents and critically evaluates the critique that Glaeser
et al. (2004) have launched against empirical studies containing institutional measures
as explanatory variables. The claim that institutions matter is inherently incomplete.
One would like to know what they matter for. This presupposes a theory linking
institutions to outcomes by one or more transmission channels. Section three proposes
a definition for the term “institution” whereas section four presents some bits and
pieces of institutional theory. In section five, the main difficulties in measuring
institutions are discussed. Section five makes a number of pragmatic proposals on how
to measure institutions. Section six gives a number of examples for flawed attempts at
measurement. Section seven concludes.

2 Have We Been Measuring Policies all these years?

Glaeser et al. (2004) have recently published an attack on the New Institutional
Economics which derives its force primarily from the reproach that much of the
empirical work pretending to measure the economic effects of institutions has indeed
not been measuring institutions, but rather policies. The authors present and evaluate
two competing views on economic growth, namely a first one according to which
“democracy and other checks on government” serve as mechanisms to secure property
rights, which would then spur investment and eventually income and growth – the
institutional view.2 According to the competing view, increased levels of human capital

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
would lead to more benign politics, less violence and more political stability, which would, in turn, lead to more secure property rights. Better institutions are here 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 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 are the conceptual flaws in the measurement of institutions that Glaeser et al. identify? Drawing on a standard definition of institutions they stress two qualities, namely (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. On top, the subjectivity of these measures would make it very likely that they “increase” when income increases. But if their ascertainment is influenced by income levels, they are not an adequate measure explaining changes in income levels.

Their critique concerning the measurement problems of institutions is well taken. Yet, some of their reasoning appears no less flawed than the measures they criticize. If the indicators used to proxy for institutions are inadequate, then these proxies are, as

leads to growth or rather growth to democracy has been going on ever since Lipset (1959) but has remained largely inconclusive (Sunde 2006 is a recent survey).

We confine our discussion to the measurement issue but cannot deny ourselves noting 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) show that the evidence in favour of the so-called modernization hypothesis forcefully formulated by Lipset (1959) completely vanishes if country fixed effects proxying for historical influences are taken into account.

It might be 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: 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.
Glaeser et al. argue, inadequate to support the hypothesis that institutions are a prerequisite for economic growth. But if the indicators are no good proxies for institutions, they are equally inadequate to support 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. 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. It seems worth trying to separate them from each other. 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 does not teach us anything about the factual constraints in the real world. And if constitutional rules are useless as a benchmark, why should we be measuring 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). In order 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 that 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.5

The constitutional variables that are suggested by Glaeser et al. as more appropriate benchmarks if one is interested in long-term constraints are the following: two dealing with electoral systems [(i) “plurality” and (ii) “proportional representation”] and two dealing with judicial constraints on government, namely (iii) judicial independence (JI)

5 This last position is remotely reminiscent of Marx’ Feuerbach-Theses (“Philosophers have hitherto only interpreted the world in various ways; the point is, however, to change it.”).
and (iv) 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 (2003).\textsuperscript{6} Judicial independence is the average of three components, namely (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 (no, limited, full). These measures are praised as “objective” which 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.” They are thus making two statements: First, the two objective measures for constitutional constraints are only weakly correlated with the institutional measures that Glaeser et al. criticize or are not even correlated with them at all. They seem to believe that this is sufficient evidence for the inferiority of the institutional measures that they criticize. But they do not stop there. They argue, secondly, that neither of the preferred objective measures are correlated with income. They seem to believe that this is sufficient to claim that it is not institutions that cause growth.

This is unconvincing for a number of reasons: (1) I do not know any theory that would claim that any of the four constitutional measures would be sufficient to observe economic growth. (2) The two judicial measures rely exclusively on \textit{de jure} information. They might, hence, not reflect much about the real world.\textsuperscript{7} (3) The two electoral measures are averages for the period 1975 to 2000, whereas the two judicial measures are supposed to reflect the situation in 1995. Their dependent variable, economic growth, is, however, the average for the period 1960 to 2000. Now, it would be truly amazing would they find that a constitutional rule claimed to be valid in 1995 would have caused growth that started three and half decades before!

(4) A further problem with their approach is that they do not sufficiently distinguish between procedural and substantive aspects of institutions. They seem to expect that a high degree in the number of constraints on the executive (a procedural aspect) should translate into secure property rights (a substantive aspect). After describing that in

\textsuperscript{6} Notice the substantial overlap in the authors of this paper and Glaeser et al.

\textsuperscript{7} Their coding is crude and only distinguishes between tenure that is less than six years (coded 0), tenure that is more than six years but less than life (coded 1), and lifelong tenure (2). Whether “lifelong” means until death or a fixed retirement age is not mentioned. Witold Henisz (2000) has constructed a variable that captures the \textit{de facto} tenure of supreme court judges with very high precision. The partial correlation coefficient between his variable and the tenure variable constructed by Glaeser et al. is only 0.366.
Polity IV dictatorships like Cuba, North Korea but also Pinochet’s Chile get the worst score and communist countries like China and the U.S.S.R. fare better, they state: “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? The ideology of the communist countries did not support private property rights and the example primarily shows how flawed their implicit assumption is.

(5) Finally, Glaeser et al. (2004) note that “If 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.” They seem to believe that this insight is incompatible with the New Institutional Economics, yet it does not need to be. Many scholars within the NIE have, time and again, emphasized the necessity of formal – or external – institutions not being fundamentally at odds with informal – or internal – institutions. If the internal institutions of a society simply do not support the factual implementation of tight (constitutional) constraints on their governments, wonderful formal institutions will, most likely, have few beneficial effects.

This section has both summarized and criticized Glaeser et al. Two important messages are worth repeating: institutional measures should explicitly take the factual enforcement of the respective institution into account and they should be as objective as possible. Unfortunately, the four constitutional constraints relied upon by Glaeser et al. turn a blind eye on factual enforcement. The next section proposes a definition of institutions and discusses a number of possible implications.

3 Defining Institutions: A Proposal

The New Institutional Economics is still a relatively young research program. No generally accepted definition of institutions has emerged yet. The vast majority of scholars defines institutions as the rules of the game. North (1990, 3), e.g., 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. Apparently, 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 etc (1990; 6, 43, 83).

We would like to propose a definition that takes the difference between informal and formal rules on the one hand, and between rules and enforcement on the other, explicitly into account. It is inspired by Ostrom. Following her (1986, 5), 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 implications of this definition warrant particular emphasis: (1) „Commonly known“ implies that purely private rules do not qualify as rules. (2) Rules are the result of human action but not necessarily the outcome of deliberate human design.  

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. The criterion of formality refers to the rule component of institutions. We 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, institutions are called “external”, if rule-breaking is sanctioned by members of society, institutions are called “internal”. Among internal institutions, a more fine-grained taxonomy could focus on who does the sanctioning: is it done by unorganized actors (due, e.g., to norms of reciprocity) or by organizations (churches, chambers of commerce, private arbitration courts)? Internal institutions thus include mores, traditions, norms and so forth.

In their critique on frequently used measures of institutions, Glaeser et al. focus on a particular subset, namely political institutions without ever explicitly defining them. They explicitly refer to “democracy and other checks on government”. A distinction between political and economic institutions has also been used by others (Acemoglu et al. 2004). There, 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, a number of differences between political and economic institutions frequently occur: usually, economic institutions leave those willing to interact a number

---

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

9 Voigt and Kiwit (1998) contain a proposal for a more fine-grained taxonomy of internal institutions.
of ways how to structure their interaction. With regard to political institutions, it is not unfrequent that they require a specific behavior. Elections are held every so many years, votes are transformed into delegates in a specific way and so forth. Another difference is this: in case a contracting partner does not comply with a contract, the sanction that she will have to bear with is more or less precisely specified ex ante. With regard to political institutions, sanctions are usually not explicitly spelt out anywhere.

Whereas North emphasizes the difference between formal and informal rules, we emphasize the difference between internal and external sanctioning of rule-breakers, and the distinction between economic and political institutions uses the kind of interaction as a classification criterion. Although the distinction between the political and the economic sphere is not very precise in many cases (e.g., how to classify institutions constraining state owned enterprises?), it is closely related with a standard classification often used by legal scholars: economic institutions structure interactions in which all involved parties act as private actors. In case of a conflict, the state could come in as a neutral arbiter (judges). These are interactions based on private law. But interactions between private parties can also be non-voluntary, car accidents or theft e.g. Here, the state usually acts on behalf of the victim. These interactions are not based on but rather adjudicated according to criminal law (which is part of public law). Finally, there are interactions between any government representative and private actors: the state taxing or regulating citizens but may be also protecting them. 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 last section served to clarify our notion of institutions. Here, we go one step further: 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 spelt out explicitly. This is done here.

---

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

11 It is, e.g., unclear what sanctions a government will face that simply refuses to hold elections at the date provided for by the constitution.
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.\textsuperscript{12} If the rule reduces the number of admissible actions of an actor, then she might prefer not to be constrained by the rule because one of the non-admissible actions promises higher expected utility than the most attractive admissible action. In the absence of \textit{any} sanction, the agent is expected not to conform to the rule then. Enter sanctions and suppose that an actor will never be sanctioned after having complied with the relevant rules (in other words: we exclude judicial and other errors for the moment). The expected utility from breaking the rule needs to be complemented by the product of the probability of being sanctioned times the utility loss of the sanction (the fine, the prison term etc.). Given that the expected utility of rule-conforming behavior is higher than that of rule-breaking behavior, we would expect the agent to conform to the rule.

In order to know to what degree institutions constrain behavior, assigning expected utilities to the situations “comply with rule/not be sanctioned” and “not comply with rule/be sanctioned” is not sufficient. In addition, we would want to know the expected utility of any rule/sanction combination. Ideally, this would enable us to identify the “compliance elasticity” defined as the percentage increase in compliance after a one percent increase in sanctions.

Unfortunately, these expected utilities cannot be ascertained directly because the utilities of both compliance with the rule and 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 elections. The margin of error in predicting how various groups will react – in other words both the probability that they will act and the damage that this will cause to 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 all actors involved to change their behavior unilaterally. Uncertainty can induce actors to deviate unilaterally once but due to their disappointed expectations, they might revert to the equilibrium soon. When trying to measure \textit{de facto} institutions, it would therefore seem to make sense to observe factual behavior over a number of periods in order not to be misled by one time deviations from equilibrium behavior.

\textsuperscript{12} For the sake of completeness, we add that institutions that are to solve (pure) coordination games do not need to be endorsed with an additional sanction as non-coordination reduces one’s payoffs already – and serves as „built in“ sanction.
Until now, we have assumed that constraints are exogenously given. But thinking about economic institutions, it becomes immediately evident that this is not the case if they are determined by political institutions. And political institutions themselves are not exogenous either. Although in many countries formal constitutional change requires supermajorities, even constitutional constraints can be changed. If investment into lobbying for different rules promises high benefits, people will invest into changing the relevant formal constraints. To assume institutions to be quasi permanent over decades or even centuries is therefore simply besides the point.

With the definitions proposed above and the simple expected utility calculus in mind, we now move on to present a number of assumptions and implications. The attempt to measure institutions needs to be driven by underlying theory. These theoretical considerations will, hence, have an influence on how we propose to make institutions measurable later on.

Assumption 1: The effects of institutions are due both to their substantial content and their factual implementation.

Institutions help to order an otherwise unordered world. They allow actors to form expectations that have a good chance of turning out to be correct. This enables actors to develop a longer time horizon, to make long-term investment, to increase the division of labor and so forth. The assumption implies that this increased level of predictability can be brought about in two interdependent ways: (1) the substantial content of a rule (e.g., little protection of intellectual property, medium protection, high protection). (2) Suppose the substantial content provides only for a little protection of intellectual property. If this is meticulously enforced, then the implementation can be important in creating 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. In principle, it is, however, possible to think of the uncertainty reducing effect of implementation given some assumption on the strength of a specific rule.

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 effects resulting from the enforcement of a rule. Refraining from measuring de jure institutions altogether would imply the assumption that they are identical everywhere which is definitely wrong.

Corollary 1: Many creators of indicators seem to assume simple linear relationships between an institution and some outcome. If that were true, then the adequate policy
Implication would be to maximize (or minimize; but not to optimize) the respective content of an institution. Things are, however, likely to be more complicated than that: up to a certain level, the strength of a rule might have positive returns and only past that level might marginal returns be negative.\textsuperscript{13}

**Assumption 2:** The constraining effect of institutions largely depends on their factual implementation and enforcement. In order 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 formally constrain behavior substantively but are subsequently not or only partially enforced are expected to constrain factually observable behavior only to a limited degree. Factual enforcement depends on the behavior of those who are to enforce an institution.

**Implication 2:** Measures aiming at including the factual enforcement of institutions need to reflect the behavior of those who participate in their enforcement. These can be judges, prosecutors, but also the police and prison staff. But it can also be the press, lobby groups, the public at large etc. This means that the measurement of institutions is messy. Glaeser et al. 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 for institutions to be enforced, a certain behavior is indeed key. If one is interested in factually enforced institutions one needs, hence, to take behavior explicitly into account.

**Assumption 3:** Factual behavior is likely to be determined by more than a single institution. Internal institutions can re-enforce, but also weaken, the constraining effect of external institutions.

**Implication 3:** Institutions serve to structure specific recurring interaction situations. In order to understand the behavior of the 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 internally and externally implemented sanctions. In order 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

\textsuperscript{13} Arruñada (2007) makes this point with regard to the measures underlying the World Bank’s Doing Business studies. In particular, he argues that the underlying method only measures costs (and disregards potential benefits), measures only ex ante costs (and no ex post ones) and measures exclusively formal costs (forgetting necessary complements).
effect (the observed behavior) exclusively to an external institution. *Ex ante*, it cannot be excluded that external and internal institutions are conflicting, i.e. following an external institution can imply breaking an internal one and vice versa.\textsuperscript{14}

The argument in favor of looking at a number of possibly relevant institutions is, however, not equivalent with a plea for grouping various institutions together and then measuring some sort of “mixed bag”. It is important to try to measure single institutions in order to be able to attribute any potential effects as precisely as possible to specific institutions. Only this allows us to identify those institutions that drive the effects – and those that are only marginally relevant.\textsuperscript{15}

**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 recurrence on political institutions. Non-compliance of members of the administration with administrative law can be challenged via administrative courts. But who can enforce formal constitutional constraints against the government? Checks and balances can be interpreted as an attempt to reduce the expected utility of non-compliance. If legislators pass a new law that is not in accordance with the constitution, constitutional review via a constitutional court can lead to annulment of this legislation. But what happens if 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.

We just saw that the enforcement of formal political institutions is highly precarious. Here, we want to emphasize that the sanctioning of non-compliance with many institutions does not need to be constrained to the state: If a merchant has breached his contractual promises, he is likely to be sanctioned by many other merchants who refrain from entering into fresh contracts with him. This sanctioning by actors other than the representatives of the state promises to be potentially most important with regard to political institutions: precisely because their formal enforcement is highly precarious, political institutions will often only be factually enforced if there is a credible threat of being sanctioned informally in case of non-compliance.

\textsuperscript{14} Voigt (2004) deals with the possibility as well as with possible consequence of the coexistence of different property rights regimes.

\textsuperscript{15} Some concepts – such as judicial independence – might not be directly observable. In cases like that, 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.
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 demonstrations by the people at large, then the government might very well expect to be better off by breaking some formal constitutional constraint. If all the actors just mentioned are, however, expected to become active as a consequence of a government not complying with a formal constitutional rule, then 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 internal institutions. In order 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. In addition, 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 relying on the state. Exclusion from a relevant group is a sanction that has been applied 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. Because for their enforcement, they do not rely on the power of the state, 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 set external institutions that have a high chance of being factually implemented could thus be seriously constrained by the relevant internal institutions. Having identified external institutions that are robustly causing some desired results is only a necessary but not a sufficient condition to demand their introduction. Only if the valid internal institutions are such that one can expect the modified external institutions to be factually enforced, would one have a sufficient argument.\(^\text{16}\)

---

\(^{16}\) On top, a number of additional conditions should be satisfied such as the expected benefit from the introduction of a specific institution outweighs all costs connected with both the
In this section, we have not only proposed a taxonomy of institutions but we have also ventured a number of conjectures on the factors that might determine whether and to what degree formal constraints – such as constitutional rules setting up checks and balances – will be factually enforced. Taken literally, they would seem to make the measurement of institutions extremely messy – if not outright impossible. In the next section, we show that pragmatic ways of measuring institutions do exist.

5 A Pragmatic Approach Towards Measuring Institutions

Before proposing a pragmatic approach towards measuring institutions, let us spell out some basic conditions that need to be met to measure institutions. First, they need to be “perceptible” or “recognizable”. With regard to institutions formally passed as legislation this should not be a problem. Yet, in some countries legislation is only published with hefty delays or never. In countries without a written constitution, it is at times hard to ascertain what the constitutional constraints precisely are. It is almost a defining characteristic of informal or internal institutions that their precise content is nowhere published formally. This might make 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."
The graph above is one way to clarify the relationship between institutions and factual behavior. It has the following implications:

1. An actor who is subject to an institution needs to know the content of the respective underlying rule and understand its implications, i.e. needs to be able to judge whether a behaviour considered by her is in conformity with the rule component or not;
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 “sanctionor” must be able to determine whether an individual has complied with the rule or not;
4. Finally, an academic observer must be able to judge all of the above!

In addition, the assumption that individual behaviour in conformity with a rule will never be sanctioned might 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, e.g., be the case with regard to freedom of opinion.

We now move on to propose a pragmatic approach towards measuring institutions. The central message is: it is essential to measure factually implemented institutions and their measurement is a lot less messy than might have been expected as a consequence of the previous discussion.
We assume all actors to have unobservable preferences. Their behavior is, however, observable. In addition, an external observer can also 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. We assume 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. We need, however, be careful not to make the reverse statement. If we observe behavior that is in compliance with the relevant rule, we do not know whether compliance is due to the threat of a sanction – or simply reflects the preferences of the actor.

We now move on to 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. They are, however, presented in a different order.

(1) In order to estimate differences between behavior expected according to some institution and factually observed behavior, we first need to pick an institution whose effects we are interested in. Before starting to measure institutions, a clear and concise conception of the institution is key. This sounds self-evident but apparently, it is not. How else could 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 or 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 still aggregate single institutions into more encompassing indicators later on, but in order 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 containing a value judgment. It has often been argued that, e.g., “human rights” are a concept firmly rooted in Western civilization. Now, if we are 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

---

17 The „rule of law“ concept is notoriously difficult to define. A number of years ago, a bibliography on the concept named no less than *** titles discussing its contents. In the meantime, many more papers have appeared.

18 In order 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. But this is a follow-up problem and need not concern us here.
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 that behavior that would be observable were actors to comply with the institution. If judges are, e.g., appointed for twelve years, we would expect average tenure to be twelve years only corrected by judges who leave their positions voluntarily 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 the behavior in accordance 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 strive 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, then this country should score well with regard to predictability, as discussed above in section 3.

(3) Next, factually observed behavior needs to be measured somehow. With regard to economic institutions, this will often appear almost impossible as the behavior of thousands or even millions of actors would have to be taken into account. With regard to political institutions, this is frequently less problematic. Usually, there is only one head of government and only so many ministers. The number of potentially relevant actors is, hence, quite limited. In some instances, empirical complexity can be reduced by choosing an appropriate sample. An example could be 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 on the basis of all of them. These would then have to be corrected by voluntary retirements, by deaths in office and so forth. If one is interested in a large cross-country sample, this could easily amount to the necessity to trace the careers of hundreds of thousands of judges. In order to make the task less tedious, we have focused on the judges on the highest court of a country alone.  

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 this can have an influence on the entire legal development of a country.

Whether political institutions are factually implemented is not a question of a single point in time but rather of some period. Suppose the constitution guarantees judges that their incomes cannot be reduced. Whether this institution is factually implemented or not will depend on its being complied with for years or even decades. Likewise with

---

19 Note that this presupposes that it is always possible to name the „highest court“ without ambiguity which might be difficult in a number of countries.
their tenure: whether their factual tenure corresponds with their *de jure* tenure can only be decided after many years. To measure the factual implementation of institutions over very long periods seems also justified because predictability cannot be created overnight. Rather, predictability is conjectured to be a function of the number of periods over which a *de jure* institution has been factually enforced.

It seems difficult to name any “optimal” period for the measurement of the factual implementation of institutions. Various considerations play in: (1) For the measurement of some institutions, there is natural minimum period. If we want to ascertain if the factual tenure of supreme court justices accord to their formally ensured tenure, we need to take at least the formal period into account (say 9 or 12 years). (2) The time period we choose 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 this reduces the number of observations too far, we might want reduce this to, say, 20 or 40 years. (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.  

(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 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. 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 seems to make more sense to describe a specific situation in which their security is at stake, and then inquire into how many days it would take and how much money would be involved to get ones’ rights. This has, e.g., been done 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.22

The collection of “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 intention will not be readily available. The availability of numbers 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 identical measures. This is, however, only possible if the criteria, the coding rules, the various components of a measure etc. are all disclosed, in other words if the construction of the measure is transparent. Unfortunately, some of the currently most frequently used measures are completely intransparent. The Freedom House indicators are a case in point.

(6) In order to identify the “distance” between behavior to be expected according to the letter of the law and factually observed behavior, some measuring rod is needed. The adequate 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 it might make sense to construct a simple dummy variable. Quite often, however, deviation from a de jure institution is, rather, 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 whether the measuring rod should have a linear or a log form.

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 difficult to control for using objective controls.

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 get 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 – and no objective measure how long it really does take to get these cases settled.
Sometimes, we are interested in the effects of a notion that is broader than just one single institution. Examples that come to mind are judicial independence or procedural formalism. In these cases, the notion one is interested in is often not directly observable. In order 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 amongst each other by different levels of correlation into one indicator follows a systematic 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 clearly keep our theoretical concepts apart. Within the factors, it is not one single variable which 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) have drawn on factor analysis to identify the principal components of procedural formalism. In our own work, we have used factor analysis to tease out the various dimensions hiding behind different indicators of federalism and fiscal decentralization (Blume and Voigt 2008).

To sum up the most important message of this section, we would like to use table 2 which is a simple 2x2 matrix: The most serious challenge in making institutions measurable lies in the dark cell. Although this will demand most resources, the cell shaded grey should also be taken into account as only data on both cells allows comparing them.

<table>
<thead>
<tr>
<th>Type of institution:</th>
<th>Way to measure:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>subjective</td>
</tr>
<tr>
<td>De jure</td>
<td></td>
</tr>
<tr>
<td>De facto</td>
<td></td>
</tr>
</tbody>
</table>
6 How Not to Measure Institutions

This section is not a comprehensive survey of widely used measures of institutions. Rather, it critically discusses only a small number of 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 that have organized their judiciary in highly inefficient ways. Trying to make any inferences from, e.g., the “budget allocated to the judicial system per inhabitant in 2004” on the quality of the judicial system seems rather daring. Input measures can be useful information for a number of topics but without information on de facto implementation, they are hard to use. Such measures are, e.g., provided by the so-called CEPEJ project of the European Council.

(2) Declared output if incentives to declare are not equally distributed. The UN Office on Drugs and Crime regularly publishes a Survey on Crime Trends and Operations of the Criminal Justice System. These surveys contain lots of highly interesting and relevant data, but they should not be used as proxies for the functioning of any institutions. 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 seems to give the data an objective blend. Before putting too much trust into them, a number of questions should be dealt with: it might be necessary not only to control for population but also for the object of some crimes. In Romania, e.g., 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 our main point still is a different one: If the police are perceived as highly unlikely to ever find any criminals and moreover perhaps even as corrupt, then the likelihood of ever notifying the police is likely to suffer. 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 (Kaufmann et al.) have been criticized both very heavily and very
explicitly.23 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. The indicators are not supplied as absolute values but as rankings. Once the number of incorporated countries changes, the ranking of a country can change without any substantial change having taken place within the country.

7 Conclusions and Outlook

In this paper, we have argued that measures of institutions should be precise, objective and take into account de jure as well as de facto elements. We have hypothesized 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 would loom large. We have also pointed out the ability to measure institutions combined with econometric findings showing their significance for explaining variation in dependent variables is no way sufficient to assume 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 measure are not sufficient to rebute the hypothesis that institutions do (not) matter. Major data collection exercises are thus ahead of us.

Here are a number of examples for areas in which better data could increase our knowledge:

(1) One of the first fields in which objective measures were introduced was central bank independence (cbi), the question, of course, whether higher degrees of cbi 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 has been used frequently is their turnover rate. This is, of course,

\[\text{Here is a list of some critics: Arndt and Oman (2006), Knack (2006), Thomas (2006), Kurtz and Schrank (2007a, 2007b). Kaufmann et al. (2007) is a reply to critics.}\]
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 any) always been followed, how has the salary of the governors developed, how the budget of the entire bank and so forth.

(2) The central bank can be interpreted as an (independent) regulatory agency endowed with the task to provide stable money. Many of the reasons for making central banks independent also apply to other regulatory agencies. This implies that their independence – 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 of the judiciary can be interpreted as solving credible commitment problems by actors wanting to make promises, be they private actors or government representatives. This means that frequently, there is also time-consistency problems involved – and the independence of the courts can be analyzed using criteria very similar to those used for the analysis of cbi. My own indicators for *de jure* and *de facto* judicial independence are examples for that. But it would be extremely helpful to have an organization like the World Bank take up these indicators and extend them to a more countries, double check the answers etc.

(4) Depending on the resources available for an extensive data gathering exercise, one might even want to rethink whether it would be possible to measure some of the central institutions making up the rule of law. These could include its generality, its abstractness, its certainty but also 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 both monetary and expert resources are needed, large international organizations like the World Bank are the most likely collectors – and disseminators – of such data.
References


Blume, L. and S. Voigt (2008); Federalism and Decentralization – A Critical Survey of Frequently Used Indicators, mimeo, Philipps University Marburg.


